

# From School Buses to Start Times: Driving Policy With Optimization

Dimitris Bertsimas, Arthur Delarue, Sebastien Martin

October 2018\*

## Abstract

Maintaining a fleet of buses to transport students to school is a major expense for school districts. In order to reduce costs by reusing buses between schools, many districts spread start times across the morning. However, assigning each school a time involves estimating the impact on transportation costs and reconciling additional competing objectives. Facing this intricate optimization problem, school districts must resort to ad hoc approaches, which can be expensive, inequitable, and even detrimental to student health. For example, there is medical evidence that early high school starts are impacting the development of an entire generation of students and constitute a major public health crisis. We present the first optimization model for the School Time Selection Problem (STSP), which relies on a new school bus routing algorithm we call BiRD (Bi-objective Routing Decomposition). BiRD leverages a natural decomposition of the routing problem, computing and combining subproblem solutions via mixed-integer optimization. It significantly outperforms state-of-the-art routing methods, and its implementation in Boston has led to \$5 million in yearly savings, maintaining service quality for students despite a 50-bus fleet reduction. Using BiRD, we construct a tractable proxy to transportation costs, allowing the formulation of the STSP as a multi-objective Generalized Quadratic Assignment Problem. Local search methods provide high-quality solutions, allowing school districts to explore tradeoffs between competing priorities and choose times that best fulfill community needs. In December 2017, the development of this method led the Boston School Committee to unanimously approve the first school start time reform in thirty years.

In the twenty-first century, school districts across the US face a wide array of challenging problems on a daily basis, from adjusting to the digital age to educating an increasingly diverse and multicultural student body. Yet perhaps the most complicated decision that administrators face is seemingly the most innocuous: determining what time each school in the district should start in the morning and end in the afternoon.

The issue of choosing appropriate school “bell times” has received increased attention in recent years, as too-early start times have been linked to a wide array of health issues among teenagers, including diminished academic achievement [6] and cognitive ability [11, 24], and increased rates

---

\*2nd round of revisions, Proceedings of the National Academy of Science (PNAS)

of obesity [7], depression [17], and traffic accidents [12]. Indeed, changes in the body’s circadian clock during puberty effectively prevent adolescents from getting adequate sleep early in the night [10]. While the American Academy of Pediatrics recommends that teenagers not start their school day before 8:30AM, a recent CDC report found that just 17.7% of U.S. high schools comply [33]. Some experts estimate that over the next ten years, the dire public health implications of early high school start times could impact the U.S. economy by over \$80 billion [19].

Moreover, research suggests that these repercussions disproportionately affect the most economically disadvantaged students [16]. As achievement gaps between students from different backgrounds remain stark [34], research has consistently found systematic biases, largely on racial lines [9], that partially explain these gaps. For example, school bell times can suffer from such biases, as is the case in Boston [29].

For decades, school districts across America have considered ways to adjust their bell times and solve these issues in a fair way. However, the sheer complexity of the problem is a major obstacle to change. School districts typically struggle with balancing many competing objectives, including student health, special education programs, parent and staff schedules, state and federal regulations, and public externalities [21].

Perhaps the greatest obstacle to adjusting school bell times is the effect of changes on school transportation. Over 50 percent of U.S. schoolchildren rely on an army of half a million yellow school buses to travel to and from school every day. In Boston, where specialized programs draw students from all over the city and traffic is often at a standstill, transportation accounts for over 10% of the district’s \$1 billion budget. To reduce transportation spending, school districts such as Boston stagger the start and end times of different schools, allowing vehicles to be re-used several times throughout the day. Because many school districts construct bus routes by hand, it is exceedingly difficult for them to evaluate the impact of bell time changes on bus costs, let alone find a set of bell times that satisfies all of the district’s objectives without inflating the budget. No matter how unpalatable, the status quo is often the only viable option. In addition, because of the impossibility of systemwide change, districts may experiment with a piecemeal approach to bell time change, where the most vocal and best connected schools may benefit the most.

The problem of school bus routing has been addressed extensively [14, 25]. It is typically decomposed into three main subproblems (see Fig. 1D-F): stop assignment, i.e., choosing locations where students will walk from their homes to get picked up; bus routing, i.e., linking stops together into bus trips; and bus scheduling, i.e., combining bus trips into a route that can be served by a

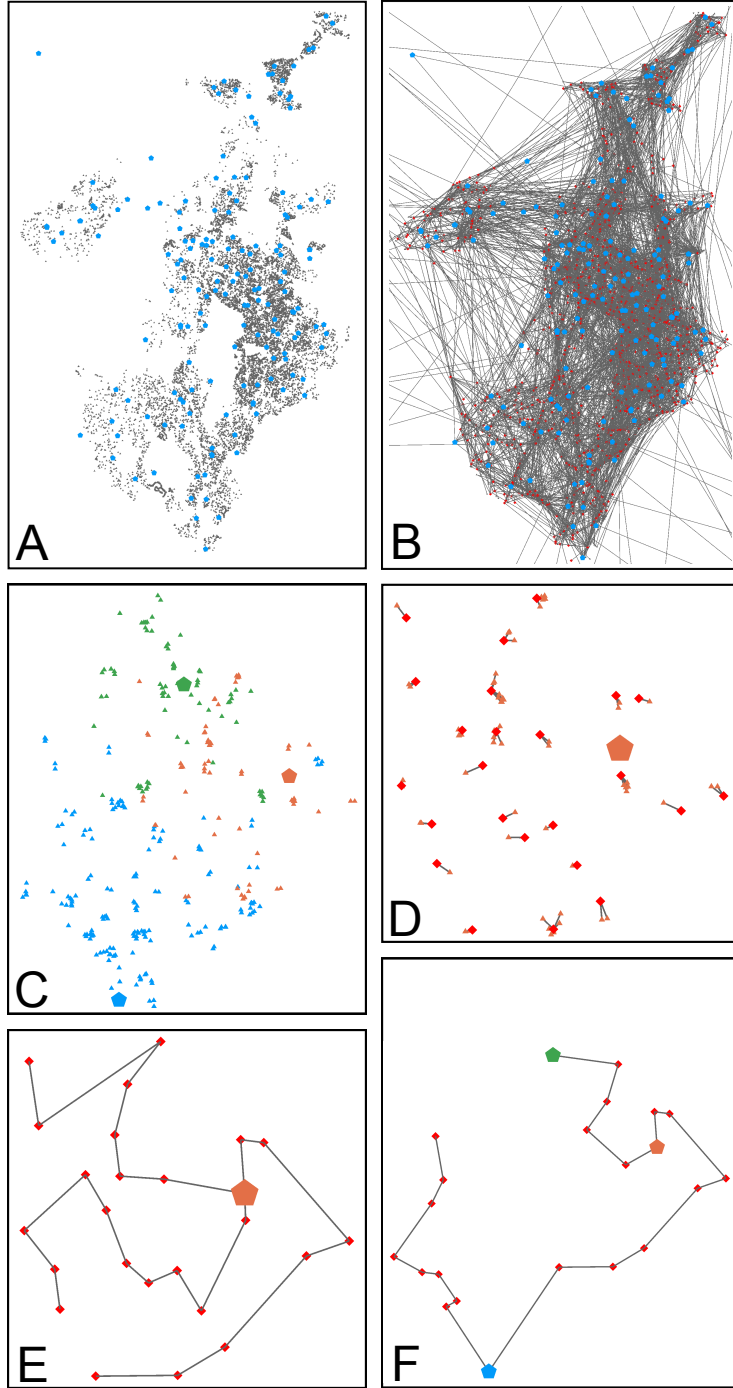


Figure 1: Geographic visualization of the school bus routing problem (and subproblems). **(A)** BPS 2017-18 data (anonymized) with gray triangles representing students and blue pentagons representing schools. **(B)** Sample BPS routing solution, with schools as blue pentagons, bus stops as red squares, and lines connecting bus stops that are served in sequence by the same bus, illustrating the complexity of Boston school transportation. **(C)** Small synthetic district (3 schools); students (triangles) are the same color as their assigned schools (pentagons). **(DEF)** Example of 3 main routing steps in this district: stop assignment **(D)**, where students (triangles) attending the orange school (pentagon) are shown connected to their assigned stops (red squares); one-school routing **(E)**, where all bus stops for the orange school are connected into bus trips; and bus scheduling between multiple schools **(F)**, where three trips (one from each school) are connected into a single bus itinerary.

single bus. State-of-the-art optimization algorithms exist for these subproblems in isolation [28, 18]. However, the literature on optimally combining subproblem solutions is less extensive. Approaches typically involve formulating the school bus routing problem as a large combinatorial optimization problem which can be solved using metaheuristics, including local search [31], simulated annealing [8], and special-purpose vehicle routing heuristics [5, 4]. Special-purpose algorithms have also been designed to address variants of the school bus routing problem, allowing “mixed loads” – students from different schools riding the bus together [31, 4, 26], bus transfers [3], or arrival time windows [18, 31, 8, 26].

Unfortunately, the approaches above are rarely applied in practice, as tractable general-purpose algorithms often fail to consider additional constraints (fleet heterogeneity, student-specific needs), resulting in practically unworkable solutions. In addition, no existing algorithms address bell time selection in conjunction with school bus routing [18]. To our knowledge, the School Time Selection Problem (STSP) has never even been mathematically formulated, much less solved.

This work presents the first model for the STSP, allowing the joint optimization of school bell times with school bus routes. We first develop a new school bus routing algorithm called BiRD (Bi-objective Routing Decomposition) which bridges the gap between standard subproblems to find better solutions. We then propose a mathematical formulation of the STSP, a multi-objective approach that can model any number of community objectives as well as transportation costs using BiRD.

BiRD outperforms state-of-the-art methods by 14% on average on benchmark data sets, and allowed Boston Public Schools (BPS) to take 50 buses off the road and save almost \$5 million in the fall of 2017, without increasing the average student’s walking or riding times. Our modeling approach to the STSP, along with the successful implementation of BiRD, led the Boston School Committee to reconsider start time policies for the first time since 1990, unanimously approving a comprehensive reform prioritizing student health in December 2017. Our STSP model was used by BPS to evaluate the impact of many different scenarios and ultimately propose new bell times for all 125 BPS schools. These start times are being reviewed at the request of parent groups, and our approach remains central to Boston start time policy.

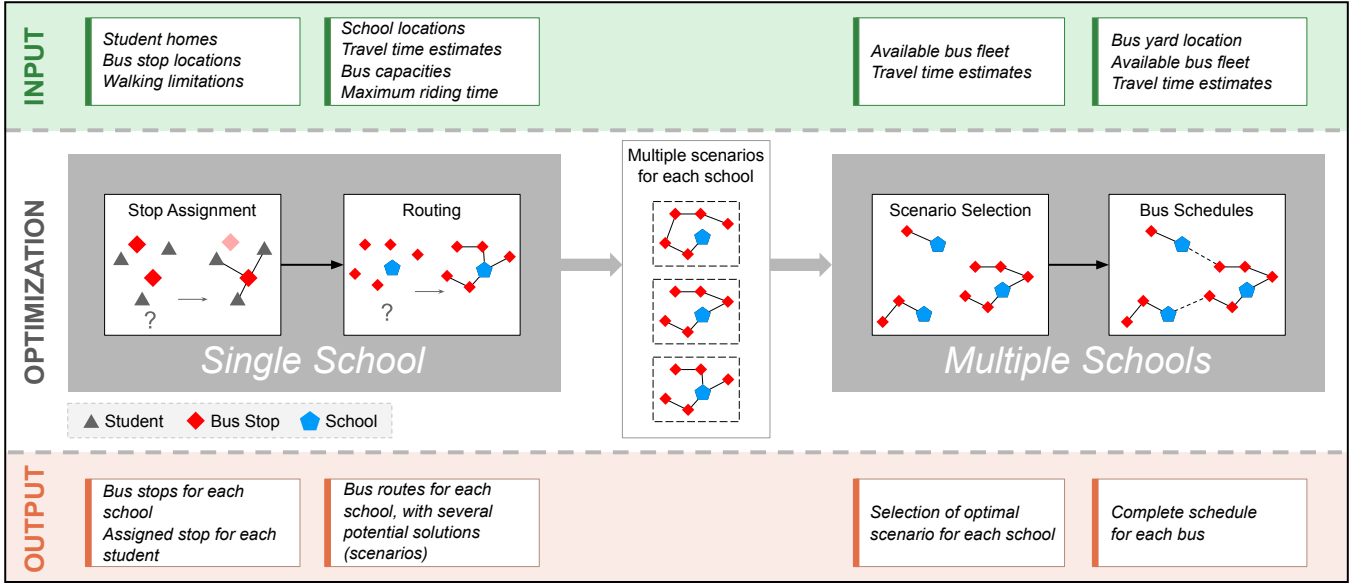


Figure 2: Overview of BiRD algorithm. On the left, the single-school problem can be divided into the two subproblems of stop assignment and single-school routing; on the right, the multi-school problem can be divided into the two subproblems of scenario selection and bus scheduling. The generation of not one, but several routing scenarios for each school, and the subsequent joint selection of a single scenario for each school, bridge the divide between the single-school and multi-school problems.

## School Transportation: A BiRD’s Eye View

Solving the school bus routing problem means assigning students to stops near their homes, selecting which bus will pick them up and in what order (keeping in mind that a bus only carries students for one school, but can serve several schools in succession thanks to staggered bell times), in a way that minimizes the overall number of buses, or another objective of interest. We show an example of a school district (BPS) in Fig. 1A and of a model school district that mimics the real setting in Fig. 1C and in the SI Appendix, Fig. S2.

The BiRD algorithm consists of several steps (see Fig. 2) for which we develop optimization-based approaches, implemented with modern software tools [2, 15] and available online [13]. For clarity, we focus on the morning problem, but our algorithm generalizes to the afternoon (see SI Appendix). Because problem details often vary between districts, it may be advantageous to adjust some steps to changes in the problem setting. BiRD’s defining feature is thus the decomposition of the problem, and in particular the scenario selection step which bridges the gap between the single-school and multi-school subproblems.

## Single-School Problem

To assign students to stops (Fig. 1D), we use an integer optimization formulation of the assignment problem, with maximum walking distance constraints. We minimize the overall number of stops because (i) it simplifies bus trips and (ii) the minimum pickup time at a stop is typically high, even if the stop has few students. When long bus routes span the entire city, as in Boston (see Fig. 1B), stop assignment has a negligible effect on the macroscopic quality of the routing solution. Our formulation can include additional objectives, such as the total student walking distance, and can exclude stop assignments that require students to cross major arteries or unsafe areas (see SI Appendix, Stop Assignment).

We then use an insertion-based algorithm to connect sequences of stops into feasible bus trips (Fig. 1E). We use integer optimization to combine these feasible trips with a minimum number of buses, with a set cover formulation reminiscent of crew scheduling problems [30] (see SI Appendix, Single-School Routing). Our method has the flexibility to handle practical modifications in the routing problem, from vehicles with different capacities to student-bus compatibility restrictions (e.g. students in a wheelchair need a bus with a special ramp/lift). In principle, the modularity of the overall algorithm means that the single-school routing algorithm can be replaced with any state-of-the-art vehicle routing method.

## Routing Multiple Schools

We use the single-school routing method to generate not one, but several varied optimized routing scenarios for each school, in order to select the best one for the system. In particular, we consider several scenarios on the Pareto frontier of two objectives (hence the name of Bi-objective Routing Decomposition), number of buses and average riding time. This tradeoff makes sense because shorter routes are more easily connected into bus schedules.

Then, we first jointly select one scenario for each school in a way that favors maximal re-use of buses from school to school (Fig. 2), by formulating an integer optimization problem with network flow structure that seeks to minimize the number of buses at the scale of the entire district (see SI Appendix, Scenario Selection). Given one routing scenario for each school, we can then solve another integer optimization problem to identify a trip-by-trip itinerary for each bus in the fleet (Fig. 1F). In this final subproblem, we optimize the number of buses jointly in the morning and in the afternoon (see SI Appendix, Bus Scheduling).

## Evaluating the Routing Algorithm

We compare BiRD to methods from the literature [4, 26, 8] on a published set of 48 benchmark data sets [26], where the objective is to minimize the total number of buses used. We outperform all existing methods on 96% of instances, with a 14% average improvement over the best existing method on each instance. The scenario selection step is central to this improvement: computational experiments (see SI Appendix, Routing Experiments) indicate that BiRD’s performance improves by 20% when we compute two different routing scenarios for each school and select the best one by considering the whole system, as opposed to using the best scenario for each school. Intuitively, what is optimal for one school may not be optimal for the entire system, motivating the bi-objective decomposition approach.

## Application in Boston

BPS has the highest transportation expenditure per student in the U.S., with rising costs due in part to narrow streets and infamous rush-hour traffic, a large fraction of special education students, and a complicated history of school desegregation. In addition, over the last decade BPS has adopted a “controlled choice” approach to school selection, which gives parents greater latitude in selecting a public school while promoting fairness across the district [1, 27]. As a result of this policy, some schools may draw students from far across the city, further complicating the school transportation problem and driving up costs.

Before we started working with BPS, bus routes for 125 public schools and over 80 private and charter schools were computed and maintained manually. BiRD’s ability to incorporate district-specific constraints (including four different bus types, and only one compatible with wheelchairs) was essential in producing a practical solution. In the end, we solved the Boston school bus routing problem using only 530 buses, against 650 for the manual solution. This represents an 18% reduction, with estimated cost savings in the range of \$10 to \$15 million. To ensure a smooth transition, BPS decided to only take 50 buses off the road in the first year of implementation, still amounting to a hefty \$5 million in cost savings [22]. Despite the smaller number of buses, the average student ride time stayed constant from 2016-2017 (around 23 minutes).

# Formulating the STSP

Selecting bell times is a complex policy problem with many stakeholders. We first focus on the interplay with transportation, since computing school bus routes is a necessary component of bell time selection. For instance, it is of interest to evaluate transportation costs when each school  $S$  is assigned a particular bell time  $t_S$ . However, there are too many possibilities to explore in practice (exponential in the number of schools). Instead, we develop a general formulation for the STSP, which contains a tractable proxy for transportation cost constructed using BiRD. We show how to include other community objectives in the next section.

## Transportation Costs

A key factor in an optimized school bus routing solution is the “compatibility” of pairs of trips, i.e., how easy it is for a single bus to serve them with minimum idle time in between. We define a trip compatibility cost that trades off (a) the feasibility of a bus serving the two trips sequentially, and (b) the amount of idle or empty driving time involved, with tradeoff parameters that depend on characteristics of the school district, and can be found using cross-validation. Then, for any pair of schools  $S$  and  $S'$ , we can define a routing pairwise affinity cost  $c_{S,t,S',t'}^{\text{routing}}$  that is the sum of the compatibility costs between every trip in every routing scenario for  $S$  at time  $t$  and  $S'$  at  $t'$  (see SI Appendix, Transportation Costs).

## Optimizing

Because its objective function only includes pairwise affinity costs, our model of the STSP is a special case of the Generalized Quadratic Assignment Problem (GQAP) [20]. Because even small instances of the GQAP can be intractable, we develop a simple local improvement heuristic that works well in practice. Given a initial bell times, we select a random subset of schools. The problem of finding the optimal start times for this subset, while fixing all other schools’ start times, is also a GQAP.

We can then solve this restricted GQAP problem using mixed-integer optimization to obtain a new set of bell times, in seconds for small enough subsets. We repeat the operation with new random subsets until convergence. Results on synthetic data suggest that a subset size of one gives near-optimal results, if the local improvement heuristic is run several times with random starting points. We note that the heuristic is interpretable: with a subset size of  $n$ , a solution obtained



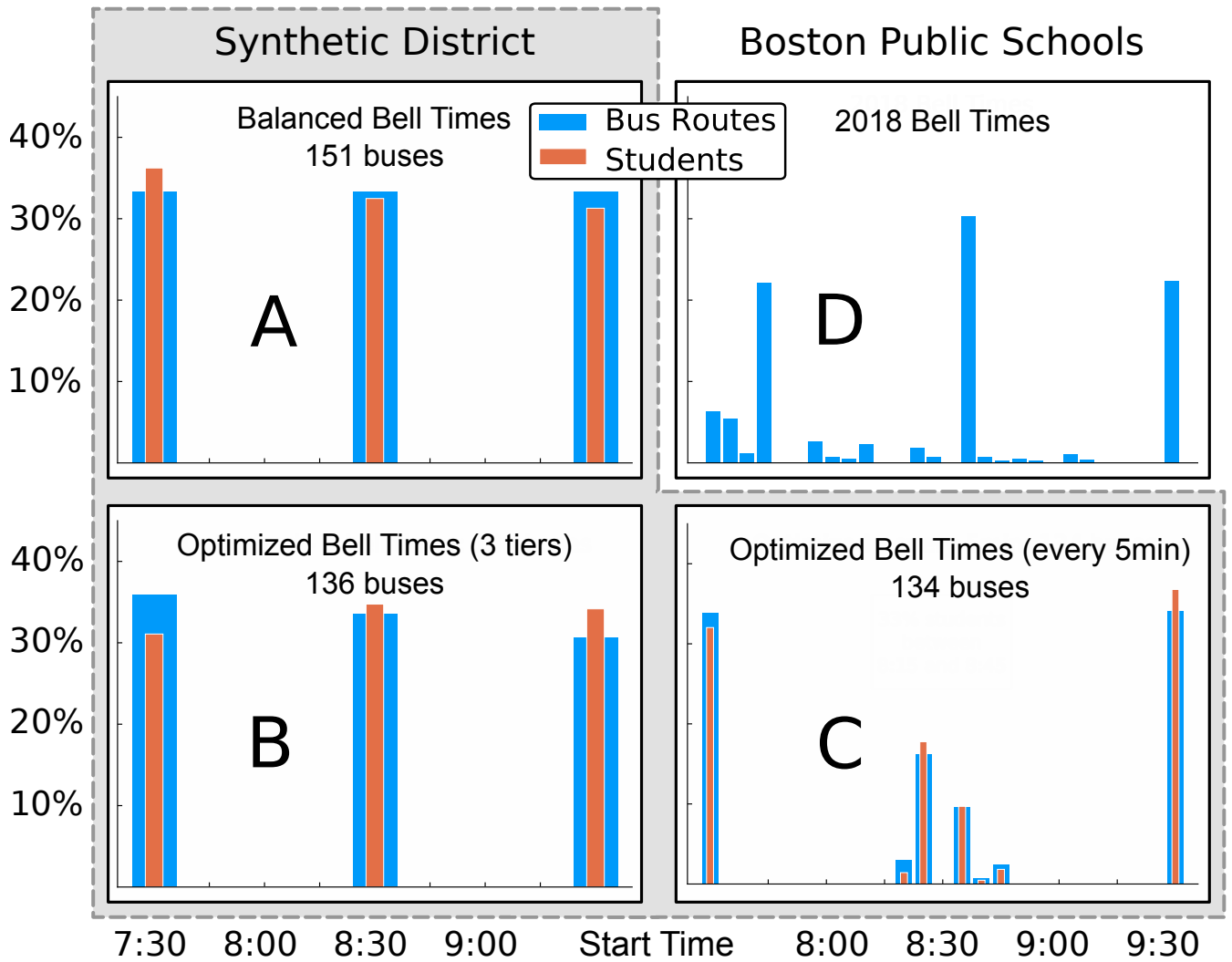


Figure 3: Bell time optimization. Comparison of 3 bell time optimization strategies on a synthetic district. When only three bell times are allowed, balancing the number of routes across bell times (A) works well, but is typically beaten by routing compatibility optimization (B). Even better solutions can be obtained by allowing more bell times in the middle tier (C). In comparison, BPS bell times are not even balanced (D).

after convergence can only be improved by changing the bell times of at least  $n + 1$  schools.

## Evaluating three-tier systems

In many districts, such as Boston (Fig. 3D), start times are separated into three equally spaced “tiers” (e.g. 7:30AM, 8:30AM and 9:30AM). Such a system allows each bus to serve up to three schools every morning [23], so districts will typically try to balance the number of bus trips across all three tiers. Our method allows us to quantify the empirical behavior of this intuitive idea.

Simulations suggest that optimizing three-tier bell times using our algorithm (Fig. 3A) yields

an 11% cost improvement over simply balancing the number of bus routes across tiers (Fig. 3B), which is already better than what school districts typically do (Fig. 3D). Distributing schools across tiers without accounting for geography/routing compatibility is suboptimal.

Furthermore, a three-tier system is not necessarily the right answer *per se*. Fig. 3C shows that allowing many possible start times for the middle tier (5-minute intervals between 8:00AM and 9:00AM) can yield a 1-2% improvement over the standard three-tier optimized solution (Fig. 3B). Interestingly, no school starts at 8:30AM in this system. Though tiered bell times are popular because of their simplicity, algorithmic tools such as ours suggest that better solutions exist. For instance, in Boston, we can find a bell time solution that requires just 450 buses, which represents a 15% improvement over the number of buses obtained without changing the bell times, and a 31% improvement over the number of buses used by BPS in the 2016-2017 school year.

## Bell Times in Practice

In a real district, bell time selection goes far beyond minimizing the number of buses, as we found in our work with BPS. For context, Boston’s existing bell time policy, enacted in 1990, split the public schools into three tiers, with start times of 7:30AM, 8:30AM, and 9:30AM, stipulating that tiers would rotate through the start times every 5 years. Unfortunately, this policy was never enforced, and the bell times assigned in 1990 mostly remain today.

These bell times are flawed. First, because they have remained static while school demographics have evolved, they have contributed to the steady rise of the BPS transportation budget over the last decade. Second, over 74% of high school students currently start school before 8:00AM. Many studies have shown that the negative effects of early high school starts are magnified in economically fragile students [16]. However, in Boston such students have worse bell times, on average, than economically advantaged students [29]. In Fig. 4, we see for example that economically disadvantaged high school students are considerably more likely to start before 7:30AM than other high school students.

## Gridlock

The Boston status quo has persisted for decades despite its shortcomings. Indeed, bell time selection is intrinsically difficult because stakeholders cannot agree on what is best for everyone. Figs. 5B and 5C show community preferences for different start times across all public schools,

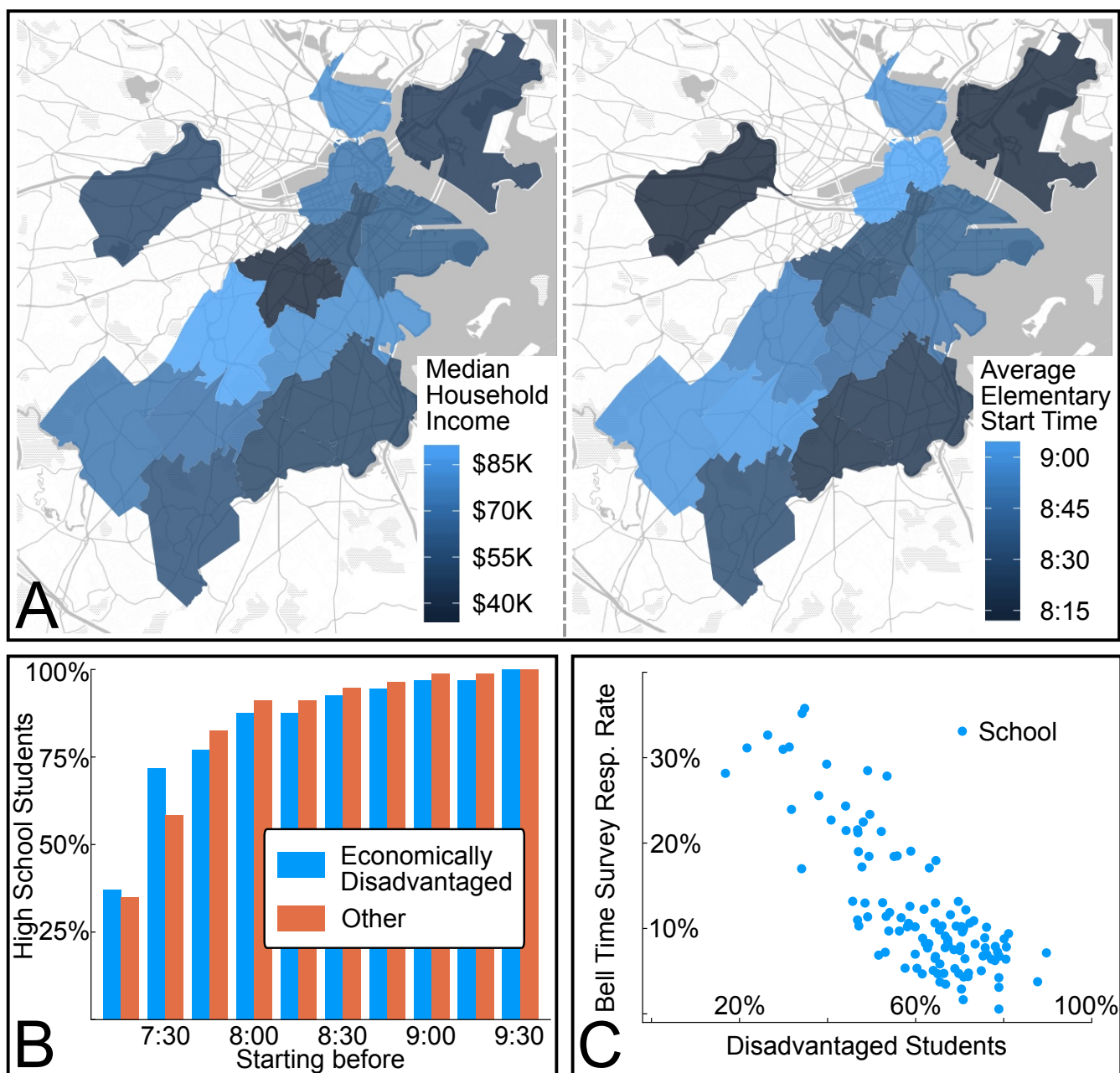


Figure 4: Equity and current start times in Boston. **(A)** Maps of Boston, with neighborhoods shaded by median household income (ACS) and average elementary school start time. Elementary school students start later in wealthier neighborhoods (0.78 correlation between household income and school start time). **(B)** Cumulative distribution of high school students starting before different times in the morning (comparing economically disadvantaged and other students). Start times skew early for economically disadvantaged high school students. **(C)** BPS Community Survey response rate by school, shown against fraction of disadvantaged students attending the school. Economically fragile populations have a lower bell time survey response rate.

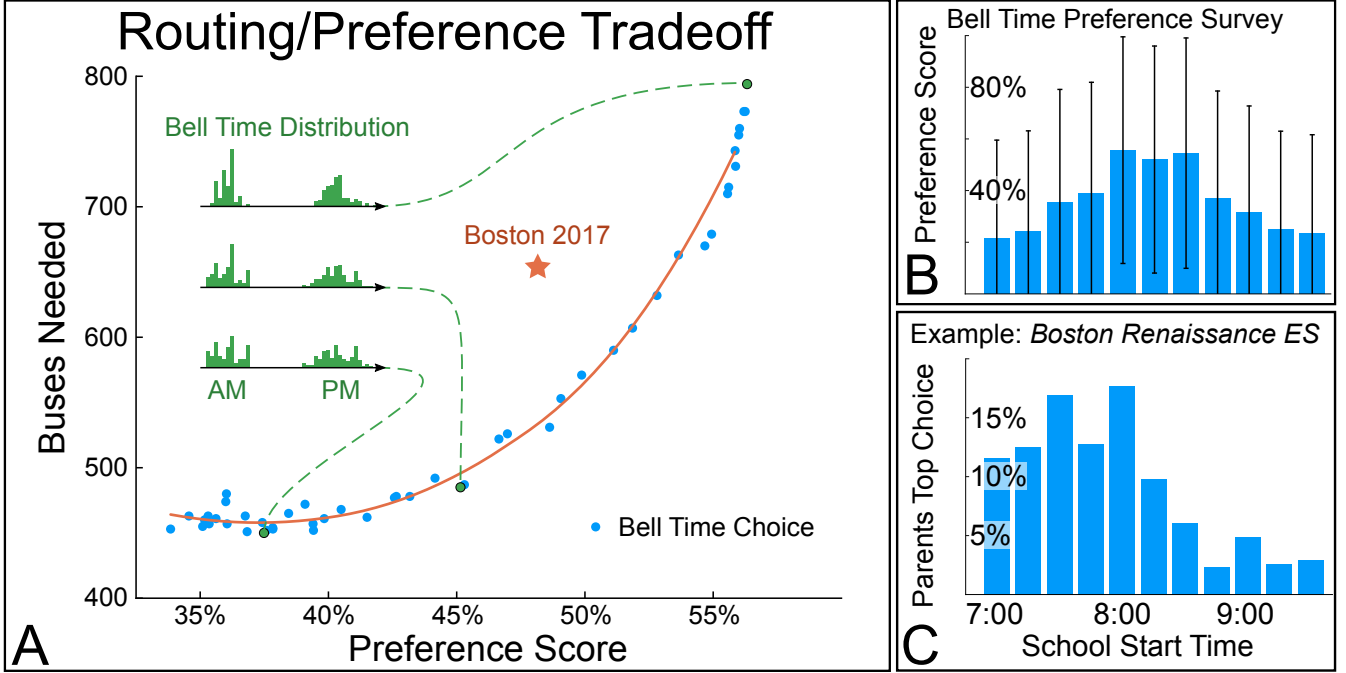


Figure 5: Optimizing preferences is hard. (A) Tradeoff curve derived by our algorithm between preference score (metric of community satisfaction – see the SI Appendix, Boston Community Survey) and transportation cost, along with three sets of bell times along the curve. Even a slight improvement in satisfaction comes at a high cost. (B) District-wide preference score of each bell time, showing that parents typically prefer 8:00AM to 8:45AM start times, with high variance. (C) Fraction of parents at a particular school who prefer each bell time. Even within a single school, agreement is hard to come by.

	Buses	Early HS	Late ES	Survey score	Bell time distribution
Current	650	74%	33%	48%	
NewRoutes	530	74%	33%	48%	
LowCost	450	43%	27%	37%	
MaxSurvey	934	0%	8%	56%	
Optimal	481	6%	15%	40%	

Figure 6: Bell time selection tradeoffs. Sample of a few scenarios considered by BPS. Current start times (with or without new routes) have many high school students starting before 8:00AM (Early HS) and elementary school students ending after 4:00PM (Late ES), mediocre community satisfaction (survey score), and a suboptimal bell time distribution both in the morning and in the afternoon (histogram weighted by students). The three other scenarios present different tradeoffs between the bell time objectives – BPS chose the “Optimal” scenario.

obtained through a BPS survey. Though families and school staff tend to favor start times between 8:00AM and 8:30AM, the displayed preferences are mostly characterized by broad disagreement, even within a single school (Fig. 5C). Any bell time for any school is sure to have both fervent supporters and vehement critics.

School districts have no hope of satisfying all, or even most, of their constituents. Moreover, the cost of even trying to satisfy the individual preferences of parents and staff can be prohibitive: Fig. 5A shows that each additional point of community satisfaction in Boston can cost dozens of additional buses and tens of millions of taxpayer dollars.

For BPS, the tradeoff curve in Fig. 5A represented a paradigm shift, the first time the district could visualize, or even quantify, any of the tradeoffs of bell time policymaking. The curve illustrates our model’s first use: providing a district the quantitative support necessary to understand the problem and make the best decision.

## The Greater Good

Though stakeholders have many competing personal priorities, they often agree on broader goals, such as having fair and equitable bell times or reinvesting saved transportation costs into schools. Starting in 2016, BPS led an engagement process aiming to understand broad community values. The results suggested four main objectives: to maximize how many high school students start after 8:00AM, minimize how many elementary school students end after 4:00PM, prioritize schools with high special education needs, and reinvest transportation savings into classrooms, while achieving these objectives in an equitable manner.

In the general case, solving the STSP in practice means optimizing a set of several objectives, such as the ones outlined above. We call an objective GQAP-representable if it can be represented using only single affinity costs  $c_{S,t}$  (representing the aversion of school  $S$  for bell time  $t$ ) and pairwise affinity costs  $c_{S,t,S',t'}$ . We find that the GQAP framework has sufficient modeling power to represent all the objectives and constraints that interest school districts in general (see SI Appendix, GQAP-Representable Objectives) and Boston in particular.

Typically, school districts will wish to balance multiple GQAP-representable objectives, including transportation costs. As is usual in multi-objective optimization, we consider that the final cost function to optimize is a weighted average of the district’s different (GQAP-representable) objectives, with weights indicating policy makers’ priorities.

We explored tens of thousands of tradeoffs for BPS, such as those presented in Fig. 6. We

notice that in Boston, reducing both the number of high school students starting too early and the number of elementary school students ending too late can be done at little to no cost.

## Application in Boston

In December 2017, the Boston School Committee unanimously approved a new policy [32], stipulating that all future bell time solutions should optimize the verifiable criteria described above, paving the way for algorithmic bell time selection. Our flexible methodology allowed us to take into account a number of very specific constraints, e.g. preventing large neighboring high schools to dismiss at the same time (which could create unsafe situations at neighboring MBTA stops). In the end, the proposed bell times (see Fig. 6) reduced the number of high school students starting before 8:00AM from 74% to 6%, and the number of elementary school students dismissing after 4:00 from 33% to 15%. The plan also led to an estimated reinvestment of up to \$18 million into classrooms. Due to the significant amount of change under this new plan, and in response to protests by some families, BPS delayed the plan’s implementation to allow more time to adjust the objective weights and constraints. As BPS continues to gather community input, the legitimate concerns raised by these families can be modeled as objectives within our general formulation and integrated within our framework.

Ultimately, using an algorithm for bell time selection at the scale of a city allows leaders to thoroughly evaluate their options, and empowers them to make decisions based not on the political whims of special interest groups, but on an objective standard agreed upon by the community.

## Acknowledgments

We would like to thank our partners at Boston Public Schools for their tireless efforts on behalf of the children of Boston. We would particularly like to express our admiration and respect for John Hanlon and Will Eger, whose single-minded devotion to public service should be an example to the world. Google Maps helped make this work possible by providing point-to-point travel time estimates for all locations in Boston. The Julia language and its JuMP extension for optimization significantly eased the implementation of our algorithms.



# References

- [1] Atila Abdulkadiroğlu, Parag A Pathak, Alvin E Roth, and Tayfun Sönmez. The boston public school match. *American Economic Review*, 95(2):368–371, 2005.
- [2] Jeff Bezanson, Alan Edelman, Stefan Karpinski, and Viral B Shah. Julia: A fresh approach to numerical computing. *SIAM review*, 59(1):65–98, 2017.
- [3] Michael Bögl, Karl F Doerner, and Sophie N Parragh. The school bus routing and scheduling problem with transfers. *Networks*, 65(2):180–203, 2015.
- [4] Jeffrey Braca, Julien Bramel, Bruce Posner, and David Simchi-Levi. A computerized approach to the new york cityschool bus routing problem. *IIE transactions*, 29(8):693–702, 1997.
- [5] Julien Bramel and David Simchi-Levi. A location based heuristic for general routing problems. *Operations research*, 43(4):649–660, 1995.
- [6] S.E. Carrell, T. Maghakian, and J.E. West. A’s from ZZZZ’s? The causal effect of school start time on academic achievement of adolescents. *American Economic Journal*, 3(3):62–71, 2011.
- [7] Xiaoli Chen, May A Beydoun, and Youfa Wang. Is sleep duration associated with childhood obesity? a systematic review and meta-analysis. *Obesity*, 16(2):265–274, 2008.
- [8] Xiaopan Chen, Yunfeng Kong, Lanxue Dang, Yane Hou, and Xinyue Ye. Exact and metaheuristic approaches for a bi-objective school bus scheduling problem. *PloS one*, 10(7):e0132600, 2015.
- [9] Geoffrey L Cohen, Julio Garcia, Nancy Apfel, and Allison Master. Reducing the racial achievement gap: A social-psychological intervention. *science*, 313(5791):1307–1310, 2006.
- [10] Stephanie J Crowley, Christine Acebo, and Mary A Carskadon. Sleep, circadian rhythms, and delayed phase in adolescence. *Sleep medicine*, 8(6):602–612, 2007.
- [11] Giuseppe Curcio, Michele Ferrara, and Luigi De Gennaro. Sleep loss, learning capacity and academic performance. *Sleep medicine reviews*, 10(5):323–337, 2006.
- [12] Fred Danner and Barbara Phillips. Adolescent sleep, school start times, and teen motor vehicle crashes. *J Clin Sleep Med.*, 4(6):533, 2008.
- [13] Arthur Delarue and Sebastien Martin. SchoolBusRouting repository. <https://github.com/adelarue/SchoolBusRouting>, 2018.
- [14] Jacques Desrosiers, Yvan Dumas, Marius M Solomon, and François Soumis. Time constrained routing and scheduling. *Handbooks in operations research and management science*, 8:35–139, 1995.
- [15] Iain Dunning, Joey Huchette, and Miles Lubin. Jump: A modeling language for mathematical optimization. *SIAM Review*, 59(2):295–320, 2017.
- [16] Finley Edwards. Early to rise? the effect of daily start times on academic performance. *Economics of Education Review*, 31(6):970–983, 2012.
- [17] Katia Fredriksen, Jean Rhodes, Ranjini Reddy, and Niobe Way. Sleepless in Chicago: tracking the effects of adolescent sleep loss during the middle school years. *Child development*, 75(1):84–95, 2004.
- [18] Armin Fügenschuh. Solving a school bus scheduling problem with integer programming. *European Journal of Operational Research*, 193(3):867–884, 2009.
- [19] Marco Hafner, Martin Stepanek, and Wendy M. Troxel. Later school start times in the U.S.: An economic analysis. Technical report, RAND Corporation, 2017.
- [20] Peter M Hahn, Bum-Jin Kim, Monique Guignard, J MacGregor Smith, and Yi-Rong Zhu. An algorithm for the generalized quadratic assignment problem. *Computational Optimization and Applications*, 40(3):351, 2008.
- [21] Susan Kohl Malone, Terra Ziporyn, and Alison M Battenheim. Applying behavioral insights to delay school start times. *Sleep Health: Journal of the National Sleep Foundation*, 3(6):483–485, 2017.
- [22] Dan O’Brien. MIT’s ”Quantum Team” wins first-ever BPS transportation challenge with revolutionary computer model. <https://www.bostonpublicschools.org/site/default.aspx?PageType=3&DomainID=4&ModuleInstanceID=14&ViewID=6446EE88-D30C-497E-9316-3F8874B3E108&FlexDataID=12431>, July 2017.

- [23] Judith Owens, Darrel Droblich, Allison Baylor, and Daniel Lewin. School start time change: An in-depth examination of school districts in the united states. *Mind, Brain, and Education*, 8(4):182–213, 2014.
- [24] Judith A Owens, Katherine Belon, and Patricia Moss. Impact of delaying school start time on adolescent sleep, mood, and behavior. *Archives of pediatrics & adolescent medicine*, 164(7):608–614, 2010.
- [25] Junhyuk Park and Byung-In Kim. The school bus routing problem: A review. *European Journal of operational research*, 202(2):311–319, 2010.
- [26] Junhyuk Park, Hyunchul Tae, and Byung-In Kim. A post-improvement procedure for the mixed load school bus routing problem. *European Journal of Operational Research*, 217(1):204–213, 2012.
- [27] Parag A Pathak and Peng Shi. Simulating alternative school choice options in Boston - Technical appendix. Technical report, MIT School Effectiveness and Inequality Initiative, 2013.
- [28] Patrick Schittekat, Marc Sevaux, and Kenneth Sorensen. A mathematical formulation for a school bus routing problem. In *Intl Conf on Serv Syst and Serv Mgmt*, volume 2, pages 1552–1557. IEEE, 2006.
- [29] Rebecca Shuster. 2018-19 school bell times equity impact. <https://www.bostonpublicschools.org/cms/lib/MA01906464/Centricity/Domain/2389/Equity%20Analysis.pdf>, December 2017.
- [30] Barbara M Smith and Anthony Wren. A bus crew scheduling system using a set covering formulation. *Transportation Research Part A: General*, 22(2):97–108, 1988.
- [31] Michela Spada, Michel Bierlaire, and Th M Liebling. Decision-aiding methodology for the school bus routing and scheduling problem. *Transportation Science*, 39(4):477–490, 2005.
- [32] Elizabeth A. Sullivan. Order on school start time realignment. <https://www.bostonpublicschools.org/cms/lib/MA01906464/Centricity/Domain/162/starttimesmotion.pdf>, December 2017.
- [33] Anne G. Wheaton, Gabrielle A. Ferro, and Janet B. Croft. School start times for middle school and high school students - united states, 2011–12 school year. Technical Report Vol. 64 No. 30, CDC, August 2015.
- [34] Greg Wiggan. Race, school achievement, and educational inequality: Toward a student-based inquiry perspective. *Review of Educational Research*, 77(3):310–333, 2007.



# From School Buses to Start Times: Driving Policy With Optimization

## Technical Appendix

Dimitris Bertsimas, Arthur Delarue and Sebastien Martin

October 2018\*

In this Supplementary Material, we present the details of the methods described in the Main Text. In particular, we introduce the mathematical notation and formalism needed to formulate and solve different optimization subproblems. We first describe the BiRD school bus routing algorithm in full detail, then we specify the setting of our computational experiments, in particular with respect to synthetic data. Subsequently, we present our mathematical formulation of the School Time Selection Problem (STSP), explain the details of our optimization algorithm and how it interfaces with school bus routing, before detailing our computational work, on both synthetic and real data.

## BiRD Routing Algorithm

We begin by giving a complete mathematical description of the BiRD (Bi-objective Routing Decomposition) algorithm for school bus routing. In the Main Text, we decompose the overall problem of school transportation into a single-school problem and multi-school problem. The single-school problem can be further decomposed into the two subproblems of stop assignment and single-school routing, while the multi-school problem can be further decomposed into the two subproblems of scenario selection and bus scheduling (the overall decomposition is detailed diagrammatically in Fig. 2). In this section, we detail the four subproblems of stop assignment, single-school routing, scenario selection and bus scheduling in order. Throughout the section, the (mixed-)integer

---

\*2nd round of revisions, Proceedings of the National Academy of Science (PNAS)

optimization problems that we formulate are solved using a mixed-integer optimization solver.

## Stop Assignment

Call  $\mathcal{S}$  the set of schools, and  $\mathcal{P}_s$  the set of pupils (students) attending each school  $S \in \mathcal{S}$ . In addition, call  $\mathcal{C}$  the set of all locations that can serve as bus stops. Each student  $p \in \mathcal{P}_S$  is associated with a set of allowed bus stops  $\mathcal{C}_p \subseteq \mathcal{C}$ . The walking distance from the home of student  $p$  to a stop  $c \in \mathcal{C}_p$  is denoted as  $d_{p,c}$ . This general setting reflects a variety of student-specific needs. For example, the allowed bus stops for younger students may be closer to their home or accessible without crossing major arteries. In addition, a pupil  $p$  with special needs may require a “door-to-door” pickup: in this case, the set  $\mathcal{C}_p$  is a singleton  $\{c\}$ , where  $d_{p,c} = 0$ .

We propose a simple integer optimization approach, where we seek to minimize the number of stops for each school and the total student walking distance. More precisely, we solve the following integer program for each school  $S$ .

$$\min \quad \sum_{c \in \mathcal{C}} z_c + \beta \sum_{p \in \mathcal{P}_S} \sum_{c \in \mathcal{C}_p} d_{p,c} y_{p,c} \quad (1a)$$

$$\text{s.t.} \quad y_{p,c} \leq z_c \quad \forall p \in \mathcal{P}_S, c \in \mathcal{C}_p \quad (1b)$$

$$\sum_{c \in \mathcal{C}_p} y_{p,c} = 1 \quad \forall p \in \mathcal{P}_S \quad (1c)$$

$$y_{p,c} \in \{0, 1\} \quad \forall p \in \mathcal{P}_S, c \in \mathcal{C}_p \quad (1d)$$

$$z_c \in \{0, 1\} \quad \forall c \in \mathcal{C} \quad (1e)$$

The binary variable  $z_c$  indicates whether stop  $c$  is selected for school  $S$ , and the binary variable  $y_{p,c}$  indicates whether student  $p$  is assigned to stop  $c$ . (1b) ensures that student  $p$  is assigned to stop  $c$  only if stop  $c$  is selected for school  $S$ , and (1c) certifies that each student is assigned to one stop. The first term in the objective corresponds to the total number of stops for school  $S$ , while the second term corresponds to the total walking distance for students attending school  $S$ . The parameter  $\beta$  controls the tradeoff between these two priorities. For large values of  $\beta$ , students will be assigned to the nearest stop to their home; as  $\beta$  tends to 0, students may walk further from their home in order to consolidate several stops (though never to an unacceptable stop that is not in  $\mathcal{C}_p$ ). We explore this tradeoff on synthetic data in Fig. 3a.

When applying this process in Boston, we added an additional constraint preventing the same

stop from having too many students at the same time. This correlates the stop assignment solutions of each school, requiring the concurrent optimization of stop assignments for all schools.

## Single-School Routing

Given that each student  $p$  has been assigned a bus stop  $c_p \in \mathcal{C}_p$ , we now turn to the problem of connecting stops into bus routes (or bus trips). Since several students may be assigned to the same stop, we can denote as  $\mathcal{C}_S \subseteq \mathcal{C}$  the set of bus stops with at least one student from school  $S$ , and call  $n_{c,S}$  the number of students from school  $S$  at a stop  $c \in \mathcal{C}_S$ .

We consider that the bus fleet is composed of several bus types, and denote the set of bus types as  $\mathcal{B}$ . All buses of a given type  $b \in \mathcal{B}$  are considered identical, with a fixed number of seats (capacity)  $Q_b$ , and a fixed number of wheelchair spots  $W_b$ . Let  $\mathcal{Y}$  designate the set of bus depots (or bus yards) where buses are stored during the night and the middle of the day.

We let  $t_{c,S}^{\text{pickup}}$  designate the length of time needed to pick up every student for school  $S$  at stop  $c$ , and  $t_S^{\text{drop-off}}$  the length of time needed to drop off every student on a bus at school  $S$ . Note that the former is a function of the number of students at the stop  $n_{c,S}$ , while the latter is independent of the number of students on the bus. For any two locations  $\ell_1, \ell_2 \in \mathcal{S} \cup \mathcal{C} \cup \mathcal{Y}$ , and given a particular time of day  $\tau$ , let  $t_{\ell_1, \ell_2, \tau}^{\text{drive}}$  designate the driving time from location  $\ell_1$  to location  $\ell_2$ , when departing location  $\ell_1$  at time  $\tau$ . Finally, we consider that all buses serving school  $S$  must arrive at school at time  $\tau_S$  to drop off their students, and are therefore free to leave the school at time  $\tau_S + t_S^{\text{drop-off}}$ . We also consider that students cannot spend more than a fixed duration  $T^{\text{max}}$  on the bus.

Throughout this paper, we consider that the travel times  $t_{\ell_1, \ell_2, \tau}^{\text{drive}}$  are deterministic and known. We note that this modeling choice makes it more difficult to account for unforeseen traffic events such as accidents. These travel times can be obtained through a commercial service like the Google Maps API, or estimated from data. In practice, school bus routes are typically constructed under a static, deterministic travel time model, and stakeholders understand that this results in buses sometimes arriving late, especially in poor weather or traffic conditions. As a simple way to mitigate the impact of traffic in our work with BPS, we artificially increased the drop-off times  $t_S^{\text{drop-off}}$  at each school, e.g. requiring buses to arrive at school 10 or 15 minutes before the beginning of school even though physically unloading students may only require 3 minutes. This “buffer” both reduces the chance that students will be late to class, and makes it less likely that delays will be compounded onto the bus’s next trip. However, no matter what travel time estimates are used, it is practically impossible for buses to always be on time.

For school  $S$ , there are many methods from the vehicle routing literature to find bus trips to serve all the stops in  $\mathcal{C}_S$ . A trip (or route) is simply an ordered sequence of stops visited, or served, by a bus. We assume that a bus serving a trip will pick up every student at each stop in the trip. A trip is considered feasible if (a) it verifies that no student spends more than a time  $T^{\max}$  between pickup and drop-off, and (b) there exists a type of vehicle in the fleet with enough capacity to transport all students assigned to the stops served by the trip. Given a feasible trip  $R$ , we let  $\mathcal{B}_R \subseteq \mathcal{B}$  designate the set of types of buses that have the necessary capacity to serve the trip, and we denote by  $T_R$  the service time of the trip, i.e. the time between arrival at the first stop and arrival at the destination school.

We use a randomized greedy heuristic (Algorithm 1) to generate a set of feasible trips  $\mathcal{T}_S$ , making sure that each stop  $c \in \mathcal{C}_S$  is served by a nonempty set of feasible trips  $\mathcal{T}_c \subset \mathcal{T}_S$ . More precisely, the heuristic returns a set of trips  $\mathcal{T}$  that covers each stop exactly once, and we run it  $N$  times to build a set of feasible trips where each stop is covered by several trips. The heuristic returns  $N$  different solutions and each stop will be covered by  $N$  different feasible trips.

We index the trips in  $\mathcal{T}_S$  by  $1, \dots, m_S$  and for a given set of trips  $\mathcal{T} \subseteq \mathcal{T}_S$  we let  $I(\mathcal{T})$  designate the subset of  $\{1, \dots, m_S\}$  corresponding to trips in  $\mathcal{T}$ . We can then find the best set of routes by solving the following minimum cover problem.

$$\min \quad \lambda \sum_{i=1}^{m_S} r_i + \sum_{i=1}^{m_S} r_i \Theta_i \quad (2a)$$

$$\text{s.t.} \quad \sum_{i \in I(\mathcal{T}_c)} r_i \geq 1 \quad \forall c \in \mathcal{C}_S \quad (2b)$$

$$r_i \in \{0, 1\} \quad \forall i \in \{1, \dots, m_S\} \quad (2c)$$

The binary variable  $r_i$  indicates whether trip  $i$  is selected. Constraint (2b) imposes that every stop must be served by at least one trip. For each trip  $i$  in  $\mathcal{T}_S$ , we have  $\Theta_i = \sum_{p: c_p \in C_i} \theta_p^{(i)}$ , where  $C_i$  represents the list of stops served by the  $i$ -th trip  $i$ , and  $\theta_p^{(i)}$  corresponds to the time spent by student  $p$  on the bus during trip  $i$ . Thus, the parameter  $\lambda$  controls the importance of the number of trips relative to the total time students spend on the bus.

In the optimal solution of the problem above, some stops may be served by more than one trip. We consider the set of trips  $\mathcal{T}_c^{\text{opt}} \subset \mathcal{T}_S$  serving each such stop  $c$ . For each  $i \in I(\mathcal{T}_c^{\text{opt}})$ , we compute the improvement  $\delta_i$  to the objective of (2) that would be obtained by removing stop  $c$  from trip  $i$ . We let  $c$  remain in trip  $j$ , such that  $j = \arg \min_{i \in I(\mathcal{T}_c^{\text{opt}})} \delta_i$ , and remove it from all other trips in

---

**Algorithm 1** Randomized greedy insertion heuristic, assuming only one type of bus with capacity  $Q$ . Input: a school  $S$  and a time of day  $\tau$

---

```

1: function GREEDYRANDOMIZED( $S, \tau$ )
2:    $\mathcal{T} \leftarrow \emptyset$  ▷ Initialize empty set of trips
3:   while  $\mathcal{C}_S \neq \emptyset$  do
4:      $c$  randomly selected from  $\mathcal{C}_S$ 
5:      $\mathcal{C}_S \leftarrow \mathcal{C}_S \setminus \{c\}$ 
6:      $R \leftarrow [c]$  ▷ Trip initialized with selected stop  $c$ 
7:      $N^{\text{students}} \leftarrow n_{c,S}$  ▷ Number of students currently on trip  $R$ 
8:     while  $\mathcal{C}_S \neq \emptyset$  do
9:        $T^{\text{new}}, c^{\text{new}}, i^{\text{new}} \leftarrow \text{BESTINSERTION}(R, \mathcal{C}_S, \tau)$ 
10:      if  $T^{\text{new}} \leq T^{\text{max}}$  and  $N^{\text{students}} + n_{c^{\text{new}},S} \leq Q$  then
11:         $R \leftarrow \text{INSERT}(R, c^{\text{new}}, i^{\text{new}})$ 
12:         $\mathcal{C}_S \leftarrow \mathcal{C}_S \setminus \{c^{\text{new}}\}$ 
13:      else
14:        break
15:       $\mathcal{T} \leftarrow \mathcal{T} \cup \{R\}$ 
16:   return  $\mathcal{T}$ 
17: function BESTINSERTION( $R, \mathcal{C}_S, \tau$ )
18:   Let  $R := [c_1, c_2, \dots, c_n]$  ▷ Name the  $n$  stops in  $R$  for clarity of notation
19:    $T^{\text{best}} \leftarrow \infty$ ;  $c^{\text{best}} \leftarrow \infty$ ;  $i^{\text{best}} \leftarrow \infty$  ▷ Initialize total trip time, stop to insert, insertion slot in trip
20:   for  $c \in \mathcal{C}_S$  do
21:     for  $i \leftarrow 0$  to  $n$  do
22:        $T \leftarrow t_{c,S}^{\text{pickup}}$  ▷ Include pickup time first
23:       if  $i = 0$  then
24:          $T \leftarrow T + t_{c,c_1,\tau}^{\text{drive}} + \sum_{j=1}^{n-1} t_{c_j,c_{j+1},\tau}^{\text{drive}} + t_{c_n,S,\tau}^{\text{drive}}$  ▷  $c$  inserted before  $c_1$ 
25:       else if  $i = n$  then
26:          $T \leftarrow T + \sum_{j=1}^{n-1} t_{c_j,c_{j+1},\tau}^{\text{drive}} + t_{c_n,c,\tau}^{\text{drive}} + t_{c,S,\tau}^{\text{drive}}$  ▷  $c$  inserted after  $c_n$ 
27:       else
28:          $T \leftarrow T + \sum_{j=1}^i t_{c_j,c_{j+1},\tau}^{\text{drive}} + t_{c_i,c,\tau}^{\text{drive}} + t_{c,c_{i+1},\tau}^{\text{drive}} + \sum_{j=i+1}^{n-1} t_{c_j,c_{j+1},\tau}^{\text{drive}} + t_{c_n,S,\tau}^{\text{drive}}$  ▷  $c$  inserted
29:         between  $c_i$  and  $c_{i+1}$ 
30:       if  $T < T^{\text{best}}$  then
31:          $T^{\text{best}} \leftarrow T$ ;  $c^{\text{best}} \leftarrow c$ ;  $i^{\text{best}} \leftarrow i$ 
32:   return  $T^{\text{best}}, c^{\text{best}}, i^{\text{best}}$ 
33: function INSERT( $R, c, i$ )
34:   Let  $R := [c_1, c_2, \dots, c_n]$ 
35:   if  $i = 0$  then
36:      $R \leftarrow [c, c_1, \dots, c_n]$ 
37:   else if  $i = n$  then
38:      $R \leftarrow [c_1, \dots, c_n, c]$ 
39:   else
40:      $R \leftarrow [c_1, \dots, c_i, c, c_{i+1}, \dots, c_n]$ 
41:   return  $R$ 

```

---

$\mathcal{T}_c^{\text{opt}}$ .

To gain tractability at the expense of full optimality, it is easy to split up this trip selection phase into  $K$  phases, where the first phase selects the best trips among the first  $N/K$  greedy solutions, then the second phase selects the best trips among the trips from the next  $N/K$  routing solutions and the optimal trips from the first phase, etc.

The output of this algorithm is a set of trips  $\mathcal{T}_S^*$  that covers every stop in  $\mathcal{C}_S$  exactly once. We can perform this iterative optimization algorithm several times for different values of the tradeoff parameter  $\lambda$  to obtain an array of varied routing scenarios for each school. Low values of  $\lambda$  will lead to scenarios with more buses but shorter trips, while high values of  $\lambda$  will produce scenarios with longer trips but fewer buses.

For each school  $S \in \mathcal{S}$ , we therefore end up with a set of routing scenarios  $\mathcal{R}_S = \{\mathcal{T}_S^h\}_{h=1}^{h_S}$ , where each scenario is a complete set of trips  $\mathcal{T}_S^h$  that covers every stop in  $\mathcal{C}_S$  exactly once. Each scenario is located in a different region of the Pareto front between two objectives, the number of buses and the average student travel time, hence the name of Bi-objective Routing Decomposition (BiRD).

Because our overall school bus routing approach is modular, the methods we propose to solve the single-school routing problem can easily be replaced. The only requirement on such substitute approaches is that they be able to produce several different solutions for each school, trading off between the average time students spend on the bus and the number of bus trips.

## Scenario Selection

We develop a bus scheduling algorithm that bridges the gap with the bus routing problem. The algorithm takes as input a set of scenarios  $\mathcal{R}_S$  of size  $h_S$  for each school  $S$ . Because what is optimal for one school may not be optimal for the entire system, our goal is to jointly select one scenario for each school in a way that minimizes the desired objective (e.g. number of buses) across the whole district.

In order to select a single-school solution for each school in a way that minimizes the total number of buses, we formulate an integer network flow problem on a graph where nodes represent bus trips, which are served when they are traversed by a unit of flow. To simplify notation, we assume at first there exists only one type of bus; we then create the following “scenario selection graph”  $(\mathcal{N}, \mathcal{A})$ . The set of nodes  $\mathcal{N}$  consists of (i) a single depot node  $y$ , (ii) a “trip node”  $\rho_{S,h,R}$  for each school  $S$ , each scenario index  $1 \leq h \leq h_S$ , and each trip  $R \in \mathcal{T}_S^h$ , and (iii) an “availability

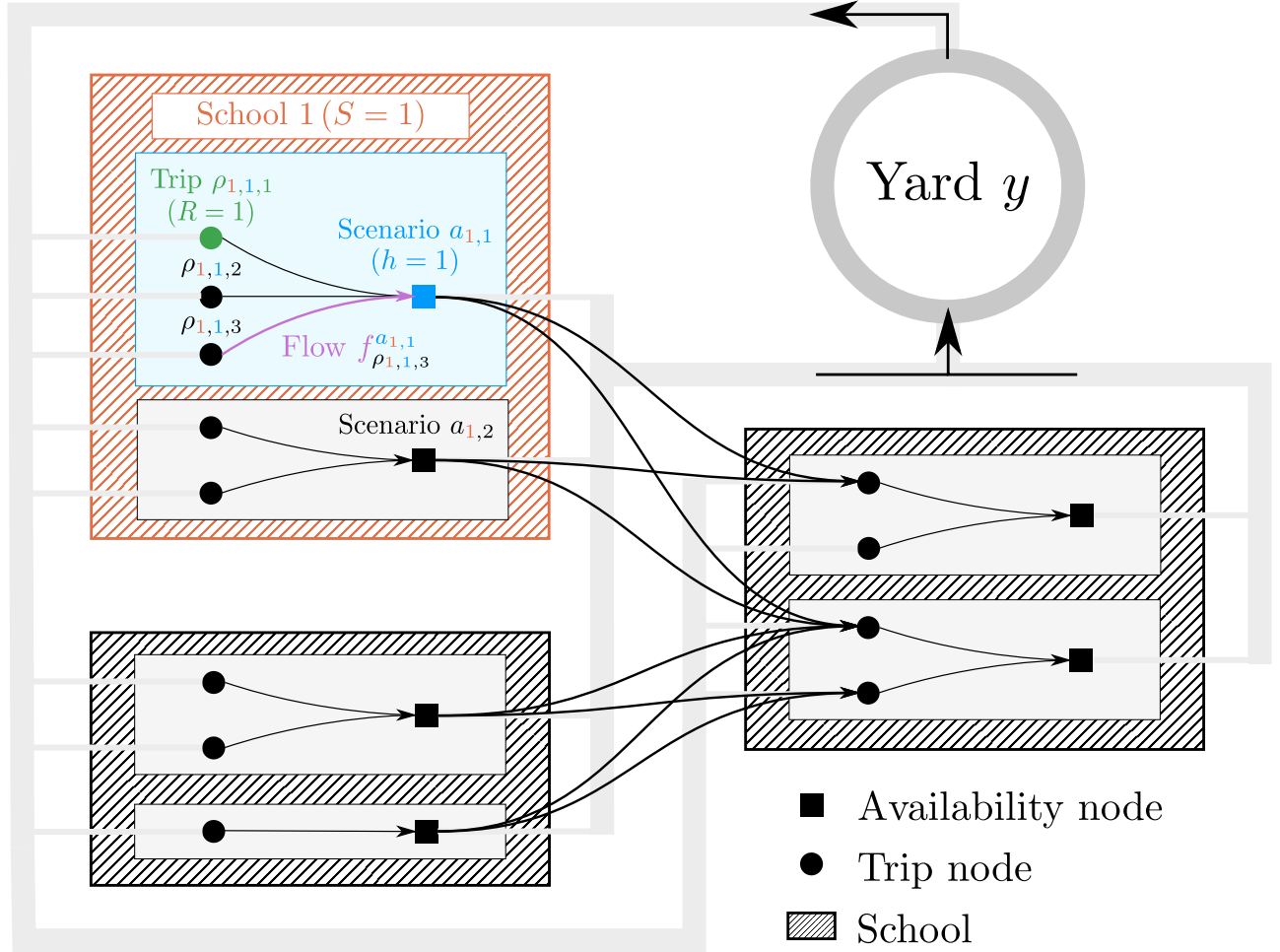


Figure 1: Diagram of the scenario selection graph for a small example of a school bus routing problem with three schools. Each school is represented by a diagonally striped rectangle, with two associated routing scenarios, represented by lightly shaded rectangles within the larger rectangle of the school. Note that for the two schools on the left, the two scenarios do not have the same overall number of trips (e.g. 3 trips vs. 2 trips for School 1), while this is not the case for the school on the right (both scenarios consist of two trips). Black arrows represent edges from trip nodes to availability nodes (within each school), and from availability nodes to trip nodes (modeling bus reuse between schools). Light gray arrows represent edges between trip/availability nodes and the yard/depot node.

node”  $a_{S,h}$  for each school  $S$  and for each routing scenario  $1 \leq h \leq h_S$ . The set of arcs  $\mathcal{A}$  includes an arc from  $y$  to each trip node  $\rho_{S,h,R}$ , from each trip node  $\rho_{S,h,R}$  to the corresponding availability node  $a_{S,h}$ , and from each availability node  $a_{S,h}$  back to the depot  $y$ . We also include an arc from each availability node  $a_{S,h}$  to each trip node  $\rho_{S',h',R'}$  where trip  $R' \in \mathcal{T}_{S'}^{h'}$  is time-compatible with a bus starting from school  $S$ . By time-compatible we mean that there is enough time for the bus to drive from school  $S$  to the first stop  $c_{R'}^{\text{start}}$  of trip  $R'$  and then make it to school  $S'$  on time, which can be expressed as  $\tau_S + t_S^{\text{drop-off}} + t_{S,c_{R'}^{\text{start}},\tau_S}^{\text{drive}} + T_R \leq \tau_{S'}$ . For a node  $i \in \mathcal{N}$ , let  $\mathcal{I}(i) \subseteq \mathcal{N}$  be the in-neighborhood of node  $i$ , and  $\mathcal{O}(i) \subseteq \mathcal{N}$  designate the out-neighborhood of  $i$ . We display a diagram of the scenario selection graph for a small example with three schools in Fig. 1.

Given the graph described above, we consider that a unit of flow traversing a series of trip nodes corresponds to a bus serving the corresponding trips in order. Therefore, minimizing the total number of buses corresponds to minimizing the total flow out of the yard node  $y$  subject to the constraints that (a) flows along all arcs must be integral and (b) given that a particular single-school routing solution  $\mathcal{T}_S^h$  is selected for school  $S$ , every trip  $R \in \mathcal{T}_S^h$  must be served by exactly one bus (i.e. each node  $\rho_{S,h,R}$  is traversed by exactly one unit of flow). We therefore formulate the following network flow problem with integer flow variables  $f_i^j$  for each arc  $(i, j)$ .

$$\min \sum_{S \in \mathcal{S}} \sum_{h=1}^{h_S} \sum_{R \in \mathcal{T}_S^h} f_y^{\rho_{S,h,R}} \quad (3a)$$

$$\text{s.t. } f_{\rho_{S,h,R}}^{a_{S,h}} = z_{S,h} \quad S \in \mathcal{S}, 1 \leq h \leq h_S, R \in \mathcal{T}_S^h \quad (3b)$$

$$\sum_{j \in \mathcal{I}(i)} f_j^i = \sum_{j \in \mathcal{O}(i)} f_i^j \quad \forall i \in \mathcal{N} \quad (3c)$$

$$\sum_{h=1}^{h_S} z_{S,h} = 1 \quad \forall S \in \mathcal{S} \quad (3d)$$

$$z_{S,h} \in \{0, 1\} \quad \forall S \in \mathcal{S}, 1 \leq h \leq h_S \quad (3e)$$

$$f_i^j \in \mathbb{Z}_+ \quad \forall (i, j) \in \mathcal{A}. \quad (3f)$$

The binary variable  $z_{S,h}$  is 1 if  $\mathcal{T}_S^h$  is the selected set of trips for school  $S$ , and constraint (3d) ensures that exactly one set of trips is selected for each school. Constraint (3c) ensures conservation of flow  $f_i^j$  at each node  $i$ , with the following interpretation at each node. At the depot node  $y$ , it means that buses leaving the depot must eventually come back. At a trip node  $\rho_{S,h,R}$ , it means that a bus serving trip  $R$  must then become available at school  $S$  at time  $\tau_S + t_S^{\text{drop-off}}$ . At an



availability node  $a_{S,h}$ , it means that a bus that is available at a school after serving a trip must either return to the yard or serve another trip. Constraint (3b) guarantees that a particular set of trips is selected if and only if each trip is assigned to exactly one bus.

The formulation above has a large number of integer variables, on the order of 2 million for a problem with 200 schools and 5 scenarios per school. However, commercial solvers such as Gurobi can solve it to optimality in less than two hours. Intuitively, the network flow formulation is quite strong, allowing the relaxation-based techniques and other heuristics implemented in modern MIO solvers to tackle it successfully.

The formulation above can easily be modified if there is more than one type of bus available ( $|\mathcal{B}| > 1$ ). All we need to do is create a new graph such that each trip node  $\rho_{S,h,R}$  maps to the set of nodes  $\{\rho_{S,h,R,b}\}_{b \in \mathcal{B}_R}$  in the new graph (one for each bus type), and similarly map the yard node  $y$  and the availability nodes for each school  $a_{S,h}$  to sets of nodes  $\{y_b\}_{b \in \mathcal{B}}$  and  $\{a_{S,h,b}\}_{b \in \mathcal{B}}$ . For every arc  $(i, j)$  in the original arc set  $\mathcal{A}$ , we create a new arc  $(i_b, j_b)$  in the new graph, and modify constraint (3b) such that only one of the trip nodes  $\{\rho_{S,h,R,b}\}_{b \in \mathcal{B}_R}$  can be traversed, effectively selecting the bus type that will serve trip  $R$ .

## Bus Scheduling

The last step of our routing methodology involves determining final bus schedules given exactly one routing scenario for each school. We use a similar approach to the one described above for scenario selection. We define a “bus selection graph”  $(\bar{\mathcal{N}}, \bar{\mathcal{A}})$ , where the set of nodes  $\bar{\mathcal{N}}$  consists of (i) a node  $y_{b,\ell}$  for each bus type  $b \in \mathcal{B}$  and each physical bus depot location  $\ell \in \mathcal{L}$ , (ii) a trip node  $\rho_{S,R,b,\ell}^{\text{AM}}$  (respectively  $\rho_{S,R,b,\ell}^{\text{PM}}$ ) for each school  $S$ , each morning trip  $R$  in the selected scenario  $\mathcal{T}_S^{\text{AM}}$  (respectively each afternoon trip  $R \in \mathcal{T}_S^{\text{PM}}$ ), each bus type  $b \in \mathcal{B}_R$ , and each depot location  $\ell \in \mathcal{L}$ .

The set of arcs  $\bar{\mathcal{A}}$  includes (i) an arc to and from the depot node  $y_{b,\ell}$  for each trip node  $\rho_{S,R,b,\ell}^{\text{AM}}$  and  $\rho_{S,R,b,\ell}^{\text{PM}}$ , (ii) an arc from each trip node  $\rho_{S,R,b,\ell}^{\text{AM}}$  (resp.  $\rho_{S,R,b,\ell}^{\text{PM}}$ ) to each trip node  $\rho_{S',R',b,\ell}^{\text{AM}}$  (resp.  $\rho_{S',R',b,\ell}^{\text{PM}}$ ) where trip  $R'$  for school  $S'$  is time-compatible with a bus starting from school  $S$ . For a node  $i \in \bar{\mathcal{N}}$ , let  $\bar{\mathcal{I}}(i) \subseteq \bar{\mathcal{N}}$  designate the in-neighborhood of node  $i$ , and  $\bar{\mathcal{O}}(i) \subseteq \bar{\mathcal{N}}$  designate the out-neighborhood of  $i$ . For depot nodes  $y_{b,\ell}$ , define  $\bar{\mathcal{O}}^{\text{AM}}(y_{b,\ell}) = \bar{\mathcal{O}}(y_{b,\ell}) \cap \{\rho_{S,R,b,\ell}^{\text{AM}} \in \bar{\mathcal{N}} : S \in \mathcal{S}, R \in \mathcal{T}_S^{\text{AM}}\}$ , and define  $\bar{\mathcal{O}}^{\text{PM}}(y_{b,\ell})$  similarly. We assign a cost  $c_{i,j}$  to each arc  $(i, j) \in \bar{\mathcal{A}}$  to represent the objective we are trying to optimize (fuel consumption, driving distance, etc.), and introduce auxiliary variables to minimize the number of buses. Solving the bus scheduling problem corresponds to solving the following mixed-integer optimization problem.

$$\min \quad \gamma \sum_{b \in \mathcal{B}} \sum_{\ell \in \mathcal{L}} c_b K_{b,\ell} + (1 - \gamma) \sum_{(i,j) \in \bar{\mathcal{A}}} c_{i,j} f_i^j \quad (4a)$$

$$\sum_{j \in \bar{\mathcal{I}}(i)} f_j^i = \sum_{j \in \bar{\mathcal{O}}(i)} f_i^j \quad \forall i \in \bar{\mathcal{N}} \quad (4b)$$

$$\sum_{b \in \mathcal{B}} \sum_{\ell \in \mathcal{L}} \sum_{i \in \bar{\mathcal{I}}(\rho_{S,R,b,\ell}^q)} f_i^{\rho_{S,R,b,\ell}^q} = 1 \quad \forall S \in \mathcal{S}, \forall R \in \mathcal{T}_S^q, q \in \{\text{AM}, \text{PM}\} \quad (4c)$$

$$\sum_{i \in \bar{\mathcal{O}}^q(y_{b,\ell})} f_{y_{b,\ell}}^i \leq K_{b,\ell} \quad \forall b \in \mathcal{B}, \ell \in \mathcal{L}, q \in \{\text{AM}, \text{PM}\} \quad (4d)$$

$$K_{b,\ell} \in \mathbb{Z}_+ \quad \forall b \in \mathcal{B}, \forall \ell \in \mathcal{L} \quad (4e)$$

$$f_i^j \in \{0, 1\} \quad \forall (i, j) \in \bar{\mathcal{A}}. \quad (4f)$$

The flow variables indicate  $f_i^j$  select the bus type and origin depot for the bus serving each trip, as well as the sequence of trips served by each bus. Meanwhile, the variables  $K_{b,\ell}$  represent the number of buses of type  $b$  coming from bus yard location  $\ell$ , and they are related to the variables  $f_i^j$  by constraint (4d). Constraint (4b) enforces flow conservation, which simply means that buses must leave the yard, serve at least one trip, and then return to the yard. Constraint (4c) enforces that each trip is served by exactly one bus, from one particular depot location and of one particular type. In the objective,  $c_b > 0$  represents the cost of a bus of type  $b$ ,  $c_{i,j}$  designates the cost of a particular edge in the graph (driving distance, fuel consumption, etc.), and  $\gamma$  controls the relative importance of the two parts of the objective (number of buses and driving distance/fuel consumption). Note that the combination of constraint (4d) and the positive cost associated with  $K_{b,\ell}$  in the objective ensure that  $K_{b,\ell}$  is the maximum of the number of morning buses and the number of afternoon buses (of type  $b$ , from location  $\ell$ ).

Readers will note that we only solve the morning and afternoon problems jointly in the last step (bus scheduling) of the BiRD algorithm. In principle, there is no reason not to solve morning and afternoon together at the scenario selection step as well. However, we found that treating morning and afternoon separately at the scenario selection step did not significantly affect the final solution while meaningfully improving tractability.

# Routing Experiments

We now describe the setting for our computational experiments, and present a more thorough overview of the results from the Main Text. We evaluate BiRD using our own synthetic examples, as well as sample problems from the literature. We first illustrate the way we generate synthetic experiments and the insights that they can provide. We then detail the process used to compare BiRD to existing approaches from the literature, using synthetic problems from Park, Tae and Kim [4].

## Synthetic Experiments and Results

In order to build intuition about the BiRD algorithmic framework, we first study its performance on our own synthetically generated examples. We consider a school district as a square (30km by 30km) in the 2D plane, in which we sample  $|\mathcal{S}|$  school locations at random. We fix the total number of students to  $|\mathcal{P}|$ , among which we sample  $|\mathcal{S}| - 1$  students  $p_{i_1}, \dots, p_{i_{|\mathcal{S}|-1}}$  uniformly at random without replacement, and enforce WLOG that  $0 = i_0 < i_1 < \dots < i_{|\mathcal{S}|-1} < i_{|\mathcal{S}|} = |\mathcal{P}|$ . Then we assign all students  $p_i$  such that  $i_{k-1} < i \leq i_k$  to school  $k$ . For any point  $x$  and any positive real number  $\rho$ , let  $\mathcal{U}_\rho(x)$  designate the uniform distribution over the disk of radius  $\rho$  centered at  $x$ .

For each school with location  $x_S$ , we select a radius  $R_S \sim \mathcal{U}(\underline{R}_S, \overline{R}_S)$ . The first student for that school is assigned a location sampled from  $\mathcal{U}_R(x_S)$ , and every subsequent student location is sampled with probability  $q$  from  $\mathcal{U}_r(x_p)$ , where  $x_p$  designates the location of the previous student, and with probability  $1 - q$  from  $\mathcal{U}_R(x_S)$ . This procedure creates small clusters of students which can be thought of as small “neighborhoods”. Each school is randomly assigned a start time of 7:30, 8:30, or 9:30, and a day length of 7, 8, or 9 hours. Finally, for each student with location  $x_p$ , we create a bus stop with location sampled from  $\mathcal{U}_{r_s}(x_p)$ . A large example of a synthetically generated district is shown in Fig. 2.

We begin with the stop assignment problem, exploring the tradeoff between the average student walking distance and the number of stops per school. We generate 100 synthetic school district instances with 100 schools and 10,000 students, and vary the stop assignment tradeoff parameter  $\beta$ . The results can be seen in Fig. 3a, showing that it is possible to reduce the average student walking distance by more than 25% without adding more than one or two stops per school.

Next, we generate 100 synthetic instances with 50 schools and 5,000 students to study the effect of the number of scenarios. For each school, we generate ten scenarios with different values

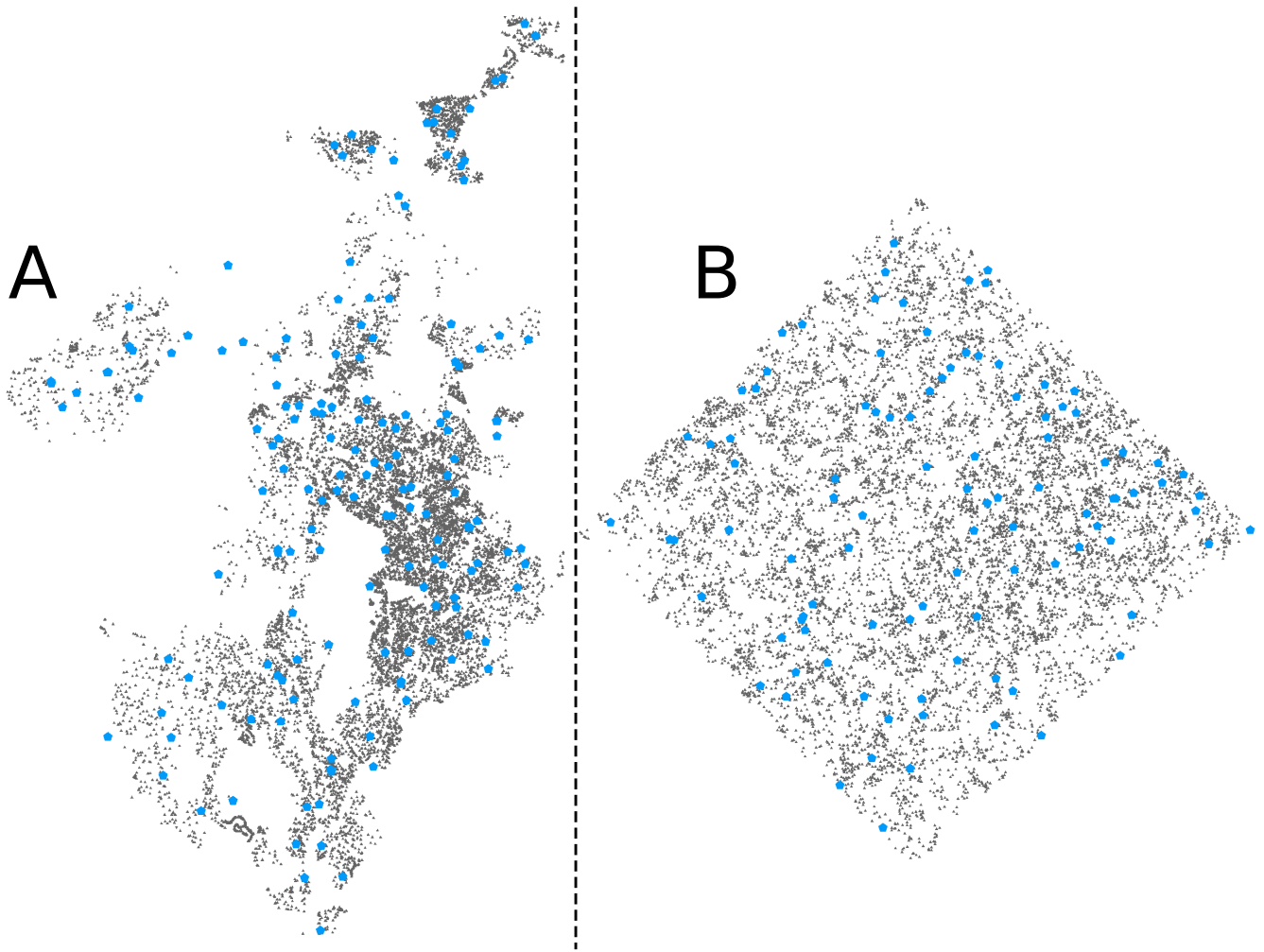
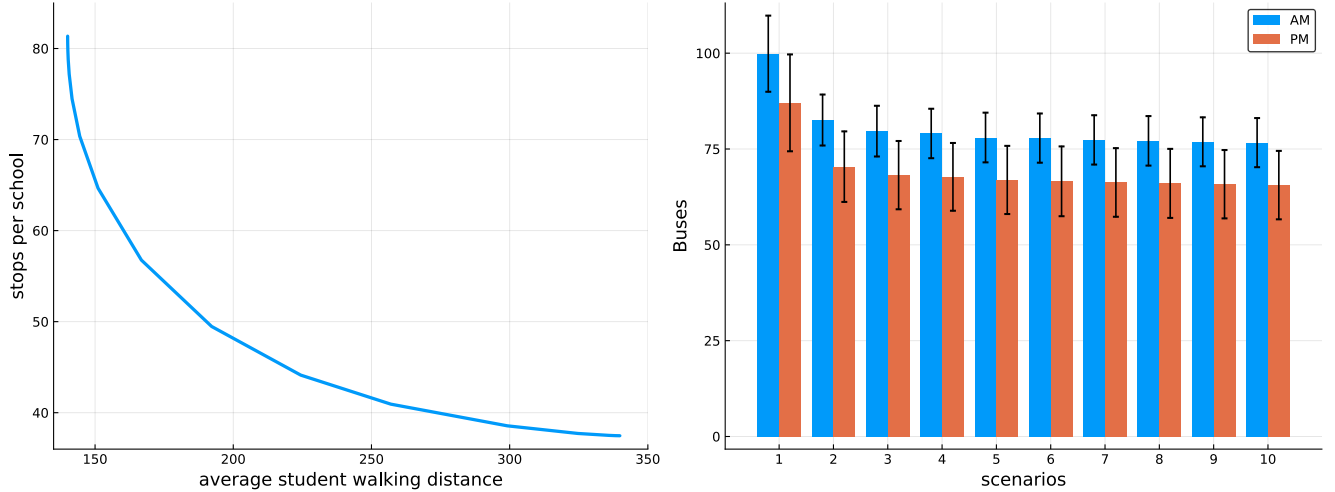


Figure 2: Geographic visualizations of two school districts. Small gray triangles represent students, while larger blue pentagons represent schools. (A) Boston Public Schools, 2017-18 school year (anonymized). (B) Synthetic school district with 100 schools, generated as described in the SI section on Synthetic Experiments.



(a) Tradeoff between average number of stops per school and average student-to-stop walking distance. Experiment on synthetic school district with 100 schools and 10,000 students. As the average walking distance increases, the number of stops decreases. The convexity of the tradeoff curve suggests diminishing returns in increasing the average walking distance.

(b) Effect of the number of scenarios on the total number of buses. As we increase the number of scenarios for each school, we create more room for optimization: the algorithm can select shorter routes for some schools, and longer routes for other schools, in a way that maximizes bus re-use. Overall, using several scenarios can yield as much as a 25% improvement in the number of buses. Using just two scenarios for school already improves the objective by 20%. Results are averaged over 100 random synthetic districts, with error bars corresponding to the standard deviation of the number of buses.

Figure 3: Analysis of performance of algorithm components on synthetic data.

of the single-school tradeoff parameter  $\lambda$ . For each scenario, we combine randomized routes such that each stop is covered by 400 routes, in 20 phases of 20 routes each. Then we solve the scenario-selection problem using  $\rho$  scenarios, for  $\rho = 1, \dots, 10$ . We use cross-validation to select the specific subset of size  $\rho$  among the considered values of  $\lambda$ . The results, shown in Fig. 3b support the intuitive statement that “what is optimal for one school may not be optimal for the entire system”, as choosing two routing options for each school yields a 20% improvement over choosing just one. In addition, the number of scenarios quickly yields diminishing returns, which is useful because it enables us to solve the scenario selection problem with two or three scenarios per school (increasing tractability) and obtain a solution that is almost as good as one computed with many more scenarios per school.

## Comparison with Existing Methods

We now compare the performance of the BiRD algorithm to methods from Braca et al. [1], Park, Tae and Kim [4], and Chen et al. [2] on benchmark synthetic data sets from Park, Tae and Kim [4].

As we mentioned before, the school bus routing problem has a large number of variants, including mixed loads and drop-off time windows. The benchmark data sets have a time window (between 10 and 30 minutes depending on the school) associated with each school corresponding to possible drop-off times, and buses serving the same school can thus arrive at different times. BiRD assumes that all buses for a given school must arrive at the same time, so we are still guaranteed to provide a feasible solution to the benchmark problems, but we have fewer degrees of freedom than some other methods. In order to quickly select a bell time within the time windows, we solve a simple optimization problem that just tries to choose the times that are the most spread. We will introduce in the next section a much more powerful bell time optimization algorithm, but it was not needed here as we want to show the performance of BiRD on its own.

Having computed a single start time for each school, we are now ready to solve the school bus routing problem. We consider two categories of benchmarks, RSRB and CSCB, which are generated slightly differently (see the original paper for details). Each one assumes that stop assignment has been performed as a preprocessing step, and can be solved assuming the maximum ride time  $T^{\max}$  is 45 minutes (2700 s) or 90 minutes (5400 s). The RSRB benchmarks have eight different problems of varying size, while the CSCB benchmarks have sixteen different problems of varying size. For each one, we use BiRD to minimize the number of buses needed to route all

Instance	Max. ride time (s)	$N_{\text{sch}}$	$N_{\text{stops}}$	$N_{\text{stud}}$	Number of buses					Improvement (over best)
					$Z^*$	$Z_{\text{Chen}}$	$Z_{\text{PML}}$	$Z_{\text{Park}}$	$Z_{\text{Braca}}$	
RSRB01	2700	6	250	3409	<b>*22</b>	26	32	35	36	15%
RSRB02	2700	12	250	3671	<b>24</b>	27	29	32	34	11%
RSRB03	2700	12	500	6844	<b>46</b>	48	61	66	77	4%
RSRB04	2700	25	500	6867	<b>43</b>	58	65	68	78	26%
RSRB05	2700	25	1000	13,765	<b>83</b>	94	106	124	118	12%
RSRB06	2700	50	1000	12,201	<b>69</b>	87	93	103	106	21%
RSRB07	2700	50	2000	26,934	<b>125</b>	149	171	185	201	16%
RSRB08	2700	100	2000	32,048	<b>133</b>	139	168	178	212	4%
CSCB01	2700	6	250	3932	<b>25</b>	32	33	38	39	22%
CSCB02	2700	12	250	3208	<b>25</b>	26	29	33	38	4%
CSCB03	2700	12	500	6861	<b>42</b>	58	55	66	69	24%
CSCB04	2700	25	500	7605	<b>44</b>	62	64	72	84	29%
CSCB05	2700	25	1000	17,161	<b>87</b>	114	125	135	157	24%
CSCB06	2700	50	1000	18,232	<b>95</b>	120	117	138	138	19%
CSCB07	2700	50	2000	27,755	<b>130</b>	173	176	203	229	25%
CSCB08	2700	100	2000	28,005	<b>136</b>	154	176	181	226	12%
CSCB09	2700	6	250	4159	<b>22</b>	-	26	32	36	15%
CSCB10	2700	12	250	4221	<b>24</b>	-	32	33	39	25%
CSCB11	2700	12	500	6006	<b>40</b>	-	55	58	68	27%
CSCB12	2700	25	500	6098	<b>51</b>	-	65	77	79	22%
CSCB13	2700	25	1000	14,242	<b>101</b>	-	120	164	158	16%
CSCB14	2700	50	1000	16,318	<b>93</b>	-	111	137	133	16%
CSCB15	2700	50	2000	27,482	<b>172</b>	-	192	243	253	10%
CSCB16	2700	100	2000	24,051	<b>132</b>	-	191	208	213	31%
Average	2700				<b>73.5</b>	85.4	95.5	108.7	117.5	18%
RSRB01	5400	6	250	3409	<b>*22</b>	27	28	31	31	19%
RSRB02	5400	12	250	3671	<b>*21</b>	23	25	30	29	9%
RSRB03	5400	12	500	6844	<b>45</b>	46	51	61	61	2%
RSRB04	5400	25	500	6867	<b>41</b>	<b>41</b>	52	56	60	0%
RSRB05	5400	25	1000	13,765	<b>83</b>	<b>80</b>	86	106	86	-4%
RSRB06	5400	50	1000	12,201	<b>62</b>	72	75	82	81	14%
RSRB07	5400	50	2000	26,934	<b>124</b>	141	147	159	163	12%
RSRB08	5400	100	2000	32,048	<b>126</b>	140	149	158	186	10%
CSCB01	5400	6	250	3932	<b>24</b>	29	26	35	33	8%
CSCB02	5400	12	250	3208	<b>22</b>	<b>22</b>	24	27	26	0%
CSCB03	5400	12	500	6861	<b>36</b>	42	46	52	50	14%
CSCB04	5400	25	500	7605	<b>38</b>	45	48	57	67	16%
CSCB05	5400	25	1000	17,161	<b>85</b>	100	99	115	134	14%
CSCB06	5400	50	1000	18,232	<b>90</b>	104	103	123	110	13%
CSCB07	5400	50	2000	27,755	<b>121</b>	146	142	154	179	15%
CSCB08	5400	100	2000	28,005	<b>120</b>	134	144	166	174	10%
CSCB09	5400	6	250	4159	<b>22</b>	-	25	26	30	12%
CSCB10	5400	12	250	4221	<b>22</b>	-	28	30	32	21%
CSCB11	5400	12	500	6006	<b>36</b>	-	45	50	50	20%
CSCB12	5400	25	500	6098	<b>48</b>	-	<b>48</b>	55	57	0%
CSCB13	5400	25	1000	14,242	<b>99</b>	-	<b>99</b>	135	124	0%
CSCB14	5400	50	1000	16,318	<b>86</b>	-	91	117	109	6%
CSCB15	5400	50	2000	27,482	164	-	<b>159</b>	215	199	-3%
CSCB16	5400	100	2000	24,051	<b>111</b>	-	144	158	148	23%
Average	5400				<b>68.7</b>	74.3	78.5	91.6	92.5	10%

Table 1: Comparison of BiRD with existing methods on synthetic data benchmarks. We exhibit an average improvement of 18% when the maximum ride time is 45 minutes and 10% when the maximum ride time is 90 minutes, even though some of the presented methods allow mixed loads. Results with a \* indicate provable full optimality.

students to school given the constraints. On all instances, we compute seven scenarios for each school (using seven different values of  $\lambda$ : 100, 500, 5000, 10000, 50000, 100000, 500000). For each scenario, we combine randomized routes such that each stop is covered by 1600 routes, in 80 phases of 20 routes each – note: for two smaller problem instances, we compute sixteen scenarios for each school, and we reduce the number of routes and iterations. We then perform the scenario selection and bus scheduling steps with the overall number of buses as the only objective. The problem solves in under an hour on the largest instance.

We compare the performance of BiRD with existing methods in Table 1. The number of buses computed by BiRD is denoted by  $Z^*$ , compared to the number  $Z_{\text{Chen}}$  obtained by Chen et al. [2], the number  $Z_{\text{Park}}$  (respectively  $Z_{\text{PML}}$ ) obtained by Park, Tae and Kim [4] without mixed loads (respectively with mixed loads), and the number  $Z_{\text{Braca}}$  obtained by using the algorithm from Braca et al. [1]. We outperform all existing methods on all instances with  $T^{\max} = 2700$  shown in Table 1, and 22 out of 24 instances with  $T^{\max} = 5400$ . On average, we outperform the best existing method on each instance by 14%, even though some of the algorithms we compare against allow mixed loads. These routing results can be replicated using our Julia package released on Github [3].

Since BiRD is a heuristic, it provides no guarantees as to the optimality of the solution. Furthermore, the question of finding lower bounds for the school bus routing problem has received very little attention and remains very much open. To our knowledge, only Park, Tae and Kim [4] have made a serious attempt at finding a lower bound, and they themselves concede that much improvement is still needed. Interestingly, on the synthetic examples presented here, BiRD matches the lower bound in 3 of the 48 instances (admittedly, three of the smallest), the only method to provide provably optimal results for any of these synthetic examples. Because solving the school bus routing algorithm exactly is intractable for large instances, and good lower bounds do not really exist, quantifying the effect of the problem decomposition on the optimality of the final solution remains an open question.

## Bell Time Selection

In this section, we present the details of our mathematical formulation for the School Time Selection Problem (STSP) and describe our synthetic experiments.



## Transportation Costs

Given the complexity of the school bus routing problem when school start times are fixed, jointly optimizing bus routes and bell times is clearly a very intractable problem, which grows exponentially in size with the number of schools. The key idea of our approach is thus to find a reasonable proxy for the transportation cost of any start time assignment. We choose to define pairwise routing costs  $c_{S,t,S',t'}^{\text{routing}}$  in a bid to balance tractability with expressivity (pairwise costs allow us to capture interaction between pairs of schools).

The main intuition behind the routing costs  $c_{S,t,S',t'}^{\text{routing}}$  is the fact that costs are lower if individual buses can serve as many trips as possible. Therefore, the main factor in reducing the number of buses is the “compatibility” of groups of trips, i.e. how easy it is for a single bus to serve a certain set of trips without wasting time waiting or driving without passengers. Given two trips  $R$  and  $R'$ , let  $\Delta t$  be the time between the end of  $R$  and the beginning of  $R'$ . We define a piecewise linear compatibility cost  $c_{R,R'}$  that is low if it is profitable for a bus to serve the two trips sequentially. More precisely:

- $c_{R,R'} = 0$  if it is impossible for a bus to serve the two trips successively
- $c_{R,R'} = 0$  if  $\Delta t \geq \bar{T}$  with  $\bar{T}$  a compatibility parameter that defines the maximal time a bus can drive between two trips for them to be “compatible”.
- $c_{R,R'} = -\frac{\bar{T}-\Delta t}{\bar{T}}$  otherwise, i.e. the cost is  $-1$  when  $\Delta t = 0$  and  $R'$  can be served immediately after  $R$  and then increases linearly to 0 as  $\Delta t$  increases to  $\bar{T}$ .

For each school and each year for which we enrollment data is available, we compute a set of varied bus routing scenarios as described earlier. The scenarios are selected so that they are likely to be used in the optimal school bus routing solution. For each school, we therefore obtain a list of scenarios that is the union of all the scenarios obtained from each year of data. Then, for two schools  $S$  and  $S'$ , we can define a compatibility cost  $c_{S,t,S',t'}^{\text{compat}}$  that is the sum of the compatibility costs  $c_{R,R'}$  and  $c_{R',R}$  for every trip  $R$  in every routing scenario for school  $S$  and every trip  $R'$  in every routing scenario for school  $S'$  when the schools bell times are respectively  $t$  and  $t'$ .

Our experiments show that the costs  $c^{\text{compat}}$  are good approximation of how the choice of bell time allows the routes of different schools to be “compatible” across the years. Choosing bell times that maximize this compatibility indirectly minimizes the future transportation costs incurred by the district. It turns out that maximizing the compatibility of different routes as described has the

unwanted tendency to lead to a reduced number of schools with early and late start times, which has a negative impact on the number of buses in the solution. This is a consequence of using a simple pairwise affinity cost that only takes into account groups of two schools. In practice, we can counteract this adversarial effect by adding a cost that encourages bell times to be spread out over all allowed values:  $c_{S,t,S',t'}^{spread} = -|t - t'|$ . The final transportation costs are therefore defined as  $c^{routing} = c^{compat}(\bar{T}) + \gamma c^{spread}$  where  $\bar{T}$  and  $\gamma$  are the two parameters that depend on fundamental characteristics of the school district, and can be found using cross-validation. Ultimately, the transportation costs do not need to be perfect: year-to-year enrollment changes mean that directional information is more than enough in practice.

Given the routing costs defined above, and temporarily ignoring all other objectives, the STSP can be formulated as a Generalized Quadratic Assignment Problem (GQAP), for example using integer optimization:

$$\min \sum_{S \in \mathcal{S}} \sum_{t \in \mathcal{T}_S} \sum_{S' \in \mathcal{S}} \sum_{t' \in \mathcal{T}_{S'}} c_{S,t,S',t'}^{routing} z_{S,t,S',t'} \quad (5a)$$

$$\text{s.t.} \quad \sum_{t \in \mathcal{T}_S} a_{S,t} = 1 \quad \forall S \in \mathcal{S} \quad (5b)$$

$$z_{S,t,S',t'} \geq a_{S,t} + a_{S',t'} - 1 \quad \forall S \in \mathcal{S}, t \in \mathcal{T}_S, S' \in \mathcal{S}, t' \in \mathcal{T}_{S'} \quad (5c)$$

$$z_{S,t,S',t'} \leq a_{S,t} \quad \forall S \in \mathcal{S}, t \in \mathcal{T}_S, S' \in \mathcal{S}, t' \in \mathcal{T}_{S'} \quad (5d)$$

$$z_{S,t,S',t'} \leq a_{S',t'} \quad \forall S \in \mathcal{S}, t \in \mathcal{T}_S, S' \in \mathcal{S}, t' \in \mathcal{T}_{S'} \quad (5e)$$

$$z_{S,t,S',t'} \in \{0, 1\} \quad \forall S \in \mathcal{S}, t \in \mathcal{T}_S, S' \in \mathcal{S}, t' \in \mathcal{T}_{S'} \quad (5f)$$

$$a_{S,t} \in \{0, 1\} \quad \forall S \in \mathcal{S}, t \in \mathcal{T}_S. \quad (5g)$$

$$(5h)$$

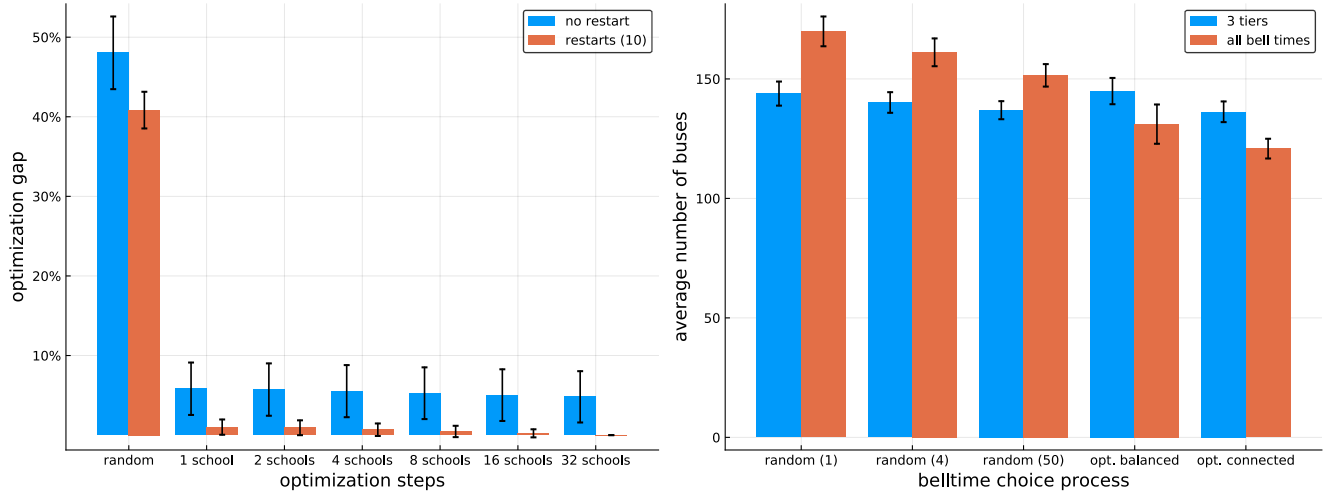
In the formulation above, the key decision variable  $a_{S,t}$  is 1 when school  $S$  is assigned time  $t$ , and 0 otherwise. Similarly, the decision variable  $z_{S,t,S',t'}$  is 1 when schools  $S$  and  $S'$  are respectively assigned times  $t$  and  $t'$ , and 0 otherwise. The set  $\mathcal{T}_S$  designates all bell times that are allowed for school  $S$  (this is a discrete, finite set, e.g. every 10 minutes between 7:30AM and 9:30AM). Constraint (5b) enforces that each school is assigned exactly one time, while constraints (5c), (5d) and (5e) enforce the relationship between the single and pairwise decision variables  $a_{S,t}$  and  $z_{S,t,S',t'}$ .

## Bell Time Optimization on Synthetic Data

We have formulated the STSP as a GQAP, and we solve it using a simple local improvement heuristic, which randomly selects a smaller subset of schools and optimizes their start times, keeping the bell times of all other schools fixed. Given an initial bell time assignment  $\{t_S^0\}_{S \in \mathcal{S}}$ , and having selected a subset  $\mathcal{S}_1 \subseteq \mathcal{S}$ , it turns out that the problem of finding the optimal start times for this subset is still a GQAP. We can formulate this GQAP as an integer program as above, solve it using a commercial solver such as Gurobi, and iterate until a stopping criterion is met.

We now turn to synthetic data to examine the effect of the parameters of the local improvement heuristic, namely the size of the subset and the number of iterations. The first tradeoff we explore is the size of the optimized subset. Clearly, as the number of schools in the optimized subset increases, the local improvement heuristic will find better solutions, but each iteration will take more time. To understand this tradeoff, we run 100 randomly-generated experiments with 100 schools. For each one, we first select a random starting point (leftmost column), then we perform 1024 iterations one school at a time, followed by 512 iterations two schools at a time, etc. Each time, we double the number of schools in the optimized subset and halve the number of iterations, so that in expectation each school features in the optimized subset the same number of times. We show results in Fig. 4a. We notice that compared to a random solution, optimizing one school at a time already drastically improves the quality of the solution, and subsequently increasing the size of the optimized subset does not have a strong effect on the quality of the solution. In addition, using random restarts, i.e. running the experiment several times with different random starting points and keeping the best one, has a very significant effect on the optimization gap. In fact, it seems that using several random restarts has a much stronger effect on the solution quality than increasing the number of schools that are optimized at each iteration.

To model transportation costs, we introduced costs  $c^{\text{routing}} = c^{\text{compat}}(\bar{T}) + \gamma c^{\text{spread}}$ . We claim that the optimal bell time assignment given these costs indeed induces a routing solution with a small number of buses. We support this claim with the experiments described in the Main Text, Fig. 3. We present another set of experiments here which also support the same point. We consider a problem instance with 100 schools, where the only objective is to minimize transportation costs, and the allowed bell times either follow 3 “tiers” (7:30, 8:30, 9:30) or encompass all 15-minute intervals between 7:15 and 9:30. We compare three optimization strategies. The first (“random”) assigns each school a bell time uniformly at random across the universe of possibilities (possibly with some random restarts to improve solution quality). The second (“balanced”) simply tries to



(a) Effect of subset size on quality of solution found by optimization-based heuristic, averaged over 100 synthetic experiments. As the subset size increases, the optimization gap decreases. In addition, random restarts have a stronger effect on the solution quality than increasing the size of the optimized subset.

(b) Comparison of different routing cost approximations for bell time optimization. Two settings are considered: three bell time tiers, i.e., schools may only start at 7:30, 8:30 or 9:30, and all bell times, when every 5-minute interval between 7:15 and 9:30 is allowed. The random assignment strategy performs well when there are only three allowed bell times, but poorly when all bell times between 7:15 and 9:30 are allowed. Approximating routing costs using the routing compatibility costs described in the SI section on Transportation Costs (“connected” experiment) gives better results in all cases than simply making sure the number of routes is balanced across tiers without regard for the actual compatibility of these routes.

Figure 4: Results of bell time optimization algorithm on synthetic data.

balance the number of routes into evenly spaced tiers, with a parameter controlling the spacing between the tiers which we can choose by cross-validation. The third (“connected”) is the one described in the Section “Transportation Costs”.

The results averaged over 100 random instances, which can be seen in Fig. 4b, show that our optimization strategy consistently and significantly outperforms the other two in both experimental settings. We notice that the random assignment strategy works quite well in the case when only three bell times are allowed, even better than the strategy that tries to balance routes into evenly spaced tiers. However, the random strategy is not able to make use of the additional allowed bell times in the second set of experiments, while both optimization strategies achieve significant improvements when the number of allowed bell times for each school increases.

## GQAP-Representable Objectives

As discussed in the “Bell Times in Practice” section of the Main Text, the start time assignment problem typically involves many objectives beyond the simple optimization of transportation costs. We present many such objectives here, and show how they can be integrated within the GQAP framework.

Many real-world objectives can be represented using single affinity costs  $c_{S,t}$ . Here are a few of the possibilities we explored in collaboration with BPS:

- Limiting change from the current bell times can be achieved by setting  $c_{S,t}$  to a positive value when  $t$  is different from the current time  $t_{\text{current}}$  for school  $S$ . The specific value can exhibit any functional dependence on  $t$  and  $t_{\text{current}}$ .
- Incorporating individual school preferences. Districts can easily quantify the preferences of various stakeholders at a particular school, from teachers and staff to parents and students, using surveys, focus groups, etc. These preferences can then easily be converted into aversion costs  $c_{S,t}$ .
- Favoring a particular school. Sometimes, a district may wish to prioritize the needs of a particular set of schools  $\mathcal{S}_0 \subseteq \mathcal{S}$ , in order to bolster academic achievements, support economically disadvantaged students, or provide a more auspicious environment for students with special needs. This objective can be achieved by penalizing less desirable times more for  $S \in \mathcal{S}_0$  than for  $S \notin \mathcal{S}_0$ . For example, in Boston, we optimized a preference score for

schools with a high number of special education students (weighted by the number of these students).

- Promoting later high school start times/earlier elementary school end times can be achieved by penalizing undesirable times for each school with a high cost.
- Interfacing with after-school programs/school-specific constraints. If a school must end before a certain time to leave time for a specific extracurricular activity, or to alleviate traffic congestion in the city, it is straightforward to compute the cost of such undesirable times and consider this as an objective.

Also allowing pairwise affinity costs  $c_{S,t,S',t'}$  increases modeling power by allowing the representation of more complicated real-world objectives:

- Allowing school partnerships. Groups of schools often partner to offer joint extracurriculars or athletic competitions. Schools can also share all or part of their transportation. Partner schools may therefore require compatible bell times, i.e. bell times that are within a particular interval.
- Considering externalities. School districts may wish to separate the start and/or end time of some pairs of schools to prevent fights between rival students or reduce the strain on public transportation.
- Ensuring equity. Single costs allow a school district to model average school satisfaction, while pairwise affinity costs also allow it to model the variance in satisfaction across neighborhoods, communities, or the entire district. Consider two sets of schools  $\mathcal{S}_1, \mathcal{S}_2 \subseteq \mathcal{S}$ , and let  $\mu_S^{(i)}$  be  $1/|\mathcal{S}_i|$  if  $S \in \mathcal{S}_i$  and 0 otherwise. Consider a metric which assigns cost  $\hat{c}_{S,t}$  to bell time  $t$  for school  $S$ . Then the squared difference  $\delta_{\mathcal{S}_1, \mathcal{S}_2}(\mathbf{t})$  of the mean of  $\hat{c}$  between the two considered subsets can be written as

$$\delta_{\mathcal{S}_1, \mathcal{S}_2}(\mathbf{t}) = \left( \sum_{S \in \mathcal{S}} \mu_S^{(1)} \hat{c}_{S,t_S} - \sum_{S \in \mathcal{S}} \mu_S^{(2)} \hat{c}_{S,t_S} \right)^2 = \left( \sum_{S \in \mathcal{S}} \left( \mu_S^{(1)} - \mu_S^{(2)} \right) \hat{c}_{S,t_S} \right)^2 \quad (6a)$$

$$= \sum_{S \in \mathcal{S}} \left( \mu_S^{(1)} - \mu_S^{(2)} \right)^2 \hat{c}_{S,t_S}^2 + \sum_{S \in \mathcal{S}} \sum_{S' \in \mathcal{S}, S' \neq S} \left( \mu_S^{(1)} - \mu_S^{(2)} \right) \left( \mu_{S'}^{(1)} - \mu_{S'}^{(2)} \right) \hat{c}_{S,t_S} \hat{c}_{S',t_{S'}}. \quad (6b)$$

The above objective is GQAP-representable, if we define single affinity costs  $\bar{c}_{S,t} = \left( \mu_S^{(1)} - \mu_S^{(2)} \right)^2 \hat{c}_{S,t}^2$  and pairwise affinity costs  $\bar{c}_{S,t,S',t'} = \left( \mu_S^{(1)} - \mu_S^{(2)} \right) \left( \mu_{S'}^{(1)} - \mu_{S'}^{(2)} \right) \hat{c}_{S,t_S} \hat{c}_{S',t_{S'}}$ , and this property

generalizes to arbitrary weights  $\mu$ , allowing districts to ensure equity across all communities and populations in a district.

We note that pairwise affinity costs are more general than single affinity costs since optimizing any single affinity cost  $c_{S,t}$  is equivalent to optimizing the corresponding pairwise affinity cost  $c_{S,t,S',t'}$ , which equals  $c_{S,t}$  when  $S = S'$  and  $t = t'$ , and 0 otherwise. For ease of notation, we choose to represent all single costs in this manner.

Our complete approach to the STSP is thus a multi-objective formulation. Specifically, given a set of GQAP-representable objectives  $\{c^\alpha\}_{\alpha=1}^{\alpha=A}$  (one of which could be the routing costs  $c^{\text{routing}}$  and corresponding priority weights  $\eta_\alpha$  (both of which are determined by the school district), we replace the routing-only objective (5a) with the weighted sum of all of the district's objectives:

$$\min \sum_{S \in \mathcal{S}} \sum_{t \in \mathcal{T}_S} \sum_{S' \in \mathcal{S}} \sum_{t' \in \mathcal{T}_{S'}} \sum_{\alpha=1}^{\alpha=A} \eta_\alpha c_{S,t,S',t'}^\alpha z_{S,t,S',t'}. \quad (7)$$

Policymakers are free to vary the priority weights  $\eta_\alpha$  to explore tradeoffs between competing objectives.

## Boston Community Survey

A typical example of a real-world objective that school districts must take into account is community satisfaction. We describe the data collected by BPS to understand the preferences of parents, teachers and staff.

When BPS began exploring the idea of bell time adjustment in the fall of 2016, they launched a community survey in order to understand the preferences of various stakeholders, including parents, teachers and staff. The survey included both an online and phone component. Parents and school staff were asked to score all bell times between 7:00 and 9:30 (every 15 minutes) between 1 (worst) and 7 (best).

To reduce noise in the survey (e.g. some parents rate all bell times as 1 or 2 while other parents rate all bell times as 6 or 7), we normalize these scores so that (a) they lie between 0 and 1, and (b) for any respondent, their favorite bell time is rated a 1 and their least favorite a 0. Then the preference score (or survey score) of a particular bell time assignment is the average (weighted by enrollment) of the preference scores of each school for their assigned bell time, where the preference score for a school  $S$  at a given bell time  $t$  is the average of the normalized preferences of all school

$S$ 's parents and staff for bell time  $t$ , where parent preferences carry twice as much weight as staff preferences (the ratio was decided by BPS). To handle schools with too few responses, we add three “dummy parents” to all schools’ responses, with preferences equal to the average of all parent preferences across the entire survey. The results in Fig. 5 and Table 1 rely on this community survey.

The main insight provided by the survey was the general disagreement of parents within each school. BPS realized that optimizing the average preference score was not very meaningful because every bell time would have both supporters and critics at every school. Therefore, they moved towards optimizing broader objectives (e.g. moving high schools later) rather than optimizing this particular preference score.

## Conclusion

We describe the BiRD algorithm, a new optimization-based heuristic approach to school bus routing which outperforms existing methods from the literature, as well as the first modeling and algorithmic solution to the School Time Selection Problem (STSP). The tools we develop can be useful to school districts, to reduce overhead operational costs and invest directly into students, in a way that fits the priorities and needs of the community.

## References

- [1] Jeffrey Braca, Julien Bramel, Bruce Posner, and David Simchi-Levi. A computerized approach to the new york cityschool bus routing problem. *IIE transactions*, 29(8):693–702, 1997.
- [2] Xiaopan Chen, Yunfeng Kong, Lanxue Dang, Yane Hou, and Xinyue Ye. Exact and metaheuristic approaches for a bi-objective school bus scheduling problem. *PloS one*, 10(7):e0132600, 2015.
- [3] Arthur Delarue and Sebastien Martin. SchoolBusRouting repository. <https://github.com/adelarue/SchoolBusRouting>, 2018.
- [4] Junhyuk Park, Hyunchul Tae, and Byung-In Kim. A post-improvement procedure for the mixed load school bus routing problem. *European Journal of Operational Research*, 217(1):204–213, 2012.