# Last Mile

Antonio Lodico, Bonnie Fan, Matt Samach, Shikha Goel

04.27.2020

## Resources Used

- Literature Review - CMU libraries - European Journal of Operational Research
- QGIS: Geoprocessing

## Team Work Plan

| Team Member | Task(s) |
|---|---|
| Antonio Lodico | Data Collection, Data Preparation, Recommendation |
| Bonnie Fan | Data Collection, Data Preparation, Implementation |
| Matt Samach | Formulation, Implementation, Analysis |
| Shikha Goel | Formulation, Implementation, Literature Review |

# Literature Review

## Article: The school bus routing problem: A review

- This article is a review of various research work on the school bus routing problem (SBRP), which is the overarching goal of this project and aims to provide a comprehensive view of the problem. SBRP is a combination of the following sub-problems:
- Data preparation: This step involves preparing the data for the other sub-problems. Data includes a road network, students(home address, destination school and whether the student is handicapped), schools(location, start and end times, maximum riding time of a student), vehicles(origin location, capacity) and an OD matrix(shortest travel times or distances between pairs of schools, student locations, and bus origin locations).
- Bus stop selection: This sub problem involves selection of bus stops and assigning students to various bus stops. This step is often excluded from literature with the assumption that bus stop locations are giving. 'The heuristic solution approaches for bus stop selection are classified into the location-allocation-routing (LAR) strategy or the allocation-routing-location (ARL) strategy'.
- Bus route generation: In this sub-problem, the routes are generated. The "route-first cluster-second" approach creates a single route using the travelling salesman algorithm and then breaks it into smaller routes. The "cluster-first route-second" approach first creates clusters of students and then each cluster has a route.
- School bell time adjustment: Most literature assumes that the start and end times of schools are constant but these can also be treated as decision variables.
- Bus scheduling: 'Route scheduling specifies the exact starting and ending time of each route and forms a chain of routes that can be executed successively by the same bus'.
- Classification of this problem can be done based on several characteristics such as number of schools(single, multiple), surroundings(rural, urban), time(morning, afternoon or both) etc.
- Objectives of the SBRP can also vary a lot like number of buses, total bus travel distance/time, total student travel distance/time, student walking distance(last mile), child's time loss etc.
- Constraints of the SRBP problem include vehicle capacity, maximum riding time, school time window, maximum walking distance, earliest pick-up time and minimum number of students to create a route.
- Solutions of 'SBRP are usually formulated as mixed integer programming (MIP) or nonlinear mixed integer programming (NLMIP) models'.However, these are often only for some of the sub-problems defined above and not all of them[1].

# Problem Statement

'Our goal is to design bus drop-off locations and bus routes, in order to maximize total social welfare'. This school bus routing problem (SBRP) project specifically considers the problem of a lack of transportation service for the last-mile distance that students have to walk (1-1.5miles for elementary and up to 2 miles for secondary children). This raises safety concerns for parents and is a barrier for using school buses.

Maximizing social welfare then implies minimizing a multi-objective objective function which includes the total bus travel distance of all students, the total walking distance (last-mile) of all students and the operating costs of the school district.

The geographic scope of the project is Pittsburgh Public Schools, Allegheny County. The project would involve the data preparation, bus stop selection and bus route generation sub-problems encountered in the literature review.

# Mathematical Formulation

## Model 1

Assumptions

- One bus
- One school

[1] Park, Junhyuk, and Byung-In Kim. "The School Bus Routing Problem: A Review." *European Journal of Operational Research*, vol. 202, no. 2, 2010, pp. 311–319., doi:10.1016/j.ejor.2009.05.017.

- Start and end at school
- One preselected stop per neighborhood
- Each stop must be visited once
- Not accounting for walking distances yet, just bus travel times, because only one bus
- Not accounting for operational costs because only one bus
- Infinite capacity on bus

Indices

$i, j, a = 1, \ldots\ldots., N$ *bus stops*

$s = 1, \ldots\ldots.., S$ *ordered trip segments*

Decision Variable(s)

$y_{ijs} = \{1, \text{ if bus travels from stop } i \text{ to } j \text{ in segment } s \ 0, \text{ otherwise } \}$

$b_s = $ *Auxiliary variable calculating the number of students on the bus*

Input Parameters

$p_j$ : *Students getting off at stop j*

$D$ : *Initial number of kids*

$b_s$ : *Number of kids remaining on bus during trip segment s*

$t_{ij}$ : *travel time between stop i and stop j*

$S$ : *total number of segments (number of bus stops + 1)*

Constraints

$b_s = D - \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{s=1}^{s-1} y_{ijs} p_j \ \forall s$ Kids on bus in during segment s

$\sum_{j=1}^{N} \sum_{s=1}^{S} y_{aj(s+1)} = \sum_{i=1}^{N} \sum_{s=1}^{S} y_{ias} \ \forall a \text{ and } \forall s$ Continuous route

$\sum_{i=1}^{N} \sum_{s=1}^{S} y_{ijs} = 1 \ \forall j$ Every bus stop is visited once

$\sum_{j=1}^{N} \sum_{s=1}^{S} y_{ijs} = 1 \ \forall i$ Every bus stop is left once

$y_{iis} = 0 \ \forall i, s$ Can't have same origin and destination

$\sum_{j=1}^{N} y_{0j1} = 1$ Start at school

$b_s$ *is positive integer* , $y_{ijs}$ *is binary*

Objective : Minimize total bus travel time of all kids

$Minimize \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{s=1}^{S} b_s t_{ij} y_{ijs}$ Total travel time of all kids

Note : This is a non-linear objective function, but Gurobi was able to implement a quadratic objective

## Model 2

Assumptions

- Multiple candidate bus stops to choose from a neighborhood, only one is chosen per hood
- One bus
- One school
- Start and end at school
- Each neighborhood must be visited once
- Not accounting for operational costs because only one bus
- Infinite capacity on

Indices

$i, j, a = 1 \ldots N$ *bus stops*

$s = 1 \ldots S$ *ordered trip segments*

$k = 1 \ldots K$ *stop zones /neighborhoods*

Decision Variables

$x_i = \{1,$ *if bus stop i is selected* $\quad 0,$ *otherwise*

$y_{ijs} = \{1,$ *if bus travels from stop i to j in segment s* $\quad 0,$ *otherwise*

$b_s =$ *Auxiliary integer variable calculating the number of students on the bus in time segment s*

Input Parameters

$D$ : *Initial number of kids on the bus*

$p_k$ : *Population in a neighborhood k*

$b_s$ : *Number of kids currently on bus during segment s*

$T_{ij}$ : *travel time between stop i and stop j*

$h_i$ : *Average walking distance (in neighborhood) for kids to get home from stop i*

$v_{ik}$ : *Binary, denotes if stop i is in neighborhood k*

Constraints

$b_s = D - \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{z=1}^{s}\sum_{k=1}^{K} y_{ijz} p_k v_{jk} \ \forall s$ — Kids on the bus in segment s

$\sum_{j=1}^{N}\sum_{s=1}^{S} y_{aj(s+1)} = \sum_{i=1}^{N}\sum_{s=1}^{S} y_{ias} \ \forall a \text{ and } \forall s$ — Continuous route

$\sum_{i=1}^{N}\sum_{s=1}^{S} y_{ijs} = x_j \ \forall j$ — A stop is a dest only if it's been selected

$\sum_{j=1}^{N}\sum_{s=1}^{S} y_{ijs} = x_i \ \forall i$ — A stop is an origin only if it's been selected

$y_{iis} = 0 \ \forall i, s$ — Can't have same origin and destination

$\sum_{j=1}^{N} y_{0j1} = 1$ — Start at school

$\sum_{i=1}^{N} x_i v_{ik} = 1 \ \forall k$ — Each neighborhood is visited exactly once

**Objective :** Minimize Total bus travel time of all kids + total last mile walk travel time of all kids

$$min \ \alpha \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{s=1}^{N} b_s T_{ij} y_{ijs} + (1-\alpha)\sum_{i=1}^{N}\sum_{k=1}^{K} x_i h_i p_k v_{ik}$$

Note : This is a non-linear objective function but the implementation in Gurobi is able to handle it and no linear transformation was needed.

# Data Summary

To explore an implementation for a single school zone with multiple neighborhoods, we took Minadeo PreK-5 as our real-world implementation case. We scraped the Minadeo school boundary geojson and performed an overlay to arrive at neighborhood zones. Each neighborhood fell within a 5 larger zip code boundaries (15207, 15208, 15217, 15218, 15221), which was what we used to define our school stop zones. Each zip code contains two candidate stops, one at each neighborhood (10 stops total). We used QGIS to generate a distance matrix in meters between the 10 stops and Minadeo itself as the 11th stop. This distance was converted to minutes with an estimated speed of 30mph for the bus. The number of children per neighborhood was estimated with a proportionate fraction (by area) out of the total school population (251). We assume all 251 children take the bus home. Walk times per zip code were calculated by summing walk times over the number of children by zip code. (See Appendix - A).

# Implementation

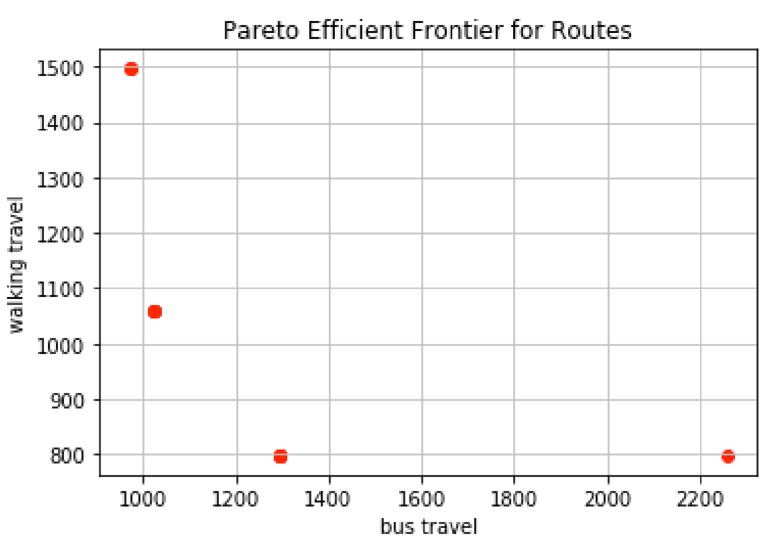
We successfully implemented both version 1 and version 2 formulations in Gurobi. We first attempted a linearized implementation (substituting for our quadratic variable with constraints) which did not yield results. We then implemented our formulation without the linearization and Gurobi produced feasible solutions. In order to test the solutions, we first used toy data and tried multiple edge cases for reasonableness check of the solutions. Once satisfied, we used the data we prepared for Minadeo in Pittsburgh and obtained our final solutions.

# Analysis

In the first model, we visited every stop and thus did not need to account for different walking times. Unsurprisingly, with more stops visited, there is a greater total amount of time spent on the bus for Model 1 compared to Model 2. However, Model 2 is visiting half the number of stops and optimizing on the walking time between students. (See Appendix for the mapped stops).

**Model 1 Stop Sequence Solution:**
**0 > 4 > 3 > 1 > 6 > 2 > 5 > 8 > 9 >10>7 >0**
**Model 1 Time on Bus = 1472 minutes**

**Model 2 Stop Sequence Solution**
**0 -> 4 -> 3 ->2 -> 8 -> 10 -> 0**
**Model 2 Time on Bus = 1024 min**




On the left, we output a Pareto frontier between total last mile walking time and bus travel time for Model 2. This shows that if we could value bus and walking time equally for around 1000 total minutes, but presumably we would want to reduce walking time - thus the solutions towards the bottom right would be desirable, with walking time around 800.

# Analysis of Bonus Questions

## 1 - Uber-like Service

One potential modification of this problem is an Uber-like solution (think Uber-Kids). Instead of having all kids ride the bus home, the school would have the option of contracting this service to have some set of kids take personal vehicles from school to home. If done right, this solution could seriously reduce total bus travel time, aggregate time travelled by kids on the bus and aggregate last-mile walking distance by kids.

As with any optimization problem, solution costs must be considered. Chief among these is that Uber-like services tend to have much higher per-rider operating costs than collective transportation solutions like bus transportation. This likely means that our model would tend to assign students to this service only in fairly extreme cases that allow the bus to operate more efficiently for the majority of other riders.

There are many potential benefits to this kind of personalized service. Some neighborhoods will have relatively low student populations or be far out of the way of other neighborhoods. In these edge cases, outsourcing to an Uber-like service would get these kids off the bus, thus freeing the route planning to travel more efficient routes through higher density neighborhoods. This would very likely decrease bus total bus travel time, aggregate student bus time as well as aggregate last mile walk time.

### 2 - Public Transit

Another modification of this problem would be to allow public transit to pick some of the slack off of the school system. This modification would also generate flexibility in terms of routing. Students could either take public transit for the entire duration of their trips home, or the school bus could drop students off at public transit stops to transfer.

There are drawbacks to this approach as well. In most cities, public transportation still costs money, meaning that either the school or the students would have to bear that financial cost. More importantly, public transit routes and scheduling are completely out of the hands of the school. This lack of control could mean high bus transit times (if routes don't quickly get to student neighborhoods) or high last mile times (if stops are far from student neighborhoods).

Benefits are similar to the Uber-service benefits. The school would be freed up to build more efficient routes, significantly decreasing the travel times of students taking the bus.

## Recommendations "All models are wrong, some are useful"

Our model, as implemented, is one that should not be directly applied or recommended for bus routes in Pittsburgh as is. That said, it could work for a very small school (<50) in areas with half mile street grids and very low (and relatively uniform) student density, as found in pockets of the exurban midwest. It's greatest value lies in as a starting point, a baseline for more robust models that would have fewer assumptions and relaxations. Below is a list of key assumptions and relaxations that could be changed in future iterations of the model to better approximate real world conditions

1. The number and location of students within each zone was generated assuming uniform distribution of students who all walked 2.5mph within square shaped zones and grid patterned walking options. With data of student residences and ages, combined with GIS software generating distances along existing walkways, we could more accurately model the walking distances to any potential bus stop. We could also find the point within a zone that minimizes walk time for students instead.
2. The model is set up currently to pick up all students in a school. We could approximate walk times of students to school, roughly by adding walking times to stop zones with slowed down bus times from zones to school. For a more robust model, we could plot walking time by GIS modeling from home to school, and have that as a bus alternative. It could also be a substitute within a 1.5 mile radius.
3. The current model had every student taking the school bus to school. In addition to adding walking as mentioned above, future models could add public transportation, ridesharing, or any combination of the above as substitutions or alternatives
4. The model currently has all students being picked up by one very large bus. We could allow for multiple buses, capping the number of students per bus. This would need to include adding a constraint or secondary goal like cost minimization or at least 40 kids/bus.
5. The model currently has only one stop option within a zone/neighborhood. Once the above recommendations were implemented, it would be possible to decouple zones and stops. Instead, stop options could be created for every street block or corner, without preassigned students to a stop option. It would require a constraint that every student who would catch a bus would need to walk to an active stop on a bus route, with the walktimes for each student to each stop would be calculated.
6. Most importantly, our current formulation and implementation only takes one goal into account: minimizing the transit times for students. A more accurate model would need to balance this against school district budgets, equity of services provided to students, parental input, ensuring no single student had a two hour bus ride for the greater good, and so on. This would require a multiple goal approach, parental input, ensuring no single student had a two hour bus ride for the greater good, and so on. This would require a multiple goal approach.

# Appendix - A Data Preparation

To narrow scope and simplify computation, we chose one school: Minadeo Elementary School.

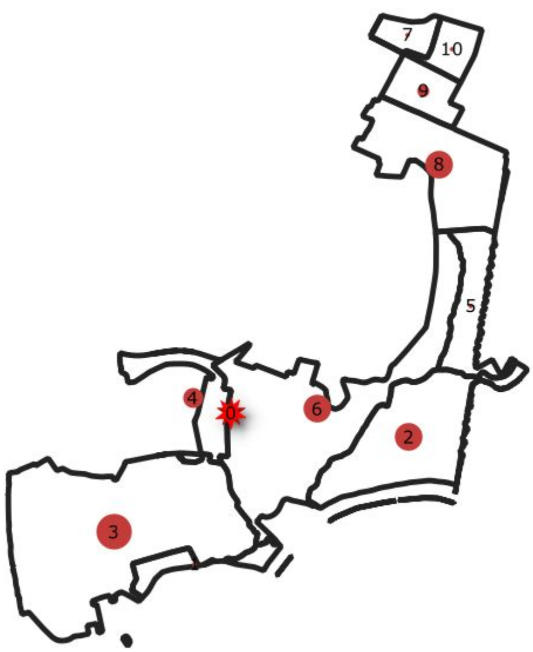
| **Parameter** | **Data form/ dim** | **Data Collection** | **Data Processing** |
|---|---|---|---|
| Bus stop i=0 | | Minadeo K-5 (stop 0) and other schools found in Pittsburgh Public School Locations - WPRDC | Joined to bus stops (see below) |
| Neighborhoods/Stop Zones k=1..K | | Found in Pittsburgh Neighborhoods GIS Boundaries and Minadeo School Boundaries Map | Scraped geojson definition from Minadeo leaflet map, performed QGIS union on neighborhood boundaries to create one stop zone per neighborhood |
| Bus stop i=1..N | 1 X N matrix | Generated with Stop zone K data | Generated centroid of stop zone with QGIS |
| D: initial number of kids on bus | | Total students enrolled by school.<br>Found in PA Future Ready Data Files | Cleaned data file, found 251 total enrolled in 2018 |
| $p_k$ : Pop. in neighborhood/stop zone k | | Students enrolled by school (above) | (Students at a school * Area ratio of stop zone ) = $p_k$ |
| *Tij :travel time b/t stops i & j* | | Generated with Bus stop i data | Generated travel distance between all stops 0..11 with QGIS in a travel matrix, calculated travel time in minutes per meter (30mph) |
| *hi :Avg dist from home of student to bus stop i* | | Generated with Stop zone k, bus stop i data | Initial estimate: square root of stop zone area (in mi)/2 = avg dist/student within a Stop zone/neighborhood (avg dist to center of a square). Converts to time by avg student walks at 2.5mph. |

# Appendix - B Map representations of Stops


Model 1 : 0 > 4 > 3 > 1 > 6 > 2 > 5 > 8 > 9 >10>7 >0


Model 2 : 0 -> 4 -> 3 ->2 -> 8 -> 10 -> 0

Below is a representation of stops by population count by neighborhood.




The size of the circles on this map represent the number of students at each stop

# Appendix - C Bus Loads over Time