Summer 2017 Transit Path Analysis Project Report

Travel Modelling Group—University of Toronto Transportation Research Institute

Faculty of Applied Sciences and Engineering | University of Toronto

By Yuan Hong (Bill) Sun, EngSci 2T0

Supervised by Prof. Eric J. Miller (Dept. of Civil Engineering) and Bilal Yusuf (TMG)

August 31st, 2017

1. Background

1.1. General Concept

In path choice modelling, one needs to compare the observed paths taken in real life from a survey to the paths generated by the computer simulation. This is to analyze the similarity of the modelled paths and the observed paths, to ensure the correctness of the modelling. The model developed will be useful for forecasting future travel demand, thus assisting with the planning of future transit infrastructure.

For a path choice analysis of the Greater Toronto Area (GTA), Emme Modeller [1] (Emme Version 4.3) was used in conjunction with XTMF [2] (Version 1.3) to allow for repeated transit assignments using different sets of parameters, and the path choices generated by each assignment is compared to the observed paths gathered by Transportation Tomorrow Survey (TTS) [3]. The methods used for comparing path similarity are detailed below in Section 1.3. The Emme base network is constructed using 2011 data (2011 Base Network), and the TTS observed data is from 2012. XTMF uses a particle swarm optimization program to generate different values for the parameters that optimize the similarity between the two data sets (parameter optimization). The list of default parameters and their values are in Appendix 8 [8].

In the Emme and TTS networks, zones and centroids defined relative to the population density. All the paths are set to go from the centroid of one zone to the centroid of another. Each centroid is connected to the nearby routes by centroid connectors generated by Emme. In the network, each zone in the GTA is represented by an integer ranging from 1 to 9999, depending on the region. A detailed zone region chart is shown in Appendix 5 below [5]. This method is called zone and centroid based transit assignment.

1.2. Path Details

Every path is defined by its origin-destination (OD) pair, as well as all the segments of public transit routes taken (all the other path details are omitted for simplicity). A proportion value, ranging from 0.0 to 1.0, represents the probability that a path is taken in Emme for a certain origin-destination pair; however, all the Emme-generated paths with proportion values that are too low (<0.1%) can be neglected since they are very unlikely to be taken, and should not significantly affect the final results. From testing, it was found that not limiting the path proportions would allow Emme to generate path files that are too large (> 1 GB) to be analyzed efficiently, as many paths with extremely small proportions are included.

The observed paths file and the Emme-generated path file are described in Section 1.4 below.

1.3. Data Comparison Methods

Three methods have been improvised to represent the similarity of the Emme-generated path data verses the TTS observed path data. However, the final method (logarithmic fitness sum) is currently the preferred method of comparison.

1.3.1. Percentage of observed paths that appear in Emme-generated paths (P)

For each OD pair, every observed path is compared with the list of Emme-generated paths for that OD pair to check if the exact observed path exists. The closer the final value is to 100%, the better.

$$P = 100\% \times \frac{\text{Number of observed paths that appear in Emme generated paths}}{\text{Total number of observed paths}}$$

1.3.2. Average max. % similarity of all observed paths with their closest-matched Emme path (S)

For each OD pair, every observed path is compared with the list of Emme-generated paths for that OD pair to check for the closest match by segments of transit routes taken (order is considered).

For example, if the observed path is [‘T001’, ‘T002’, ‘T003’], and the two Emme-generated paths for the same OD pair are [‘T000’, ‘T002’, ‘T003’, ‘T004’] and [‘T001’, ‘T005’, ‘T004’], then the maximum percent similarity is 2/3 as two out of three segments of the observed path appear in the first Emme path in the same order (‘T002’ & ‘T003’). Moreover, if the observed path appears exactly in the list of Emme-generated paths for the same OD pair, then the maximum percent similarity is 1.0.

The sum of all the percentages is divided by the total number of observed paths. The closer the final value is to 100%, the better.

$$S = 100\% \times \frac{\sum (\text{Max no. of path segments that match in order / No. of segments in the observed path})}{\text{Total number of observed paths}}$$

1.3.3. Logarithmic fitness sum:

For every OD pair, the probability that the observed paths are chosen is computed (exact method will be discussed in section 2 below). A small beta value (β), usually 0.1 or less, is taken. The fitness sum should be negative. The smaller the magnitude of the final value (i.e. closer to 0), the better.

$$Fitness = \sum ln\ (\frac{\text{P (path chosen in Emme)} + \beta}{1.0 + \ \beta})$$

$$\text{P (path chosen in Emme)} = \sum(\text{Proportion of Emme path if (Emme path == observed path)})$$

1.4. Input Data Details

1.4.1. Emme to TTS Line ID File (.csv):

For every transit route, the route ID in Emme and TTS are shown. This is needed to convert the transit route IDs in the Emme-generated paths. A sample data sheet is shown on the right.

	A	B
1	emme_id	tts_id
2	B001A	BR01
3	B001B	BR01
4	B001C	BR01
5	B001D	BR01
6	B001E	BR01
7	B001F	BR01
8	B001G	BR01
9	B001H	BR01
10	B001I	BR01
11	B001J	BR01
12	B002A	BR02
13	B002B	BR02
14	B002C	BR02

1.4.2. TTS Observed Paths File (.csv):

The first two columns are the origins and destinations. The next 6 columns (Access mode, Egress mode, Subway on, Subway off, Go on, Go off) are ignored, and then the transit routes (Route_1 - Route_6) are stored. A sample data sheet is shown below.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	OriginZon	Destinatic	AccessMo	EgressMo	SubwayO	SubwayOf	GoOn	GoOff	Route_1	Route_2	Route_3	Route_4	Route_5	Route_6
2	3869	9998	W	W	0	0	0	0	XT99	0	0	0	0	0
3	3338	9998	W	W	0	0	0	0	XG99	0	0	0	0	0
4	3338	9998	W	W	0	0	0	0	XG99	0	0	0	0	0
5	3338	9998	W	W	0	0	0	0	XG99	0	0	0	0	0
6	2443	9998	W	W	0	0	GS40	SS38	GT07	XR99	0	0	0	0
7	1093	9998	D	W	0	0	0	0	XR99	0	0	0	0	0
8	1093	9998	D	W	0	0	0	0	XR99	0	0	0	0	0
9	597	9998	D	W	0	0	0	0	XT99	0	0	0	0	0
10	597	9998	P	W	0	0	0	0	XT99	0	0	0	0	0
11	313	9998	P	W	0	0	0	0	XT99	0	0	0	0	0
12	295	9998	W	P	0	0	GS09	SS38	GT01	XR99	0	0	0	0
13	218	9998	W	W	0	0	0	0	XT99	0	0	0	0	0
14	152	9998	P	W	0	0	0	0	XT99	0	0	0	0	0
15	152	9998	P	W	0	0	0	0	XT99	0	0	0	0	0
16	152	9998	P	W	0	0	0	0	XT99	0	0	0	0	0
17	131	9998	W	P	0	0	0	0	T032	0	0	0	0	0
18	129	9998	W	W	0	0	0	0	XT99	0	0	0	0	0
19	101	9998	W	W	SS11	SS48	0	0	T596	T593	XT99	0	0	0
20	21	9998	W	W	SS17	SS41	0	0	T596	T594	XH99	0	0	0
21	5181	5252	W	W	0	0	0	0	HA98	0	0	0	0	0

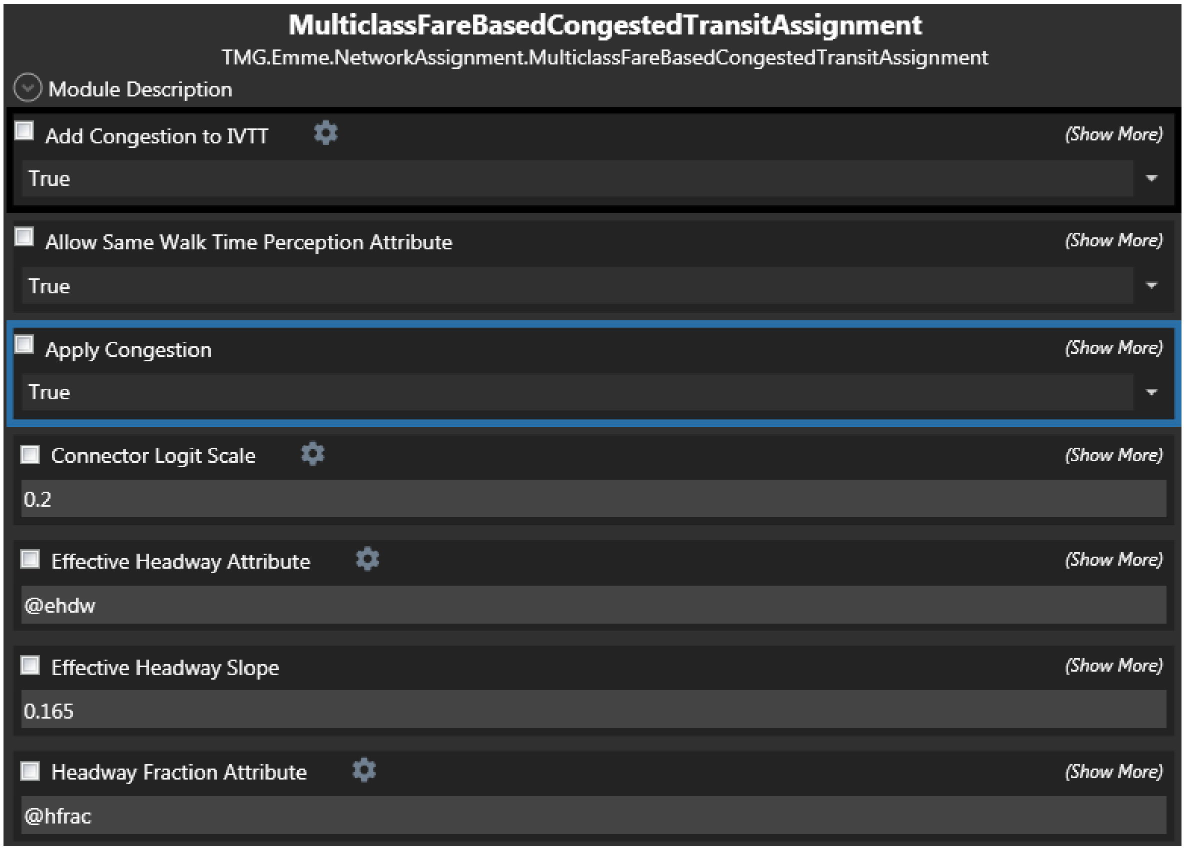
1.4.3. Emme-Generated Paths File (.txt):

The first two columns are the origins and destinations. The third and fourth columns are proportions and volumes, respectively. Each path segment includes mode, transit line ID (or a dash if the mode [5] of the segment is not transit), and the destination node. Each individual path generated by Emme is unique. However, when reading in the file, only the segments with transit modes are considered, and only the transit routes are stored. This is to cut down the amount of data stored in memory and increase runtime efficiency. A sample data sheet is shown below.

```
c
c paths with proportion in 0.001000,9999.000000
c
c orig dest pathnum prop vol orig <aux. transit> <transit>
c      <aux. transit>  mode -     node
c      <transit>       mode line node
c
17 1 1 0.0393 2.34 17 v - 13470 w - 10367 w - 11256 t - 101395 s 'T501Ba' 103598 t - 13948 v - 1
17 1 2 0.0363 2.163 17 v - 13470 w - 10367 w - 11256 t - 101395 s 'T501Ha' 103598 t - 13948 v - 1
17 1 3 0.0108 0.6455 17 v - 13470 w - 10367 w - 11256 t - 101395 s 'T502Aa' 101696 t - 11588 w - 12239 v - 1
17 1 4 0.327 19.47 17 v - 12998 t - 102936 s 'T501Ba' 103598 t - 13948 v - 1
17 1 5 0.3 17.84 17 v - 12998 t - 102936 s 'T501Ha' 103598 t - 13948 v - 1
17 1 6 0.0784 4.669 17 v - 12998 t - 102936 s 'T502Aa' 101696 t - 11588 w - 12239 v - 1
17 1 7 0.0825 4.912 17 v - 10400 w - 10371 w - 11506 t - 101621 s 'T501Ba' 103598 t - 13948 v - 1
17 1 8 0.0767 4.569 17 v - 10400 w - 10371 w - 11506 t - 101621 s 'T501Ha' 103598 t - 13948 v - 1
17 1 9 0.0289 1.721 17 v - 10400 w - 10371 w - 11506 t - 101621 s 'T503Ba' 101696 t - 11588 w - 12239 v - 1
17 1 10 0.0205 1.221 17 v - 10400 w - 10371 w - 11506 t - 101621 s 'T502Aa' 101696 t - 11588 w - 12239 v - 1
36 1 1 0.187 2.333 36 v - 10973 w - 10974 w - 10343 t - 100408 s 'T501Ba' 103598 t - 13948 v - 1
36 1 2 0.168 2.089 36 v - 10973 w - 10974 w - 10343 t - 100408 s 'T501Ha' 103598 t - 13948 v - 1
36 1 3 0.0271 0.3382 36 v - 10973 w - 10974 w - 10343 t - 100408 s 'T502Aa' 101696 t - 11588 w - 12239 v - 1
36 1 4 0.232 2.894 36 v - 10971 w - 10277 t - 100331 s 'T501Ba' 103598 t - 13948 v - 1
36 1 5 0.207 2.581 36 v - 10971 w - 10277 t - 100331 s 'T501Ha' 103598 t - 13948 v - 1
36 1 6 0.0285 0.3548 36 v - 10971 w - 10277 t - 100331 s 'T502Aa' 101696 t - 11588 w - 12239 v - 1
36 1 7 0.0163 0.2032 36 v - 10276 t - 97001 m 'TS01Gb' 97002 t - 100331 s 'T501Ba' 103598 t - 13948 v - 1
36 1 8 0.0145 0.1812 36 v - 10276 t - 97001 m 'TS01Gb' 97002 t - 100331 s 'T501Ha' 103598 t - 13948 v - 1
36 1 9 0.002 0.02491 36 v - 10276 t - 97001 m 'TS01Gb' 97002 t - 100331 s 'T502Aa' 101696 t - 11588 w - 12239 v - 1
```

1.5. Transit Assignment Tool

The tool that was used for transit assignments is called Multiclass Fare Based Congested Transit Assignment, in the TMG Toolbox in Emme Modeller. This tool can only be called and specified within XTMF. In all the assignments, congestion was applied (highlighted in blue below).

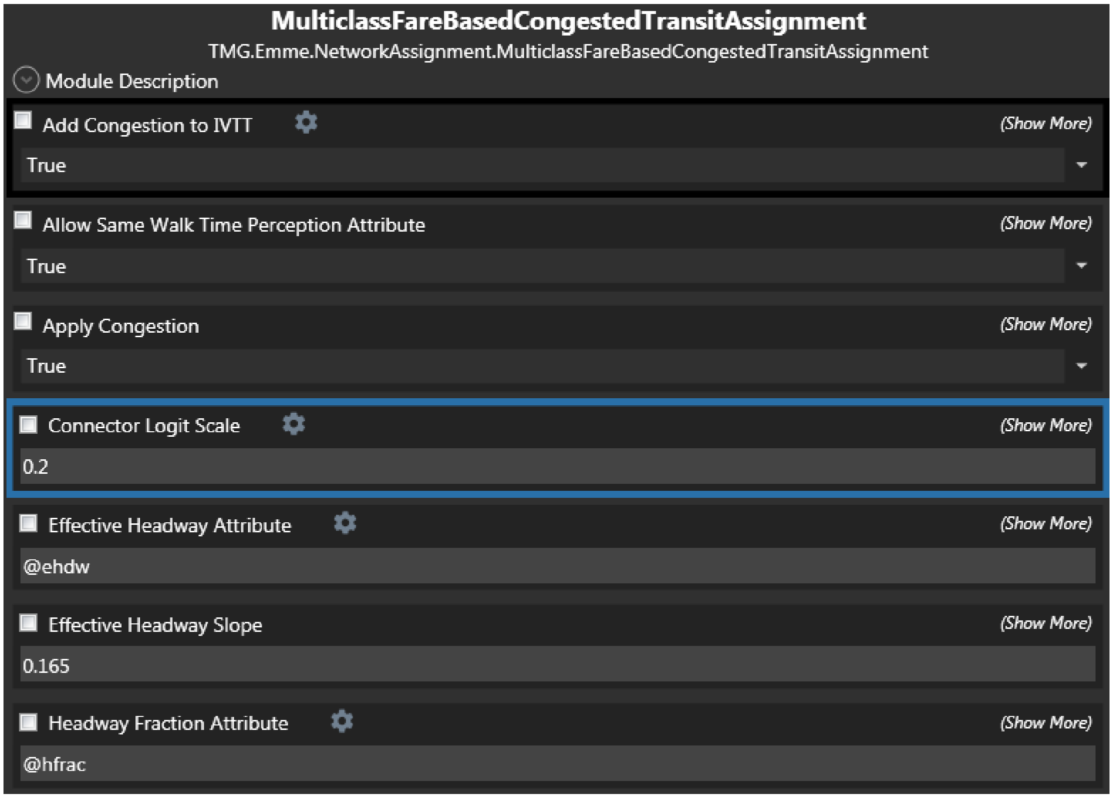


1.5.1. Connector Logit Scale

The connector logit scale (highlighted in blue below) is a parameter used in the multinomial logit model [7] to control the probability of the centroid connectors taken. A value of 1 means no logit model at all and only the best centroid connector was taken. A value of 0 means all the connectors are used equally.

During experimental model runs, the connector logit scale was lowered to 0.001 and it was found that more paths were generated, and that the P and S values (Section 1.3) were significantly higher as more observed paths are found in the Emme-generated paths. For example, using the 2011 Base Network and the default parameters in Appendix 8 [8], using a connector logit scale of 0.2, P = 51.7% and S = 69.7%. Using a connector logit scale of 0.001, P = 58.0% and S = 73.6%.

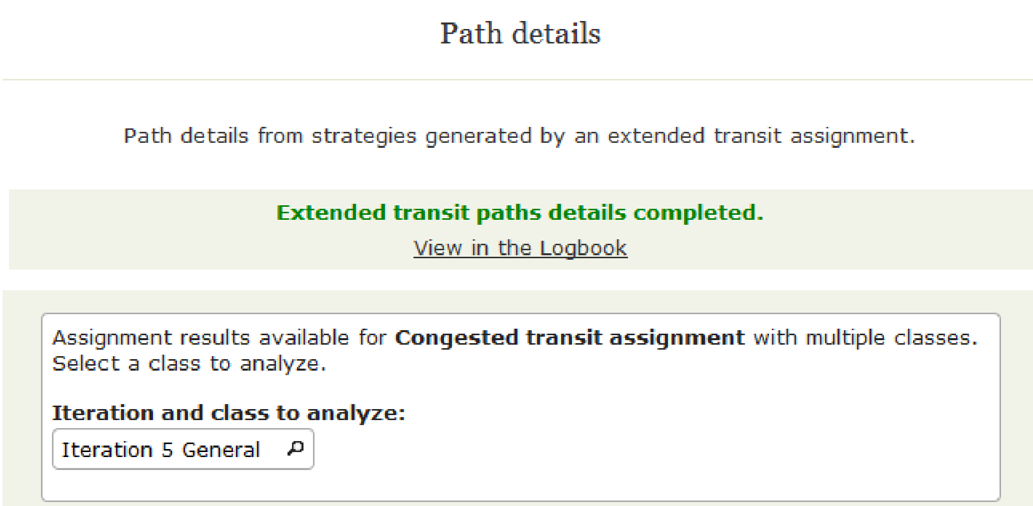
However, in order to achieve a balance in the system (realistically there is no equal chance for each centroid connector to be used), a default value of 0.2 was used for all the model runs.



1.5.2. Iteration Number

The congested transit assignment is ran with 6 iterations per generation, numbered from 0 to 5. From test runs, it was noted that as the iterations progress, the path data usually becomes better (i.e. more similar to the observed path data) as Emme better organizes them. For example, using the 2011 base network and the default parameters, with all the other conditions staying the same, iteration 0 gives P = 51.2% and S = 69.8%; iteration 5 gives P = 51.7% and S = 69.7%.

Therefore, to get the best possible path data, the default iteration number is set to 5 for all model runs.



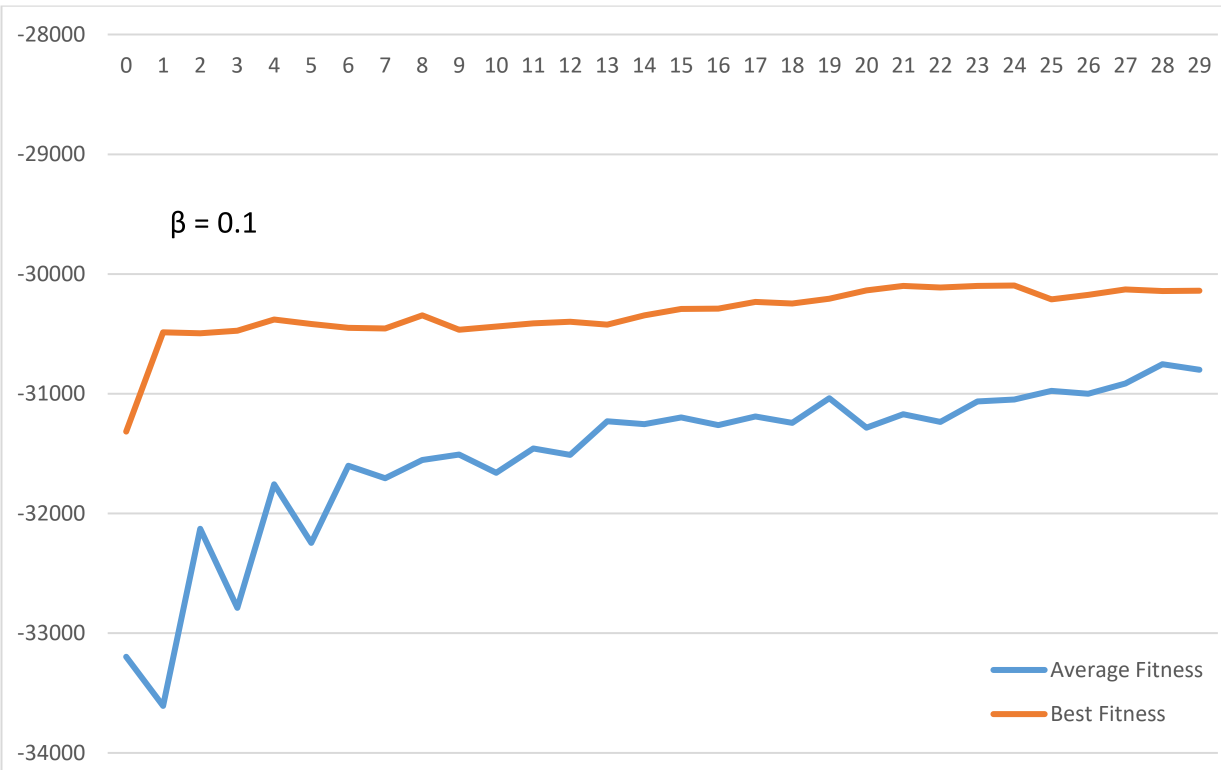
2. Problem

Through parameter optimization by XTMF, is has been observed that Emme has trouble precisely modelling path choices using zone and centroid based path assignment.

Currently at the best, using the default parameters and settings as noted in Section 1.5, up to 56.05% of the observed paths are generated by Emme (P). The best average maximum percent similarity of all observed paths with their closest-matched Emme path was up to 71.91% (S). A satisfactory result would be at least 80% for both of these values.

2.1. Initial Estimation Run

As shown in the graph below, in the final path choice estimation run, using β = 0.1, the best fitness value from all generations could only approach -30000. A desirable result should be as close to 0 as possible. In addition, from Generation 0 to Generation 24 (where the best fitness value occurred), the best fitness value for each generation only converged by 3.895%.



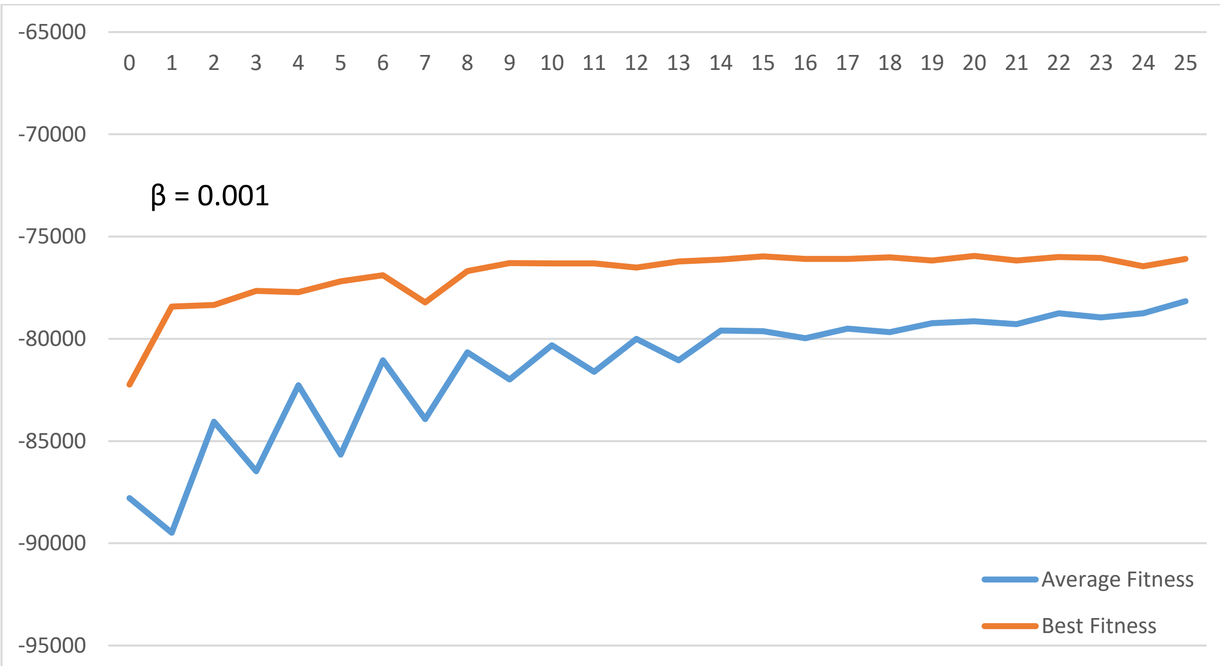
2.2. Modified Estimation Run

It was examined that using a smaller β value would result in larger differences in the fitness (1.3.3). For example, if for two observed paths in an OD pair, P (path 1 chosen) = 0.5, P (path 2 chosen) = 0;

Using β = 0.1, Fitness (path 1) = ln ((0.5+0.1) / (1+0.1)) = -0.606
Fitness (path 2) = ln ((0+0.1) / (1+0.1)) = -2.398 the difference is 1.792

Using β = 0.001, Fitness (path 1) = ln ((0.5+0.001) / (1+0.001)) = -0.692
Fitness (path 2) = ln ((0+0.001) / (1+0.001)) = -6.909 the difference is 6.217

Therefore, using a smaller β value, any differences between the observed paths and the Emme-generated paths would be more clearly reflected, by showing a larger change in the fitness value for each path. Thus, β = 0.001 was used for this modified path choice estimation run. The results are shown in the graph below.



Again, this is not a desirable result since it is far from 0. However, from Generation 0 to Generation 20 (where the best fitness value occurred), the best fitness value for each generation converged by 7.651%. This is a slight improvement from the original run, but still far from an optimal result.

Overall, these results showed that parameter optimization still makes Emme generate significantly different paths compared to the observed paths from TTS. This means that parameter optimization is not the only solution to resolving this issue, and that other methods are needed to estimate the path choices. Further investigation is shown in detail below.

3. Hypothesis

The fact that Emme cannot completely model path choices across the GTA is probably because the limitations of zone and centroid based path assignment. Since each zone covers a rather large area (up to 2-3 km across in some cases; see the case study (Section 4.2) below for a map of the zones), or have a strange shape other than a square/rectangular block, starting or ending a trip from different ends of the zone can result in significant differences in the path choice. The TTS path data is also presented with only the start and end zones, not the exact locations.

It is also hypothesized that this problem is worse in the inner city of Toronto since there are more transit routes available in a smaller area. In the suburbs, the limited transit options over a larger area meant that there are less possible paths from the variability of start and end locations; therefore, it like more likely for entire paths to match (or completely not match). Zone and centroid based path assignment raises these potential differences, by assigning all the paths to start and end from the centroids only, and not any random point.

Another potential reason is that most centroids of a zone only have 4 connectors to the surrounding main roads; these centroid connectors are determined by the population density of the region (See map below in Section 4.2 for details). This forcibly allows the paths to only traverse through these centroid connectors and the labelled main roads, not anywhere else on the map. It limits many potential path options, especially paths that go through places without centroid connectors.

Other possible factors that explain the differences include walk perception, transferring between lines, and the fact that a few people may take straight paths.

4. Observations and Analysis

4.1. Origin-Destination Paths Analysis

To examine the differences between the Emme-generated paths and the TTS observed paths, the paths from selected OD pairs are taken and manually compared.

To verify the limitations of zone and centroid based path assignment, as well as the positioning of the centroid connectors, Google Maps was used to show the change of most convenient path (least time by transit) by moving start and end locations within their zones.

To look at all the OD pairs that have observed paths that do not appear in the list of Emme-generated paths, a Python program is used to collect all the path data, and compare the differences.

4.1.1. Example: (133, 181)

The exact locations of the centroids of Zone 133 and Zone 181 are in shown in Figure 1 below (represented by the green circles), and Figure 2 shows the area covered by the zones (highlighted in light blue). Zone 133 is on the left side while Zone 181 is on the right side.

A Python program was made to analyze the Emme-generated path data for that OD pair, by adding up the proportions for all the paths that are have the same transit segments. All the paths and their proportions are printed out in the order of decreasing proportions, as shown in Figure 4 below.

On Google Maps, at first, the start and end locations were set to the exact location of the centroids. It suggested that the most convenient path is to take bus route 52 (T052) east, and then take bus route 7 (T007) north (highlighted in colour in Figure 3). This corresponds exactly with the most travelled path as suggested by Emme, which has a proportion of 0.4479 (44.79%), as shown in Figure 4.

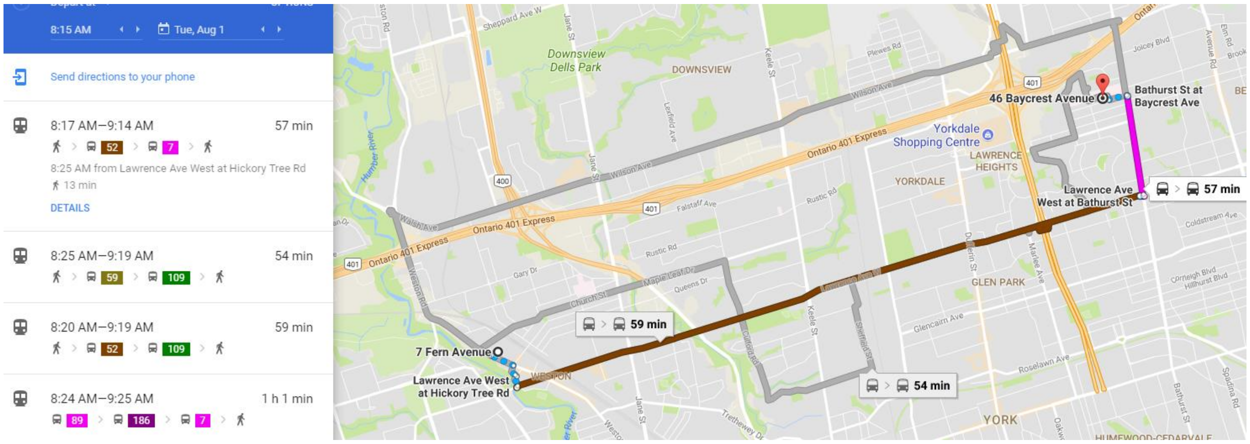


Fig. 3

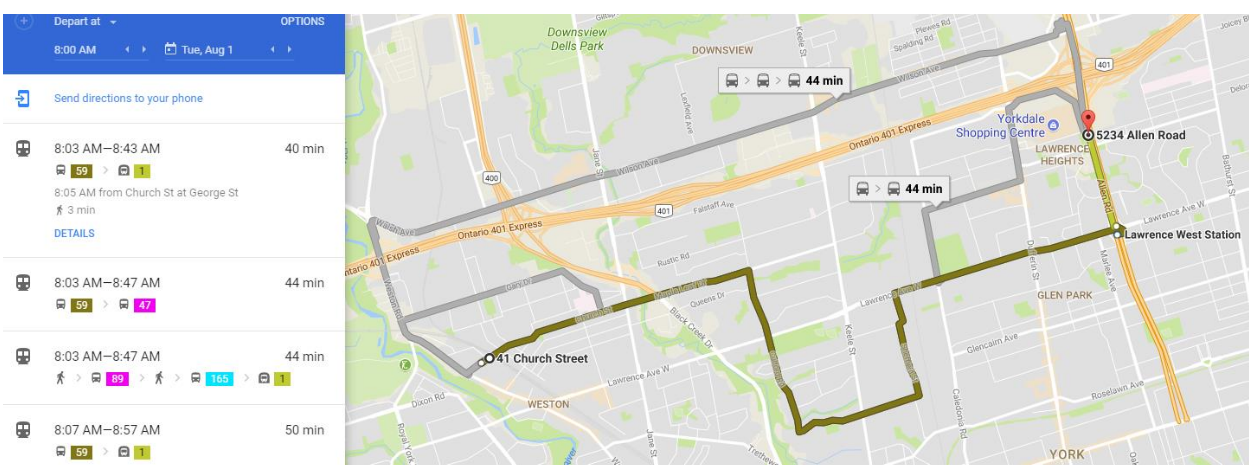
```
(133, 181)
['T052', 'T007']: 0.4479
['T089', 'T052', 'T007']: 0.1951
['T089', 'T165']: 0.12213
['T089', 'T096']: 0.11503
['T058', 'T109']: 0.0327439
['T058', 'T052', 'T007']: 0.0255236
['T059', 'T035', 'T096']: 0.0224349
['T059', 'T035', 'T165']: 0.018227
['T089', 'T058', 'T109']: 0.0116191
['T089', 'T058', 'T052', 'T007']: 0.00906035
['T089', 'T096', 'T096']: 0.00030252
['T089', 'T096', 'T165']: 0.00022967
['T089', 'T096', 'T160']: 5.33e-05
```

Fig. 4

However, this does not match with the observed paths for this OD pair (Figure 5), which involves taking bus route 59 (T059) to Lawrence West Station, and then taking the University-Spadina Subway (T593) north to Yorkdale Station. In fact, this path is nowhere to be seen on the list of Emme-generated paths.

133	181	W	W	SS47	SS48	0	0	T059	T593	0	0	0	0
133	181	W	W	SS47	SS48	0	0	T059	T593	0	0	0	0
133	181	W	W	SS47	SS48	0	0	T059	T593	0	0	0	0

Using Google Maps, the start location was shifted north from the original centroid in Zone 133, while the end location was moved west of the original centroid in Zone 181 (Figure 6). Both points remain in the same zones. The fastest route suggested by Google Maps now matches with the observed path.



This demonstration shows that for some OD pairs, the variability of start and end locations of a path within the zones, as well as the positioning of the centroid connectors, may be the reason why the Emme paths and observed paths do not match. A similar pattern has been observed for other OD pairs.

4.2. Numerical Analysis of All OD Pairs

A numerical analysis was performed on all OD pairs across the GTA to see if the pattern observed from the case study applies to other paths. The Emme-generated path file was the file that gave the highest percentage of observed paths that appeared in Emme-generated paths (See Section 1.3.1).

A Python program was used to count the total number of OD pairs, the number of OD pairs with no observed paths found in the Emme-generated paths, and the number of OD pairs which none of the segments in the observed paths appear in the Emme paths. The numbers are as follows (using the default parameters and settings from Section 1.5):

Total number of OD pairs	OD pairs with no obs. paths found	OD pairs with no segments found
17027	7608	1490
	44.68%	8.75%

P = 51.70%   S = 69.74%

These numbers show that for the OD pairs which no observed paths were found in the Emme-generated paths, most of these paths have at least some segments found in the corresponding Emme paths.

4.3. Numerical Analysis of All OD Pairs by Zone Region

4.3.1. Using the path file that gave the highest P and S values:

The same analysis above was performed for 3 groups of paths: Paths that start and end within Toronto, paths that start or end in Toronto, and paths that start and end outside of Toronto. This is to see if the density of the transit routes affect the correct modelling of path choices. The results are as follows:

All paths:

Total number of OD pairs	OD pairs with no obs. paths found	OD pairs with no segments found
17027	7608	1490
	44.68%	8.75%

P = 51.70%   S = 69.74%

Start and end within Toronto:

Total number of OD pairs	OD pairs with no obs. paths found	OD pairs with no segments found
12036	5386	839
	44.75%	6.97%

P = 52.65%   S = 72.29%

Start or end in Toronto:

Total number of OD pairs	OD pairs with no obs. paths found	OD pairs with no segments found
2305	1100	235
	47.72%	10.20%

P = 45.95%   S = 65.63%

Start and end outside Toronto:

Total number of OD pairs	OD pairs with no obs. paths found	OD pairs with no segments found
2686	1122	416

P = 51.50%   S = 59.83%

This set of data clearly shows that as the paths get away from the central area of Toronto, more observed paths have none of their segments found in the corresponding Emme-generated paths, but more entire paths are found instead. It shows that in areas where transit is sparser, Emme is more likely to get either the entire path correct, or completely wrong. This verifies the hypothesis that it is more difficult to model the paths in areas where transit is less dense.

4.3.2. Using the path file that produced the best fitness value (β = 0.1):

All Paths:

Total number of OD pairs	OD pairs with no obs. paths found	OD pairs with no segments found
17027	7244	1494
	42.54%	8.77%

P = 54.45%   S = 70.49%

Start and end within Toronto:

Total number of OD pairs	OD pairs with no obs. paths found	OD pairs with no segments found
12036	5063	844
	42.07%	7.01%

P = 56.09%   S = 72.80%

Start or end in Toronto:

Total number of OD pairs	OD pairs with no obs. paths found	OD pairs with no segments found
2305	1076	244
	46.68%	10.59%

P = 47.05%   S = 67.70%

Start and end outside Toronto:

Total number of OD pairs	OD pairs with no obs. paths found	OD pairs with no segments found
2686	1105	406
	41.14%	15.52%

P = 52.05%   S = 60.71%

These values are improved compared to the base case in 4.3.1. The same trend from above is observed in the different zone regions.

4.3.3. Using the path file that produced the best fitness value (β = 0.001):

All Paths:

Total number of OD pairs	OD pairs with no obs. paths found	OD pairs with no segments found

17027	7030	1339
	41.28%	7.86%

P = 56.05% S = 71.91%

Start and end within Toronto:

Total number of OD pairs	OD pairs with no obs. paths found	OD pairs with no segments found
12036	4885	712
	40.59%	5.92%

P = 57.82% S = 74.49%

Start or end in Toronto:

Total number of OD pairs	OD pairs with no obs. paths found	OD pairs with no segments found
2305	1082	246
	46.94%	10.67%

P = 46.97% S = 66.68%

Start and end outside Toronto:

Total number of OD pairs	OD pairs with no obs. paths found	OD pairs with no segments found
2686	1063	381
	39.58%	14.18%

P = 54.35% S = 62.75%

These values are better than those observed from 4.3.2, and further improved compared to the base case in 4.3.1. The same trend from 4.3.1 is observed in the different zone regions, which verifies the hypothesis from Section 3. This also shows that lowering the beta value can slightly improve the percent of observed paths modelled by Emme.

4.4. Analysis

Emme tries to compensate for these differences by generating all combinations of the path segments over the origin-destination. In order to create a match with the observed paths, Emme generates increasingly more and complex paths. The longer the distance, the more paths Emme generates, although a majority of those paths have extremely small proportions and can be neglected.

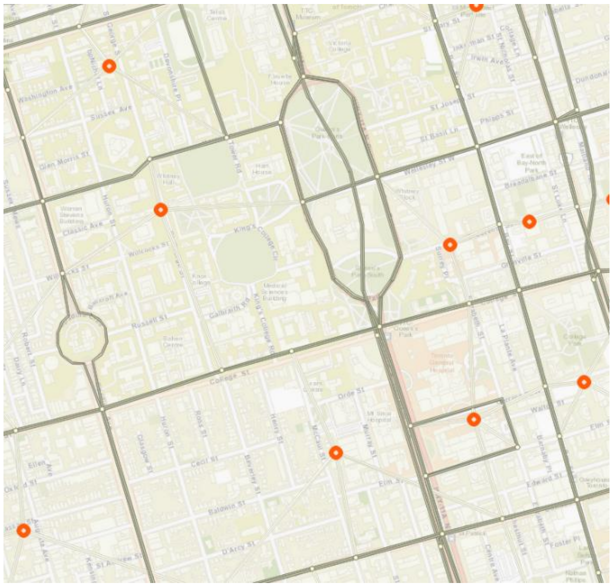
TTS data only showed start and end zones as well, not specific points. Some seemingly bizarre observed path choices are verified by Google Maps that the person may be travelling from one corner of a zone to the corner of another zone, instead of from centroid to centroid.

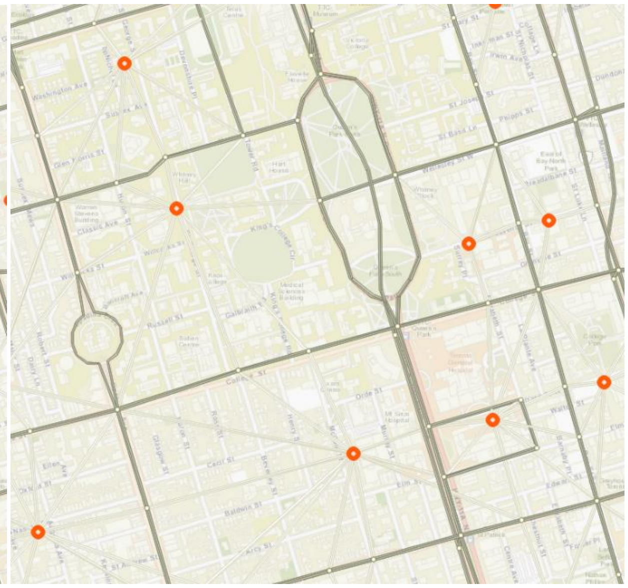
Also, since each centroid has only 4 centroid connectors to the nearby transit routes, it limits the paths that can be taken. To compensate for this, the number of centroid connectors can be increased.

4.5. Experiment: Increasing the number of centroid connectors

In this experiment, the number of centroid connectors per centroid is increased from 4 to 10 for those situated within the metropolitan area. The centroid connectors were generated by Emme. As expected, a noticeable improvement was observed. Using the default parameters and settings, P = 52.28%, S = 70.25%. This is slightly better than result of P = 51.70% and S = 69.74% from the original run (Section

4.2), which proves the hypothesis from above. However, these results are not as good as the best results from the parameter optimization runs (Sections 4.3.2, 4.3.3), which means that these results can be improved further if a parameter optimization was run on the new network with more centroid connectors. However, due to time constraints, this operation has not been performed.


Before


After

5. Conclusions

From all the analyses and experiments, several conclusions can be made so far:

1. The variability of start and end locations of a path within the zone area, as well as the positioning of the centroid connectors, reduces the chances that certain observed paths are modelled by Emme.
2. In areas where transit is less dense, Emme is more likely to get either the entire path correct, or completely wrong. Thus, it is more difficult to model the paths in areas where transit is sparser.
3. Increasing the number of centroid connectors would increase the probability that certain observed paths are modelled.
4. Lowering the beta value (as close to 0 as possible) in the path choice estimation can improve the percent of observed paths modelled by Emme, shown by the P and S values.
5. It is still inconclusive whether parameter optimization on a zone and centroid based network is sufficient to model the path choices.

6. Possible Next Steps

6.1. Improving the Emme Algorithm

6.2. Changing Centroid Connectors

A parameter optimization similar to the one shown in Sections 2.1 and 2.2 can be run using the network with more centroid connectors to improve the Emme path data.

Also, since all centroid connectors currently have a fixed length depending on their locations, this might affect how the paths are taken. Instead, the lengths of all centroid connectors can be set to zero to compensate for these possible differences.

6.3. Disaggregated Transit Assignment

Disaggregated transit assignment uses a coordinate system by assigning paths to go from point to point instead of from zone to zone. This can possibly model observed paths that currently cannot be found in Emme using zone and centroid based transit assignment.

However, there is not yet a tool in Emme Modeller (Emme Version 4.3) to run a disaggregated transit assignment. Currently, it can only be run manually from Emme Prompt. (Please read Section 5.35 on Emme 2 Manual to see how to do this)

6.4. Shrinking Zones

To increase the precision of the analysis, the size of the zones can be shrunk by breaking down the current zones into smaller zones and assigning centroids to each one of them.

However, this operation is difficult as it requires a lot of manual work changing the zones (Emme cannot do that, and the zones are defined by TTS), and it would require a lot more computation to process an increased number of zones, centroids, and connectors.

Appendix

1. Emme
https://www.inrosoftware.com/en/products/emme/

2. XTMF
http://www.ecf.utoronto.ca/~miller/TMG-XTMF-Documentation.pdf

3. TTS
http://www.transportationtomorrow.on.ca/

4. TTS Transit modes:

Code	Type	Description
C	*Auto*	Personal vehicle, any occupancy
E	*Auxilliary Auto*	Light/medium truck
F	*Auxiliary Auto*	Heavy truck
H	*Auxiliary Auto*	HOV2+ personal vehicle
I	*Auxiliary Auto*	HOV3+ personal vehicle
J	*Auxiliary Auto*	LOV (<2 or <3 depending on HOV definition used)
B	*Transit*	Local bus: 9m, 12m or articulated bus
G	*Transit*	Highway coach bus: GO Buses and intercity buses
L	*Transit*	LRT (light rail operated in exclusive right-of-way)
M	*Transit*	Subway
P	*Transit*	Premium bus service (not GO or intercity)
Q	*Transit*	BRT (bus on exclusive right-of-way)
R	*Transit*	Commuter rail
S	*Transit*	Streetcar (light rail operated in shared right-of-way)
A	*Auxiliary Transit*	Auto access to transit
K	*Auxiliary Transit*	Bicycle
T	*Auxiliary Transit*	Transfer between two lines for the same transit agency
U	*Auxiliary Transit*	Transfer between two transit agencies
V	*Auxiliary Transit*	Walk mode on centroid connector
W	*Auxiliary Transit*	Walk mode on road network link
Y	*Auxiliary Transit*	Walk from park & ride lot to transit station
X	*Unassigned*	Reserved for internal use

Currently unassigned: D,N,O,Z

5. TTS Zone Regions:

Zones	**Region**
1-1000	Toronto
1001-2000	Durham
2001-3000	York
3001-4000	Peel
4001-5000	Halton
5001-6000	Hamilton
6001-7000	Niagara
7001-8000	Waterloo
8001-9000	Brantford, Brant, Guelph, Wellington, Orangeville, Dufferin, Orillia, Simcoe, Barrie, Kawartha Lakes, Peterborough County, City of Peterborough

Source: http://www.dmg.utoronto.ca/pdf/tts/2011/dataguide2011.pdf

6. Map of TTC routes: https://ttc.ca/PDF/Maps/TTC_SystemMap.pdf

7. Multinomial Logit Model: https://en.wikipedia.org/wiki/Multinomial_logistic_regression

8. Default parameter values for parameter optimization:

Parameter	Default Value
Wait Time Perception	3.487829
Boarding Penalty Perception	1
Walk Perception Value: PD1	1.044555
Walk Perception Value: Toronto	1.193238
Walk Perception Value: Non-Toronto	1.00458
Walk Perception Value: Toronto Centroid Connectors	3.903624
Walk Perception Value: Non-Toronto Centroid Connectors	3.468211
Walk Perception Value: Subway	1
Fare Perception	12.33352
Boarding Penalty: Brampton	8.4
Boarding Penalty: Durham	2.516
Boarding Penalty: GO Bus	9.994
In Vehicle Time Perception: GO Train	0.8
Boarding Penalty: GO Train	0.534
Boarding Penalty: Halton	7.88
Boarding Penalty: Hamilton	0.288
Boarding Penalty: MiWay	1.44
Boarding Penalty: YRT	1
Boarding Penalty: VIVA	7.3
Boarding Penalty: Streetcar	4.74
Boarding Penalty: TTC Bus	3.75