Issues and Strategies Involved in Developing Agent-Based Multimodal Network Simulation Model for Transportation Planning: Lessons from a Case Study on the Greater Toronto and Hamilton Area

Adam Weiss, B.A.Sc
M.A.Sc. Candidate, Department of Civil Engineering
University of Toronto
Email: adam.weiss@mail.utoronto.ca

Mohamed S. Mahmoud, M.Sc.
PhD Candidate, Department of Civil Engineering
University of Toronto
mohamed.mahmoud@utoronto.ca

Peter Kucirek, M.A.Sc.
Research Associate, Department of Civil Engineering
University of Toronto
Email: peter.kucirek@mail.utoronto.ca

Khandker Nurul Habib, PhD, P.Eng.
Professor, Department of Civil Engineering
University of Toronto
Email: khandker.nurulhabib@utoronto.ca

Paper prepared for presentation
at the Best Practices in Transportation Planning Session

of the 2013 Conference of the
Transportation Association of Canada
Winnipeg, Manitoba

Abstract

The paper presents the issues and strategies involved in developing an agent-based multimodal network simulation model for the Greater Toronto and Hamilton Area (GTHA). The model was developed by using a Java-based open source simulation platform: MATSim. The issues and strategies presented utilize a geocoded automobile network and General Transit Feed Specification (GTFS) data of multiple transit agencies within the study area. While network simulation model for automobile network is common, an integrated multimodal network that combines auto and transit network (physical network and daily transit schedules) has not been developed for large study area, such as the GTHA by anyone. A key challenge is to integrate the GTFS data seamlessly in the multimodal framework. The GTFS data allowed meshing the auto network and the transit network together, creating a fully functioning multimodal network. The main challenge associated with this task is the determination of network resolution. The auto network is at times at too low of a resolution relative to the transit network, while the transit network often contained too much detail to be relevant for traffic simulation for a region as large as the GTHA. The paper presents guidelines and example of resolving these issues and overcoming the challenges.

1. *Introduction*

The Greater Toronto and Hamilton area (GTHA) is situated to the north west of Lake Ontario in the province of Ontario, forming Canada’s largest urban region. The GTHA’s current population is over 6 million, with projected growth to approximately 8.6 million by 2031 [1]. The GTHA, as shown in Figure-1, has eight local transit systems and a regional transit service operating under the administration of *Metrolinx,* a provincial government agency that was created to improve the coordination and integration of different modes of transportation in the region. Metrolinx, in their regional transportation plan for the GTHA, The Big Move [1], have identified a number of challenges facing the region as it continues to grow. The combination of population growth, auto-centric development, insufficient investment in transportation infrastructure and disconnected transit services providers as factors contributing to increased reliance on cars, which in turn creates congestion. These issues all indicate that an increase in transportation infrastructure and public transit spending are required such that the GTHA remain an economically competitive, attractive, vibrant and healthy region. This in turn raises the question of where resources should be allocated such that they will have the most benefit. This question highlights the importance of transportation planning and modelling in order to understand the trade-offs between potential infrastructure improvements.

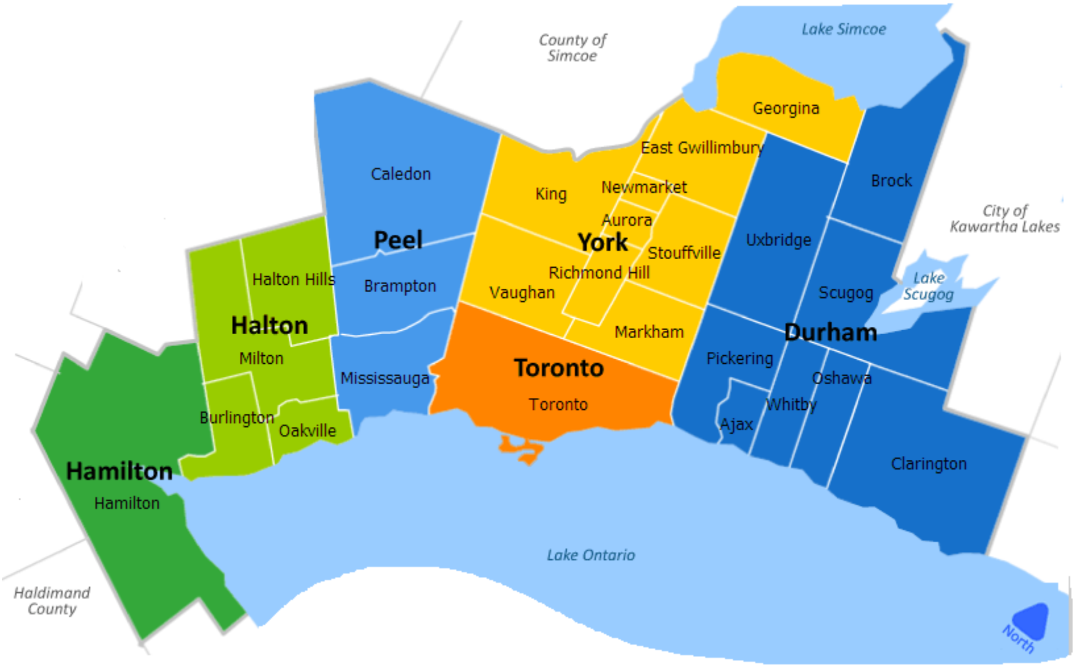

Figure-1 The Greater Toronto and Hamilton Area (GTHA) [1]

Traditional transportation planning approaches apply an aggregate four-stage model to mimic individuals’ travel choices within urban regions. These aggregate models allow for planners and engineers to answer simple questions about the impact that potential changes or polices might have on the transportation system. Traditional models are not capable of effectively testing complex policies, as these models do not provide sufficient level of detail. Therefore interest has grown in agent-based approaches, permitting more complex questions to be answered. These agent based approaches track individuals' trips with unique origins and destinations; a disaggregate approach which more accurately reflects population heterogeneity with respect to route choices.

Historically, the four-stage model is used by transportation planners to answer policy questions with respect to travel behaviour. This modelling procedure is limited by the use of aggregate zone based nature and is thus ill suited to answering detailed questions about specific policies. It is clear that by considering daily travel-activity chains or tours, it is possible to represent travel behaviour more accurately when compared to modelling peak period trips only. Daily activity-travel chains can easily be modelled in terms of chain of choices made by the travellers for different travel decisions, e.g. time of day choice, mode of travel choice, destination choice, trip purpose choice, etc. Such chain of choices can then be passed to a trip assignment tool in order to simulate individuals' route choices and thereby obtain trip level of service attributes for each trip during the course of day. In order to maximize the effectiveness of these activity based modelling approaches, the assignment procedure should not aggregate all these trips into zones, as is typically done for the four stage model. Therefore it is important to couple an agent-based simulation tool with a disaggregate activity based demand models. Such an integrated framework could be used to test and evaluate potential policy measures such as: changing pricing scenarios (e.g., congestion pricing and increasing transit fares), introducing new transit services and improving station facilities (e.g., parking lots, ease of transfer, bike stands, etc.).

As of 2010, the MATSim platform, an open source dynamic agent-based assignment tool, has provided support for joint transit and automobile assignment and with the advent of modern desktop computers and the ability to simulate a small percentage of traffic while still producing useable disaggregate results; the assignment procedure becomes more reasonable from a computational standpoint [2]. These features, coupled with an existing auto only MATSim network for the GTHA, prompted the exploration into the feasibility of developing a multimodal agent-based dynamic traffic simulation framework for the region.

For a region as large as the GTHA, obtaining schedule data for all the different transit agencies that operate within the region can be a challenge. This information is essential for performing multimodal assignment as it provides the details of not only where transit stops are located, but also when transit vehicles will visit these stops and in what order. Fortunately, the advent of Google Transit Feed Specification (GTFS) data, a common format for Transit agencies to publish their schedules for developers solves the problem of divergent data sources.

The Remainder of this paper will address existing applications of GTFS data, as well as a number of multimodal development projects, the methods that were undertaken to integrate the GTFS data with the geocoded EMME [3] automobile network for the GTHA.

2. Literature Review

The literature review focuses on existing applications of GTFS data, with a particular focus on applications specific to multimodal network development. In particular, a discussion on using GTFS within the MATSim framework will be presented, looking at both the cases of Singapore and the City of Toronto.

2.1. Traditional GTFS Uses

The unified structure of the GTFS data has been used by mobile and internet applications developers to create trip planning tools for transit riders. Keynote examples of such applications include Mapnificent [4] and OneBusAway [5]. These types of applications are attractive for two reasons: the GTFS data is open source and it has a standardized data format, which makes it readily available for many municipalities. Such applications are universal and can be used for any region by updating the application database with GTFS data for the new region.

2.2. Network Modelling Approaches

Despite the traditional uses of GTFS data, other applications exist, particularly in the field of multimodal network development. Through map matching procedures, transit stop locations (obtained through GTFS or other data sources) can be spatially matched to existing automobile-only bi-directional geocoded networks. This allows for both transit and automobile assignment to be performed concurrently, thereby capturing the interaction between transit and automobile modes during assignment.

The map-matching tool developed by Ordonez and Erath [6], was created to match bus line stop locations to a high-resolution geocoded network. This work focuses specifically on a bus only transit service and a high-resolution network for the City-State of Singapore. It should be noted that this approach was used for the MATSim Singapore network development and therefore was initially very attractive for implementation with the GTHA MATSim network, however a number of GTHA specific problems with this approach were identified. In the case of the GTHA, the use of regional rail services, streetcars, light rail transit (LRT) and subway services, makes the methods presented by Ordonez and Erath difficult to apply properly. Furthermore, because of the lower resolution of the auto only network in the GTHA, many of the methods and specific issues related to the high-resolution map matching procedure were inapplicable in the case of the GTHA. Nevertheless, their work identifies key issues associated with combining an auto only network with transit services through the use of a map matching approach. Furthermore, much of the script code written to process GTFS data for MATSim in Singapore was applied for the GTHA MATSim network.

Similar to the work of Ordonez and Earth, the work of Li [7] examines a procedure to match bus stops to road links in a geocoded auto network. Rather than examining each stop individually, the map matching procedure outlined by Li looks at a collection of sequential stops within a route and uses a shortest path algorithm selecting the links that are most likely to be used by the transit vehicle to travel between successive stops. This approach is effective for map matching surface transit services with short distances between stops, to a high-resolution network, where multiple candidate links are in the vicinity for each potential stop. Despite the effectiveness of this approach, it is not applicable for map matching in the GTHA; however, the approach presented by Li present a solid foundation for understanding the map-matching problem.

Kucirek [8] presented an initial attempt and discussion on the development of a multimodal network for the GTHA. At the time the work was conducted, only the 2006 auto planning level network was available. Despite this, GTFS data from March 2012 for the TTC was used to create a proof of concept that full multimodal assignment could be performed within the GTHA. This allowed the existing GTFS schedule generator developed for the Singapore MATSim assignment procedure to be used once the stop map matching had been completed. Because this work was developed as a proof of concept and a comparison between MATSim and the more traditional EMME assignment, little to no network resolution editing was performed prior to the actual map matching resulting in a simplified network.

3. Data Sources

3.1. EMME/MATSim Auto Network

The existing MATSim auto network was imported from a geocoded planning-level network developed for the City of Toronto [3] for use with the EMME traffic assignment software package. The EMME and the MATSim networks consist of links and nodes, which are a virtual representation of the physical road infrastructure. Each node is associated with an identification number as well as a coordinate value and each link is associated with an identification number, a “from” node, a “to” node (representing the start and end nodes of each link), a length, a free flow speed, a capacity, and a list of modes which are permitted to use that link.

3.2. GTFS Data

General Transit Feed Specification is a standardized format for organizing public transport schedules and transit infrastructure geographical location. This standardized data format allows multiple transit agencies to publish their data such that developers can create streamlined applications, which use these standardized data sets. In the case of the GTHA, GTFS datasets were obtained for the City of Toronto, the Region of York, the City of Mississauga, the City of Brampton, the City of Hamilton and the GO Regional Transit Service, operated by Metrolinx. In cases where the transit authorities made their GTFS data publicly available, the GTFS data were retrieved from a GTFS database [9], otherwise, the feeds were obtained by request from the transit agency. Three transit agencies, namely, Durham Region Transit, Oakville Transit, and Burlington Transit, did not have GTFS data available at the time of writing and therefore these services were omitted from this study.

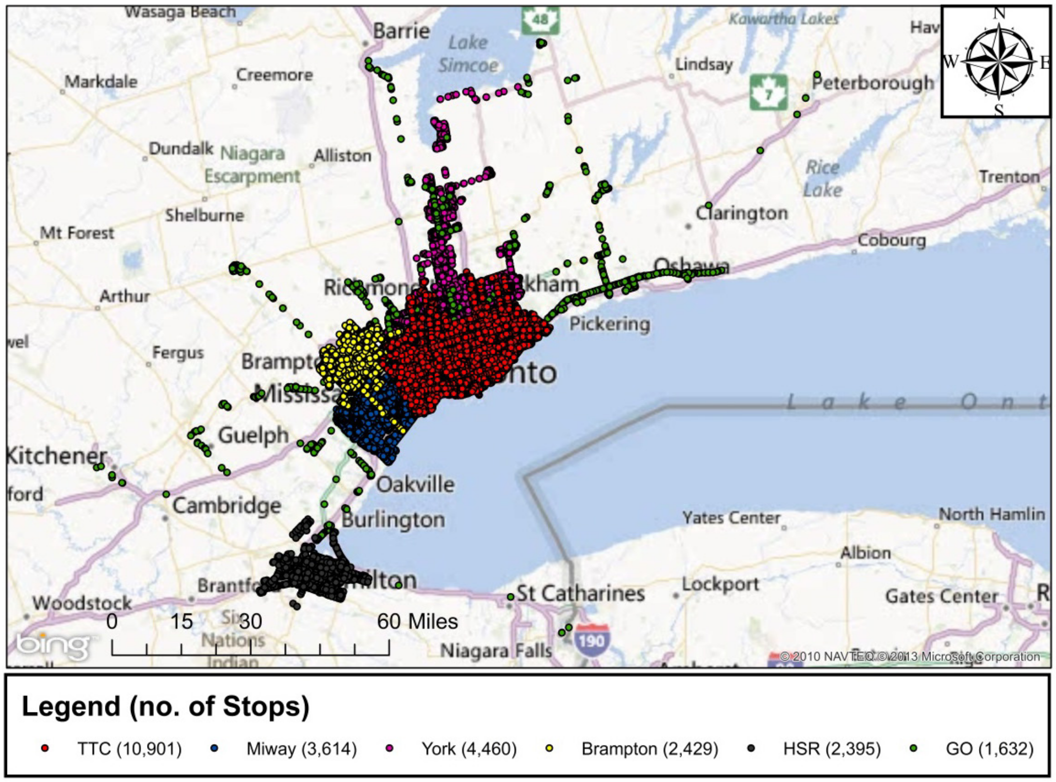

Figure-2 – GTFS Transit Stop Locations Within the GTHA

4. *Problem Definition:*

Figure-2 shows all the GTFS transit stop locations as well as the number of stops for each operator. In total there are 25431 stops split between all 6 operators. This section will outline the issues that arise from integrating these stops with the existing network such that a realistic multimodal assignment can be performed.

Figure-3 shows a schematic diagram of the network representation. The existing geocoded network, which is at a lower resolution compared to the actual road infrastructure, is shown in Figure-3-(a).The key challenge in matching GTFS data with the existing planning-level network was a difference in spatial resolution; the location of transit stops are defined more precisely in the GTFS data relative to the nodes in the geocoded network. Existing geocoded nodes represent major intersections while GTFS transit stops are shown at their exact locations, which are not necessarily located at major intersections.

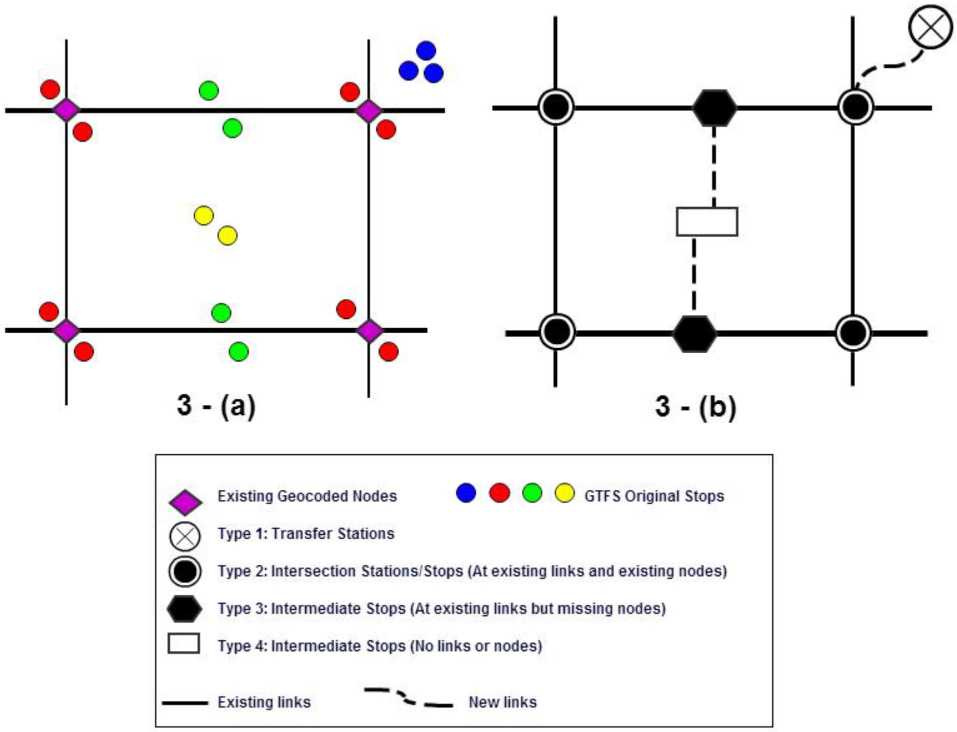

Figure-3 – Schematic of GTFS Map Matching Problem and Proposed Hybrid Network Solution

Matching the resolution between the GTFS stops and the geocoded nodes is important to ensure that the simulation functions properly because MATSim only permits passengers to access or egress a transit vehicle at the end of links. If transit stops are not located at the end of a link, they will be skipped during the assignment procedure. To avoid this, three alternative solutions were proposed:

1. Modify the geocoded network by increasing its resolution to match the resolution of the GTFS data;
2. Modify the GTFS data by removing stops to match the lower resolution of the geocoded network; and
3. Reference multiple GTFS stops to their corresponding geocoded network links without modifying the network resolutions.

Alternative three was originally attempted by Kucirek [8] with limited success and therefore was not considered. On the basis that it is less onerous to decrease the resolution of a detailed network compared to the work involved in increasing the resolution of an aggregate network, alternative one was eventually selected. Therefore, a map matching procedure is needed to represent both geocoded nodes and GTFS data in a hybrid network as shown in Figure-3-(b). This hybrid network will contain geocoded nodes which are proximate to the location of each

GTFS transit stop. This will facilitate the mapping of each GTFS transit stop to a node in the hybrid network and by extension a link terminating at that node.

In order to facilitate the generation of this hybrid network, four GTFS data types were identified and spatially grouped to ease the matching procedure. Type one stops, which refer to surface terminals or interconnected transfer points between different transit lines, are represented by a cluster of stops, with each stop representing a single bus bay. Type two stops are surface (bus and streetcar) transit stops which are located around nodes that are already present in the geocoded network. Type three stops are surface (bus and streetcar) transit stops which are located along an existing network link that is present in planning level network however the link does not have nodes segmenting it at the location of actual stops. Type four stops are surface (bus and streetcar) transit stops which are located along road segments that are not defined in the planning-level network. These road segments include new residential subdivisions or minor local roads with transit services which were not coded into the planning-level network. Type two, three and four stops were occasionally found clustered together, with each stop in a cluster representing a stop for one direction of travel.

Moving from the original geocoded network, as in Figure-(3-a), to a hybrid GTFS/geocoded network, as in Figure-(3-b), increases the network resolution for types one, three and four stops by adding new links or/and modifying exiting links within the geocoded network, while maintaining the existing sufficient resolution for type two stops. Figure-4 shows a sample of the original geocoded links, nodes and a sample of GTFS stops labeled by their type and Figure-5 shows a section of the hybrid network after matching the GTFS stops to the geocoded nodes. As shown in Figure-5, all nodes are connected to the hybrid network links except transfer stations which are connected to the network via transit-only links.

5. Map Matching Methods

An overview and discussion of the procedure used to obtain the hybrid network is presented in section 5.1. The procedure is designed to minimize manual editing by grouping and matching station based on spatial proximity, which facilitates visual inspection. A detailed explanation of each step taken in the procedure is outlined in section 5.2.

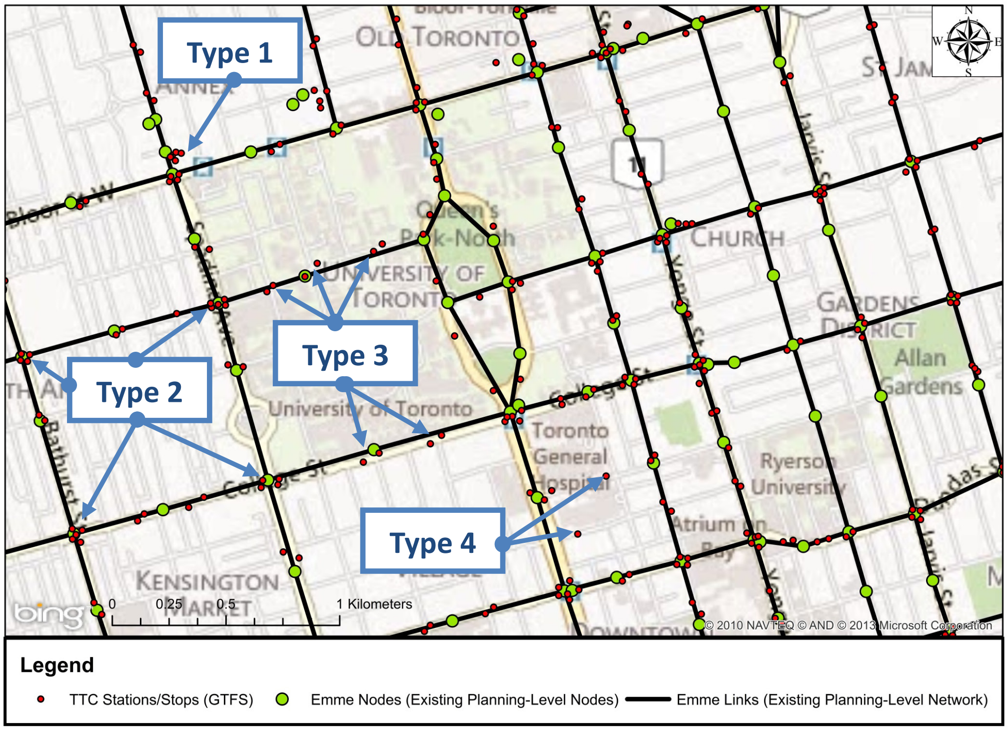

Figure-4 Example of Each Stop Classification

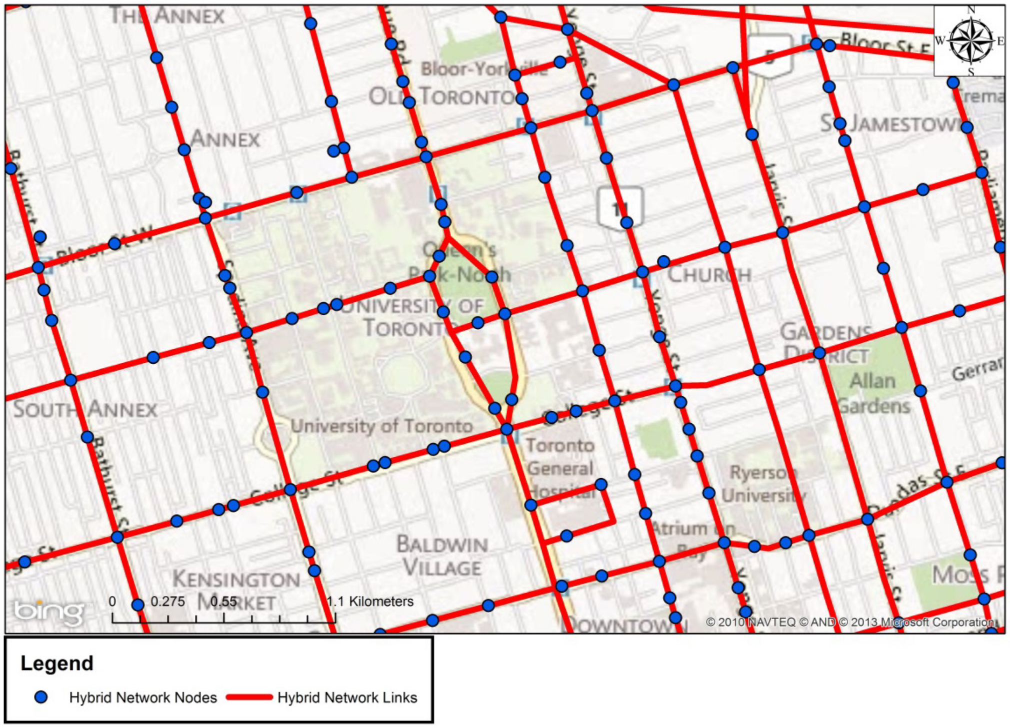

Figure-5 Final Hybrid Network

5.1. Procedure and Discussion

Before the map matching procedure was performed, a set of particular GTFS stops were identified and excluded from the semi-automated procedure to be dealt with on a case-by-case basis. This set includes transfer stations (type one) and stops along transit lines with an exclusive right of way, express routes stops or night transit stops. The procedure for type one stops was to group them into stop groups based on special proximity, and then based on the location of the stop group (the centroid of all stops in the stop group), create a special transit only node with transit only links that connect the transfer node to the rest of the network. Stops along routes with an exclusive right of way such as subways or regional trains were provided with their own separate network. This was done in order to prevent traffic on local roads interfering with the operation of these transit lines. The separate right of way networks were overlaid onto the multimodal network. Express routes were excluded because the stop spacing was at lower resolution then the existing planning level network. In this case, the work developed by Ordonez Erath [6] for use on the highly detailed Singapore MATSim network was employed to assign the express route stops to the network. Stops servicing night transit service were also removed, as these services would often create complications in the development of the transit schedule for assignment due to irregular transit headways. Although potentially feasible to include these stops at a later date, they fall outside of the scope of the plans for the network in the near future.

The remaining stops were aggregated into stop groups based on their spatial proximity as represented in Figure 4. Each stop group was first snapped to a corresponding planning-level network node if it was within a specific distance. These stops were classified as type two. All remaining stop groups were then snapped to a corresponding planning-level network link if the perpendicular distance between the stop group and link was within a specific distance. These stops were classified as type three. Any remaining stops that were not snapped to the network were flagged as being type four stops. These stops were handled manually by adding new nodes and links to the network. The remaining stop groups were then snapped to the new network. Links with stop groups along their edge were split at the location of these stop groups and new nodes were added. The combination of the original planning level network, the stop groups coded to existing nodes, the new nodes added as a result of splitting links and the new nodes and links added to accommodate type four stops form the basis of the hybrid network.

The hybrid network now contains nodes, links and stop groups. The stop groups have been snapped to links as seen in figure 5. Within each stop group is a set of GTFS stops, each of which in turn needs to be mapped to the end of a link. In cases where multiple links terminate at a single node a procedure must be developed to determine the correct link to assign each stop. For example in figure 6-(a) the node has four links terminating at it and three stops within that nodes corresponding stop group. As can be seen in figure 6-(b), the north south and westbound links each have a stop associated with them, whereas the eastbound link does not have a stop associated with it. This results in the final required formulation seen in figure 6-(c). Because it is infeasible to determine this by visual inspection for all GTFS stops, an automated procedure was developed. This procedure involved determining the stop sequence of each transit line and then using a modified version of the Dijkstra's shortest path algorithm for the MATSim framework [10 & 11] to find the links between nodes whose stop groups have two consecutive stops in a transit line. The second stop in the pair of consecutive stops was then matched to the final link in the path between the two stops. This procedure was performed for all transit lines in the GTFS schedule data such that all of the GTFS transit stops were assigned to the end of a link.

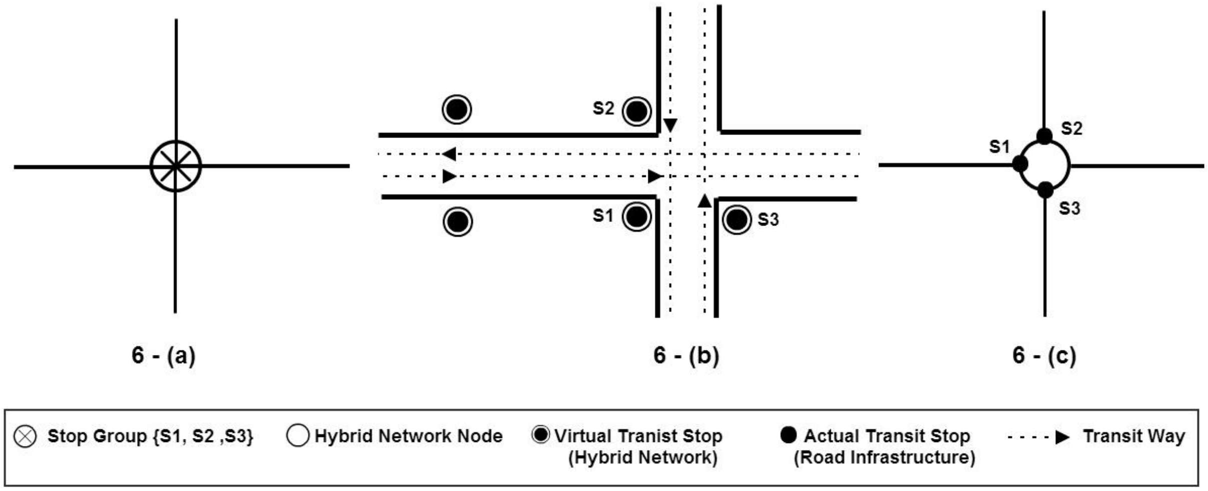

Figure-6 The Problem of Assigning Stops to Links

Fortunately, the link to stop mapping and the link path between consecutive stops is information that is explicitly required as an input in a schedule file for MATSim to perform transit assignment and thus once this information had been extracted, it was used to generate the MATSim schedule file.

5.2. Detailed Procedure

Each detailed step presented here results in a new or modified geometry, which can be visually and analytically checked against the prior dataset to validate the results from the previous step.

1) Stops serving exclusive ROW lines – in this case subway and commuter rail lines – were filtered from the set of stops. They were grouped by proximity.
2) From the remaining surface stops, Type 1 stop groups were identified and grouped manually, and then removed from the set of ungrouped stops. Type 1 stops were identified through a combination of methods including examining stops in close proximity to subway or rail stations as well as large points of attraction such as shopping malls. These stops were put aside to be added to the network during a later step.
3) The remaining ungrouped stops for each transit agency were grouped by proximity, based on special proximity. It worth mentioning that we tested different tolerance buffers for each area based on the density of stops and their spacing. For example, the TTC had a higher stop density and therefore a lower distance between stops was used when compared to Brampton.

4) All of the different transit agency stop groups were combined together and then another grouping was performed to combine stop groups from different agencies, which use the same or proximate infrastructure.
5) The nearest network node to each stop group was located. Stop groups whose nearest network node was within 50 meters were identified as type two stops and snapped to that node. Stop groups whose nearest network node was more than 50 meters away were classified as either Type 3 or Type 4.
6) The nearest network link to the remaining type three or four groups was located. If the perpendicular distance between the stop group and the link was within 150 meters that stop group was classified as Type 3. If the perpendicular distance between the link and the stop group was greater than 150 meters, then the stop group was classified as type 4.
7) Type three stop groups were snapped to the nearest network link. The link was split at the location of the snapped stop group and a new node was added at this location.
8) Type 4 stop groups required manual editing of the network, since many of them corresponded to new network links. The speed limit and number of lanes of each new link were identified using Google Maps and Street View; most links were local streets with speed limits or either 40 or 50 km/hr. Link per-lane capacities were based on adjacent links in the same regional municipality with similar properties.
9) During the manual editing process, it was occasionally necessary to expand certain stop groups to accommodate logical nearby stops which were not identified by the automated procedure. It was also occasionally necessary to re-classify some groups as Type 1, typically smaller bus-bays serving smaller shopping malls or local health centres.
10) A second pass of the type classification procedure was performed to map each stop group to its corresponding intersection.
11) For Type 1 stop groups, a network node was created at the stop group centroid, and connected via a two-way link to an appropriate nearby network node; thus representing a network loop.
12) Stop groups were subdivided into sets of stops with each stop to be mapped to the one of the links ending at the stop's stop group's node. Based on the shortest path between consecutive stops in a transit line, each stop was assigned to the final link between it and the previous stop on the path of the transit vehicle.
13) Finally, express routes were processed using MATSim's existing module for GTFS mapping, which allows large routes to be adjusted in detail one-at-a-time. For more details on this procedure, see [5] and [7].

6. Conclusion

This paper examines the issues and challenges associated with performing map matching of high resolution transit stop locations to a lower resolution planning level network. This map matching procedure uses a transit stop classification system based on spatial proximity to both other stops and the network. This work extends on existing map matching procedures and develops a procedure to expand lower resolution networks such that they are fully able to accommodate high-resolution transit data using standard map matching procedures. The outcome of this procedure resulted in both a methodology for similar cases of matching high-resolution coordinates to a low resolution network as well as a fully functioning multimodal network ready for simulation. The next section will discuss future plans for accomplishing this simulation

7. Future Work

In order to perform a multimodal traffic assignment, a sample population is required. Based on other work done with multimodal MATSim assignment for the City-State of Singapore [9] a 25 percent expanded sample population is appropriate in order to minimize cases of bus overcrowding compared to traditionally smaller percent sample used for auto assignment (in the case of the GTHA, a 5% sample is used). The bus-overcrowding problem occurs when the capacity of the transit vehicle is reduced to reflect the population sample size, which can cause a single extra rider to result in the bus exceeding capacity when it should not.

A 24-hour trip diary for the sample population of the GTHA was extracted from the 2006 Transportation Tomorrow Survey (TTS) dataset, which was collected by the Data Management Group (DMG) at the University of Toronto [12]. TTS is a trip-based household survey conducted every five years in the Greater Toronto and Hamilton Area (GTHA) among 5 percent of its population. The latest TTS survey datasets available is the 2006 TTS survey dataset, which will be used for this study. Once the 2011 data is available, trip diaries can be simply updated accordingly.

The discussed map-matching procedure is a main step in the development of a multimodal network for the GTHA. However, three of the GTHA transit operators did not have GTFS data available at the time of writing and therefore these services were omitted from this study. This omission creates a gap within the study area. For that reason, the map-matching procedure was designed in a generic way to accommodate future extensions and ensure universal transferability to other study areas.

The developed network will be validated using traffic screen line counts obtained from the Data Management Group at the University of Toronto [12]. Then, an extensive network cleaning process will follow. For instance, the process includes identification of “bus-pocket” locations or road infrastructure where bus stopping does not impede traffic. These stop locations will be flagged such that during assignment, the busses will not impede the flow of other vehicles.

An advanced transit fare calculator tool is currently under development. This tool is capable of calculating fares for trips that involve using one or more transit modes, which can be operated by different agencies. In addition, the fare calculator accounts for fare integration polices such as discounted fares for connecting trips.

8. Acknowledgement

This research was partially funded through an NSERC Discovery Grant and the MEDI Early Researcher Award that are awarded to the last author.

9. References

[1] Metrolinx. (2008). The Big Move: Transforming Transportation In The Greater Toronto And Hamilton Area: Draft regional transportation plan.

[2] Erath, A., Fourie, P., van Eggermond, M., Ordonez, S., Chakirov, A., & Axhausen, K. (2012). Large-Scale Agent-Based Transport Demand Model For Singapore. Working Paper.

[3] Travel Modelling Group. (2013). EMME 2011 Network.

[4] Cawley, J. (2011, Jun 20, 2011). Online Application Puts Hampton Roads On Map. Daily Press, pp. A.4.

[5] Ferris, B., Watkins, K., & Borning, A. (2009). OneBusAway: A Transit Traveler Information System. Proceedings of Mobicase 2009, San Diego California USA. 92-106.

[6] Ordonez, S., & Erath, A. (2011). Semi-Automatic Tool For Map-Matching Bus Routes On High-Resolution Navigation Networks. 16th International Conference of Hong Kong Society for Transportation Studies, Hong Kong.

[7] Li, J. (2012). Match Bus Stops To A Digital Road Network By The Shortest Path Model. Transportation Research Part C: Emerging Technologies, 22, 119-131.

[8] Kucirek, P. (2012). Comparison between MATSim & EMME: Developing A Dynamic, Activity-Based Microsimulation Transit Assignment Model For Toronto. (Master Of Applied Science, University of Toronto).

[9] Czebotar, J. (2013). GTFS data exchange

[10] Dijkstra, E. W. (1959). A Note On Two Problems In Connexion With Graphs. Numerische Mathematik, 1(1), 269-271.

[11] Balmer, M., Edelhoff, T., Mohring, R. H., & Schilling, H. (2006). Optimal route assignment in large scale micro-simulations. (). Institute of Mathematics, TU Berlin, Germany.

[12] Data Management Group (DMG). (2008). Transportation Tomorrow Survey. Data Management Group.