Dynamic Modelling of Transit Operations and Passenger decisions

Oded Cats

Doctoral thesis in Transport Science with specialisation in Transport systems

KTH – Royal Institute of Technology
Department of Transport Science
Division of Transport and Logistics
Centre for Traffic Research

December 2011
Acknowledgements

This thesis summarizes my PhD research in the dual-doctoral program of the Technion-Israel Institute of Technology and KTH – Royal Institute of Technology, Sweden. The seed of this cooperation was the common interest of Tomer Toledo from the Division of Transportation and Geo-Information Engineering at the Technion and Haris Koutsopoulos and Wilco Burghout from the Division of Transport and Logistics, KTH in developing a transit operations simulation model. This thesis was supervised by them.

I would like to express my gratitude to my supervisors. Tomer is my mentor in the academic life since I have started my transport studies. The combination of sharp critic and sincere interest and support encouraged me to further pursue and realize my research interests. Haris devotion and insightful comments provoked me to tackle challenging issues. His focus on sound methodology along with practical considerations contributed to a wider research perspective. Wilco guided me into the traffic simulation modeling details and Mezzo code in particular. His willingness to have discussions in unordinary conditions are appreciated.

The financial support of the Technion during the first year of my studies is greatly acknowledged. My employment at the centre for traffic research (CTR) at KTH was funded by DYMOBUS project. The project was financially supported by VINNOVA (the Swedish governmental agency for innovation systems) and the traffic department of the city of Stockholm. SL, the regional transit authority, supported this project in kindness with data and advice. Their support is highly appreciated.

I specially want to thank Ingmar Andreasson, Shlomo Bekhor, Haim Aviram, Kalle Kottenhoff and Avi Ceder who have expressed genuine interest in my work during these years. I would like to thank my dear colleagues from the Technion and KTH for the stimulating and pleasant working environment. I am thankful for the fruitful discussions I had with people from the transport research and industry community in Sweden and Israel. The dual-doctoral program allowed me not only to experience different research environments, but moreover to meet many colleagues and friends that I would always cherish.

I am grateful to my parents support. Thank you for encouraging me to pursue my curiosity and realize my dreams. Last but not least, Oz, my love. Thanks to your companion and support, the walk of life is always worthwhile.
List of Publications

Parts of this thesis have been previously published in the form of journal publications and scientific conferences:

Journal Publications:
Cats O., Burghout W., Toledo T. and Koutsopoulos H.N. (2010), Mesoscopic modeling of bus public transportation. Transportation Research Record, 2188, 9-18.

Refereed Conference Proceedings:

In addition the following papers were submitted for publication:
Cats O., Burghout W., Toledo T. and Koutsopoulos H.N. Modeling real-time transit information and its impacts on travelers’ decisions. Submitted to Transportation Research Record, 2012.
Contents

Abstract .............................................................................................................................................. 1
1. Introduction .................................................................................................................................... 2
  1.1 Transit systems and Advanced Public Transport Systems (APTS) ........................................... 2
  1.2 Transit modeling spectrum ........................................................................................................ 3
  1.3 Transit simulation models ......................................................................................................... 6
  1.4 Research objectives and approach ......................................................................................... 7
  1.5 Thesis outline .......................................................................................................................... 8
2. Literature review of transit assignment and simulation models ................................................. 9
  2.1 Conventional transit assignment models .................................................................................. 9
    2.1.1 The transit assignment problem ...................................................................................... 9
    2.1.2 Frequency-based assignment models ............................................................................. 11
    2.1.3 Schedule-based assignment models .............................................................................. 14
  2.2 Transit simulation models ....................................................................................................... 17
    2.2.1 Background .................................................................................................................... 17
    2.2.2 Transit operations simulation models ............................................................................ 18
    2.2.3 Transit assignment simulation models ............................................................................ 19
  2.3 Transit path choice modeling ................................................................................................... 22
    2.3.1 Choice modeling ............................................................................................................ 23
      2.3.1.1 The role of choice-set and choice model classification ............................................ 23
      2.3.1.2 Non-compensatory one-stage models .................................................................... 24
      2.3.1.3 Compensatory one-stage models .......................................................................... 25
      2.3.1.4 Semi-compensatory two-stage models ................................................................. 26
      2.3.1.5 Compensatory two-stage models .......................................................................... 27
    2.3.2 Choice-set generation models .......................................................................................... 29
      2.3.2.1 Choice-set generation model properties ................................................................. 29
      2.3.2.2 Transit choice-set generation models ..................................................................... 30
    2.3.3 Random utility discrete choice models ............................................................................. 32
      2.3.3.1 Discrete choice models properties ......................................................................... 32
      2.3.3.2 Transit path choice models .................................................................................... 33
  2.4 Synthesis and modeling issues ................................................................................................ 35
3. Dynamic transit model framework ............................................................................................. 39
  3.1 Model components ................................................................................................................... 39
  3.2 Traffic dynamics ...................................................................................................................... 40
5. Transit path choice model estimation........................................................................................................106
  5.1 Questionnaire design.......................................................................................................................................106
    5.1.1 General information................................................................................................................................110
    5.1.2 Chosen transit path .................................................................................................................................110
    5.1.3 Transit consideration-set .......................................................................................................................112
    5.1.4 Transit path choice..................................................................................................................................113
    5.1.5 Transfer stop choice................................................................................................................................119
  5.2 Summary statistics ........................................................................................................................................119
    5.2.1 General information................................................................................................................................119
    5.2.2 RP transit path ..........................................................................................................................................122
    5.2.3 SP transit consideration-set ...................................................................................................................125
    5.2.4 SP transit path choice ............................................................................................................................133
    5.2.5 SP transfer stop choice..........................................................................................................................134
  5.3 Transit choice set generation model .............................................................................................................136
  5.4 Path utility function estimation.....................................................................................................................143
    5.4.1 Model specification and estimation results .............................................................................................144
    5.4.2 Discussion..............................................................................................................................................147
6. Evaluation of holding control strategies .......................................................................................................150
  6.1 Background and related studies ....................................................................................................................150
  6.2 Case study I: bus line 51, Tel Aviv ...............................................................................................................154
    6.2.1 Experiment description ............................................................................................................................154
    6.2.2 Analysis of the current conditions and a validation ..............................................................................157
    6.2.3 Scenario design .......................................................................................................................................160
    6.2.4 Assessment of holding strategies ..........................................................................................................161
  6.3 Case study II: bus line 1, Stockholm ............................................................................................................168
    6.3.1 Line description .......................................................................................................................................169
    6.3.2 Analysis of the current performance ....................................................................................................170
    6.3.3 Experiment description ..........................................................................................................................177
    6.3.4 Results and discussion ..........................................................................................................................181
    6.3.4.1 Service regularity .................................................................................................................................181
    6.3.4.2 Operational considerations ................................................................................................................181
    6.3.4.3 Robustness ..........................................................................................................................................187
    6.3.5 From simulation to a field study ............................................................................................................191
  6.4 Optimizing the number and location of time point stops ............................................................................199
    6.4.1 Related work ..........................................................................................................................................199
6.4.2 Problem formulation ................................................................. 200
6.4.3 Greedy algorithm ................................................................. 201
   6.4.3.1 Description .................................................................. 201
   6.4.3.2 Results and discussion ............................................. 202
6.4.4 Genetic algorithm ................................................................. 205
   6.4.4.1 Description ................................................................ 205
   6.4.4.2 Results and discussion ............................................. 207
6.4.5 Conclusions ........................................................................ 211
6.5 Interaction of holding control strategies and boarding regimes on a common corridor. 216
7. Impacts of real-time information .................................................. 218
   7.1 Potential benefits of real-time information provision .............. 218
   7.2 Modeling real-time information in BusMezzo ...................... 220
      7.2.1 Modeling approach .................................................... 220
      7.2.2 Generation of real-time information ......................... 222
      7.2.3 Real-time information and traveler decisions ............. 225
7.3 Case study I: Stockholm metro network .................................. 228
   7.3.1 Description and design .................................................... 228
   7.3.1.1 Experiment description ............................................. 228
   7.3.1.2 Scenarios design .......................................................... 229
   7.3.1.3 Choice specification ................................................ 231
   7.3.2 Results and discussion ................................................. 232
7.4 Case study II: Stockholm rapid transit network ...................... 238
   7.4.1 Description and design .................................................... 238
   7.4.2 Results and discussion ................................................. 240
7.5 Conclusions ............................................................................. 244
8. Discussion .................................................................................... 246
   8.1 Research contribution ........................................................ 246
   8.2 Limitations and further research ........................................... 248
References ....................................................................................... 251
Appendix A: Object-Oriented Framework ........................................... 274
Appendix B: Inputs and outputs of the simulation model .................. 275
Appendix C: Screenshots from the web-based survey ...................... 278
### List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Choice model classification</td>
<td>24</td>
</tr>
<tr>
<td>3.1</td>
<td>Transit model components</td>
<td>40</td>
</tr>
<tr>
<td>3.2</td>
<td>Mezzo GUI screen</td>
<td>42</td>
</tr>
<tr>
<td>3.3</td>
<td>Dynamic assignment procedure in Mezzo</td>
<td>45</td>
</tr>
<tr>
<td>3.4</td>
<td>Dwell time determinants</td>
<td>47</td>
</tr>
<tr>
<td>3.5</td>
<td>Illustration of fleet management schemes</td>
<td>54</td>
</tr>
<tr>
<td>3.6</td>
<td>Flowchart of the transit simulation process (simplified)</td>
<td>62</td>
</tr>
<tr>
<td>3.7</td>
<td>Transit simulation model inputs and outputs</td>
<td>65</td>
</tr>
<tr>
<td>4.1</td>
<td>Two-stage modeling approach</td>
<td>67</td>
</tr>
<tr>
<td>4.2</td>
<td>Path choice decision process</td>
<td>70</td>
</tr>
<tr>
<td>4.3</td>
<td>Illustration of transit path alternative components</td>
<td>72</td>
</tr>
<tr>
<td>4.4</td>
<td>Flowchart of the choice-set generation model</td>
<td>76</td>
</tr>
<tr>
<td>4.5</td>
<td>Recursive search illustration</td>
<td>77</td>
</tr>
<tr>
<td>4.6</td>
<td>Stop choice decision structure</td>
<td>92</td>
</tr>
<tr>
<td>4.7</td>
<td>Alternative stop choice decision structures</td>
<td>94</td>
</tr>
<tr>
<td>4.8</td>
<td>Boarding decision structure</td>
<td>95</td>
</tr>
<tr>
<td>4.9</td>
<td>Alighting decision structure</td>
<td>96</td>
</tr>
<tr>
<td>4.10</td>
<td>Decision tree</td>
<td>98</td>
</tr>
<tr>
<td>5.1</td>
<td>Questionnaire structure flowchart</td>
<td>109</td>
</tr>
<tr>
<td>5.2</td>
<td>A satellite image of the city of Haifa with the locations of CBS and train stations</td>
<td>111</td>
</tr>
<tr>
<td>5.3</td>
<td>Relative household income distribution of survey respondents</td>
<td>120</td>
</tr>
<tr>
<td>5.4</td>
<td>Car ownership distribution among survey participants</td>
<td>120</td>
</tr>
<tr>
<td>5.5</td>
<td>Percentage of respondents that use a travel mode on a regular basis</td>
<td>121</td>
</tr>
<tr>
<td>5.6</td>
<td>Frequency of trips to the Technion</td>
<td>122</td>
</tr>
<tr>
<td>5.7</td>
<td>Number of respondents per origin travel zone</td>
<td>123</td>
</tr>
<tr>
<td>5.8</td>
<td>Travel ticket type share</td>
<td>123</td>
</tr>
<tr>
<td>5.9</td>
<td>Number of trip legs distribution</td>
<td>124</td>
</tr>
<tr>
<td>5.10</td>
<td>Cumulative distribution function of number of declared and specified alternatives</td>
<td>125</td>
</tr>
<tr>
<td>5.11</td>
<td>Number of specified path alternatives</td>
<td>126</td>
</tr>
<tr>
<td>5.12</td>
<td>Number of selected path alternatives</td>
<td>127</td>
</tr>
<tr>
<td>5.13</td>
<td>The cumulative density function of the ratio between the selected consideration-set and the given master-set</td>
<td>128</td>
</tr>
<tr>
<td>5.14</td>
<td>Inclusion probability and the average rating of path alternatives by the number of transfers</td>
<td>129</td>
</tr>
<tr>
<td>5.15</td>
<td>The cumulative density function of the ratio between the total in-vehicle time of paths selected to the consideration-set and reported paths relative to the total in-vehicle time of the shortest path</td>
<td>130</td>
</tr>
<tr>
<td>5.16</td>
<td>Choice frequency by relative attribute levels</td>
<td>134</td>
</tr>
<tr>
<td>5.17</td>
<td>Average importance grade of transfer stop attributes</td>
<td>135</td>
</tr>
<tr>
<td>5.18</td>
<td>Objective function value as function of the in-vehicle time ratio</td>
<td>137</td>
</tr>
<tr>
<td>5.19</td>
<td>Comparison of inclusion rate vs. simulated probability to board</td>
<td>142</td>
</tr>
<tr>
<td>5.20</td>
<td>Comparison of average rating vs. simulated probability to board</td>
<td>143</td>
</tr>
<tr>
<td>Section</td>
<td>Title</td>
<td>Page</td>
</tr>
<tr>
<td>---------</td>
<td>----------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>6.1</td>
<td>The route of bus line 51 in Tel-Aviv metropolitan area</td>
<td>155</td>
</tr>
<tr>
<td>6.2</td>
<td>Schematic route and load profile for line 51</td>
<td>156</td>
</tr>
<tr>
<td>6.3</td>
<td>Headway distribution at stop 28 on the inbound direction and at stop 4 on the outbound direction</td>
<td>158</td>
</tr>
<tr>
<td>6.4</td>
<td>A partial trajectory of inbound route</td>
<td>159</td>
</tr>
<tr>
<td>6.5</td>
<td>Coefficient of variation of the headway under various holding strategies</td>
<td>162</td>
</tr>
<tr>
<td>6.6</td>
<td>Time-space diagram of buses on service in line 51</td>
<td>164</td>
</tr>
<tr>
<td>6.7</td>
<td>Distribution of holding time at time point stops under various holding strategies</td>
<td>166</td>
</tr>
<tr>
<td>6.8</td>
<td>Trade-off between passenger in-vehicle delay and waiting time under various holding strategies</td>
<td>167</td>
</tr>
<tr>
<td>6.9</td>
<td>The route of bus line 1 in Stockholm inner-city</td>
<td>169</td>
</tr>
<tr>
<td>6.10</td>
<td>Load profiles of line 1</td>
<td>170</td>
</tr>
<tr>
<td>6.11</td>
<td>A snapshot of the current bus driver display</td>
<td>171</td>
</tr>
<tr>
<td>6.12</td>
<td>Vehicle trajectories on eastbound between 16:00-17:00 on 26-May-2008</td>
<td>172</td>
</tr>
<tr>
<td>6.13</td>
<td>On-time performance along the line</td>
<td>173</td>
</tr>
<tr>
<td>6.14</td>
<td>Correlation of speeds with schedule deviation along the line</td>
<td>175</td>
</tr>
<tr>
<td>6.15</td>
<td>Total travel time distribution</td>
<td>176</td>
</tr>
<tr>
<td>6.16</td>
<td>Headway distribution at terminals and time point stops along the line</td>
<td>177</td>
</tr>
<tr>
<td>6.17</td>
<td>Coefficient of variations of the headway along the westbound route under various holding strategies</td>
<td>182</td>
</tr>
<tr>
<td>6.18</td>
<td>Headway distribution under various time point layouts and holding strategies combinations</td>
<td>183</td>
</tr>
<tr>
<td>6.19</td>
<td>Trade-off between passenger in-vehicle delay and waiting time under various time point layouts and holding strategies combinations</td>
<td>187</td>
</tr>
<tr>
<td>6.20</td>
<td>Total travel time distribution under various time point layouts and holding strategies combinations</td>
<td>188</td>
</tr>
<tr>
<td>6.21</td>
<td>Schedule adherence distribution at the relief point ('Fridhemsplan') under various time point layouts and holding strategies combinations</td>
<td>191</td>
</tr>
<tr>
<td>6.22</td>
<td>An illustration of the trial period driver display</td>
<td>195</td>
</tr>
<tr>
<td>6.23</td>
<td>Greedy algorithm flowchart</td>
<td>196</td>
</tr>
<tr>
<td>6.24</td>
<td>Objective function value for the complete run of the greedy algorithm</td>
<td>203</td>
</tr>
<tr>
<td>6.25</td>
<td>Evolution of the holding and waiting times</td>
<td>204</td>
</tr>
<tr>
<td>6.26</td>
<td>Evolution of the coefficient of variation of the headway</td>
<td>205</td>
</tr>
<tr>
<td>6.27</td>
<td>Time point stop location in the complete greedy algorithm run</td>
<td>206</td>
</tr>
<tr>
<td>6.28</td>
<td>Genetic algorithm flowchart</td>
<td>207</td>
</tr>
<tr>
<td>6.29</td>
<td>Average, standard deviation and minimum objective function value per generation</td>
<td>208</td>
</tr>
<tr>
<td>6.30</td>
<td>Average and standard deviation of the total passenger-waiting time per generation</td>
<td>209</td>
</tr>
<tr>
<td>6.31</td>
<td>Average and standard deviation of the total passenger holding time per generation</td>
<td>209</td>
</tr>
<tr>
<td>6.32</td>
<td>Frequency of time point locations at various generations</td>
<td>210</td>
</tr>
<tr>
<td>6.33</td>
<td>Individuals obtained by the multi-objective genetic algorithm by various generations</td>
<td>211</td>
</tr>
<tr>
<td>6.34</td>
<td>Time components of selected solutions</td>
<td>212</td>
</tr>
<tr>
<td>6.35</td>
<td>Histograms of the objective function value for different number of time points, based on all the individuals generated during the genetic algorithm</td>
<td>213</td>
</tr>
<tr>
<td>Section</td>
<td>Title</td>
<td>Page</td>
</tr>
<tr>
<td>---------</td>
<td>-----------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>6.36</td>
<td>Histograms for all the feasible solutions of single and double time</td>
<td>214</td>
</tr>
<tr>
<td></td>
<td>point layouts</td>
<td></td>
</tr>
<tr>
<td>7.1</td>
<td>Framework for real-time information modeling in BusMezzo</td>
<td>221</td>
</tr>
<tr>
<td>7.2</td>
<td>A scheme for the real-time information generator of remaining time</td>
<td>225</td>
</tr>
<tr>
<td></td>
<td>until next arrival</td>
<td></td>
</tr>
<tr>
<td>7.3</td>
<td>Stockholm Metro network and the inner-city part of the network</td>
<td>229</td>
</tr>
<tr>
<td>7.4</td>
<td>Network configuration and travel attributes of the relevant trip</td>
<td>233</td>
</tr>
<tr>
<td></td>
<td>components</td>
<td></td>
</tr>
<tr>
<td>7.5</td>
<td>Path choice distribution</td>
<td>235</td>
</tr>
<tr>
<td>7.6</td>
<td>Coefficient of variation of the number of boarding passengers at</td>
<td>238</td>
</tr>
<tr>
<td></td>
<td>Fridhemsplan on the Green line</td>
<td></td>
</tr>
<tr>
<td>7.7</td>
<td>Stockholm’s rapid network as displayed by BusMezzo</td>
<td>239</td>
</tr>
<tr>
<td>7.8</td>
<td>Passenger load on rush hour trips of the eastbound light rail train</td>
<td>243</td>
</tr>
<tr>
<td></td>
<td>22 and outbound Metro line 14</td>
<td></td>
</tr>
<tr>
<td>A.1</td>
<td>Object-oriented framework for the transit simulation structure</td>
<td>274</td>
</tr>
<tr>
<td>C.1</td>
<td>Choice-set selection</td>
<td>278</td>
</tr>
<tr>
<td>C.2</td>
<td>Choice-set rating</td>
<td>278</td>
</tr>
<tr>
<td>C.3</td>
<td>Instructions on the dynamic path choice section</td>
<td>279</td>
</tr>
<tr>
<td>C.4</td>
<td>An example of the format of the dynamic path choice questions which</td>
<td>280</td>
</tr>
<tr>
<td></td>
<td>was given as part of the instructions preceding this section</td>
<td></td>
</tr>
<tr>
<td>C.5</td>
<td>Transfer choice attributes ranking</td>
<td>281</td>
</tr>
</tbody>
</table>
## List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Transit planning spectrum and the corresponding modeling characteristics</td>
<td>4</td>
</tr>
<tr>
<td>5.1</td>
<td>Questionnaire sections</td>
<td>107</td>
</tr>
<tr>
<td>5.2</td>
<td>Route choice experiment factors</td>
<td>116</td>
</tr>
<tr>
<td>5.3</td>
<td>Experimental design of the choice scenarios</td>
<td>117</td>
</tr>
<tr>
<td>5.4</td>
<td>Factor values on choice scenarios</td>
<td>118</td>
</tr>
<tr>
<td>5.5</td>
<td>Number of trip legs by origin and location</td>
<td>124</td>
</tr>
<tr>
<td>5.6</td>
<td>The share of chosen and consideration-set paths that satisfy various criteria</td>
<td>131</td>
</tr>
<tr>
<td>5.7</td>
<td>Regression analysis results</td>
<td>133</td>
</tr>
<tr>
<td>5.8</td>
<td>Matching between reported and generated consideration-set</td>
<td>138</td>
</tr>
<tr>
<td>5.9</td>
<td>Optimal solutions under all possible activation combinations</td>
<td>139</td>
</tr>
<tr>
<td>5.10</td>
<td>Performance measures of the choice-set generation model based on survey data</td>
<td>141</td>
</tr>
<tr>
<td>5.11</td>
<td>Estimated coefficient values for the simplified MNL model</td>
<td>145</td>
</tr>
<tr>
<td>6.1</td>
<td>Service measures of performance for inbound direction under various scenarios</td>
<td>165</td>
</tr>
<tr>
<td>6.2</td>
<td>Experimental design for holding scenarios</td>
<td>181</td>
</tr>
<tr>
<td>6.3</td>
<td>Service measures of performance under various holding scenarios</td>
<td>185</td>
</tr>
<tr>
<td>6.4</td>
<td>Transit level of service based on headway regularity (from TCRP, page 3-48)</td>
<td>185</td>
</tr>
<tr>
<td>6.5</td>
<td>Operational measures of performance under various holding scenarios</td>
<td>189</td>
</tr>
<tr>
<td>6.6</td>
<td>Effects of holding implementation on headway coefficient of variation</td>
<td>193</td>
</tr>
<tr>
<td>6.7</td>
<td>Summary of results for selected solutions</td>
<td>212</td>
</tr>
<tr>
<td>7.1</td>
<td>Summary of previous studies on bus prediction models</td>
<td>223</td>
</tr>
<tr>
<td>7.2</td>
<td>Experimental design summary</td>
<td>231</td>
</tr>
<tr>
<td>7.3</td>
<td>Average passenger journey time components</td>
<td>234</td>
</tr>
<tr>
<td>7.4</td>
<td>Average passenger journey attributes and the relative change compared with the base scenario</td>
<td>241</td>
</tr>
<tr>
<td>B.1</td>
<td>Transit-related input for BusMezzo</td>
<td>275</td>
</tr>
<tr>
<td>B.2</td>
<td>Transit-related output of BusMezzo</td>
<td>276</td>
</tr>
</tbody>
</table>
ABSTRACT

Efficient and reliable public transport systems are fundamental in promoting green growth developments in metropolitan areas. A large range of Advanced Public Transport Systems (APTS) facilitates the design of real-time operations and demand management. The analysis of transit performance requires a dynamic tool that will enable to emulate the dynamic loading of travelers and their interaction with the transit system.

BusMezzo, a dynamic transit operations and assignment model was developed to enable the analysis and evaluation of transit performance and level of service under various system conditions and APTS. The model represents the interactions between traffic dynamics, transit operations and traveler decisions. The model was implemented within a mesoscopic traffic simulation model. The different sources of transit operations uncertainty including traffic conditions, vehicle capacities, dwell times, vehicle schedules and service disruptions are modeled explicitly.

The dynamic path choice model in BusMezzo considers each traveler as an adaptive decision maker. Travelers’ progress in the transit system consists of successive decisions that are defined by the need to choose the next path element. The evaluations are based on the respective path alternatives and their anticipated downstream attributes. Travel decisions are modeled within the framework of discrete random utility models. A non-compensatory choice-set generation model and the path utility function were estimated based on a web-based survey.

BusMezzo enables the analysis and evaluation of proactive control strategies and the impacts of real-time information provision. Several experiments were conducted to analyze transit performance from travelers, operator and drivers perspectives under various holding strategies. This analysis has facilitated the design of a field trial of the most promising strategy. Furthermore, a case study on real-time traveler information systems regarding the next vehicle arrival time investigated the impacts of various levels of coverage and comprehensiveness. As passengers are more informed, passenger loads are subject to more fluctuation due to the traveler adaptations.
1. INTRODUCTION

1.1 TRANSIT SYSTEMS AND ADVANCED PUBLIC TRANSPORT SYSTEMS (APTS)

The steady growth in population, motorization and demand causes great traffic problems, mainly in large metropolitan areas. Transport authorities focus on more effective utilization of existing transport infrastructure by applying operation strategies and demand management schemes. It is well recognized that transit systems have a pivotal role in developing a more sustainable and efficient transport systems (Schrank and Lomax, 2005). Consequently, the improvement of transport services and management is one of the foundations of the EU transport policy (European Commission for Transport, 2009). An important challenge facing transport policy makers and planners is to design attractive alternatives to the private car. These efforts focus on improvements in terms of door-to-door times, reliability and comfort while at the same time minimizing operating costs.

An additional policy priority that targets the need for more efficient transport system is the further incorporation of intelligent transportation systems (ITS). ITS include a large range of such applications, among them electronic toll payment, traveler information and freeway management. The development of advanced technologies for transport systems contributes also to the improvement of transit systems. The set of ITS that is aimed to improve transit performance and level of service is known as advanced public transport systems (APTS). APTS are generally classified into four categories of systems: fleet management, traveler information, electronic payment, and demand management (Morgan, 2002). Instantaneous data collection and communication technologies enable the design and application of real-time monitoring and control schemes. The implementation of these schemes has the potential to improve transit performance and level of service. An example of APTS application is the provision of real-time arrival information at stops based on automatic vehicle location (AVL) systems, which provide passengers with real-time departure information (FHA and FTA, 2000; FTA, 2006). The implementation of AVL systems also supports applications of various schedule monitoring techniques (such as holding, skipping and dispatching decisions) and transit signal priority (TSP) schemes. The Federal Transit Administration
reports that APTS implementation increased by over 70% between 1995 and 2000 (FTA, 2000). The intensified adoption of APTS calls for methods that will represent their operation and passengers’ response to them in order to evaluate them and refine their design.

1.2 Transit Modeling Spectrum
The need to integrate and operate increasingly complex, diverse and technology-oriented transit services poses a challenge to both planners and operators. Furthermore, as new technologies and applications are proposed, tools to assist in their development and evaluation prior to field implementation are needed. This results in a growing need for tools to assist policy planning and to analyze and evaluate operations and management schemes of transit systems. There is large range of methods and tools aimed to support transit agencies and operators decision making with regards to various applications. Table 1.1 summarizes the attributes of three levels of applications with regards to transit modeling – from the strategic planning level through the operation and management level and down to the implementation details – and their respective modeling characteristics. There is an inverse relationship between the decision horizon and the appropriate level of detail (Lee, 1994).

Long-term strategic transport planning is typically based on the classic four steps model. The conventional four steps model was extended and revised in recent years to accommodate activity-based modeling and trip departure choice (Ortuzar and Willumsen, 2001). The four-step planning model is aimed for strategic planning and policy making and has therefore to take into account long-term processes as land-use development, socio-demographic trends and future infrastructures and services. There are several commercial packages that are commonly used for predicting traffic and transit conditions based on the four-step models (e.g. TRANSCAD (Caliper Co., 1996), EMME/2 (INRO, 1999), VIPS (VIPS, 2000)). These models are useful for long-term planning, where the input is approximated and the output is interesting at the network-wide aggregated level. However, those models are not suitable for mid- and short-term transit planning and operation analysis, where the dynamic evolution of system conditions is the main interest.
<table>
<thead>
<tr>
<th>Decision horizon</th>
<th>Strategic planning</th>
<th>Operations and management</th>
<th>Implementation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Scale</strong></td>
<td>Long</td>
<td>Intermediate</td>
<td>Short</td>
</tr>
<tr>
<td>Large areas (e.g. region)</td>
<td>Middle-size areas (e.g. city)</td>
<td>Local (e.g. intersection, corridor)</td>
<td></td>
</tr>
<tr>
<td><strong>Example applications</strong></td>
<td>Network design; Transit oriented development (TOD)</td>
<td>Control strategies; Transit priority; Timetables</td>
<td>TSP schemes; Stop capacity; Driver scheduling</td>
</tr>
<tr>
<td><strong>Details</strong></td>
<td>Macro-level</td>
<td>Intermediate</td>
<td>Micro-level</td>
</tr>
<tr>
<td><strong>Tools and methods</strong></td>
<td>Travel habit survey; Four steps model; Static assignment models</td>
<td>Dynamic models; Simulation models; Mathematical programming</td>
<td>HCM and TCQSM; Field tests; Designated software (crew scheduling, signal planning)</td>
</tr>
<tr>
<td><strong>Traffic dynamics</strong></td>
<td>Macroscopic - equilibrium conditions are important</td>
<td>Interactions with other dynamic factors and variability</td>
<td>Detailed fine-tuning (acceleration and deceleration, spillbacks)</td>
</tr>
<tr>
<td><strong>Transit operations</strong></td>
<td>Very simplified and deterministic across the network (e.g. fixed headways, no capacity constraints)</td>
<td>Factors that affect transit performance and level of service (e.g. dwell time, trip chaining)</td>
<td>Detailed at the local level (e.g. door configuration) and approximations for external factors (e.g. headway distribution)</td>
</tr>
<tr>
<td><strong>Transit demand</strong></td>
<td>Mode choice and induced demand are of interest. Simplified behavioral assumptions.</td>
<td>The total demand for transit is given at some level (e.g. OD matrix in terms of stops).</td>
<td>Taken as given at the segment level (external).</td>
</tr>
<tr>
<td><strong>Main measures</strong></td>
<td>Mode share; Passenger flows.</td>
<td>Reliability; Travel times.</td>
<td>Efficiency; Capacity.</td>
</tr>
</tbody>
</table>
Analytical models are used to study a large range of transit-related issues at various stages of planning. Operations research methods and tools were developed and applied for transit-related problems such as network design, timetable optimization, vehicle, driver scheduling and the design of control strategies. For a comprehensive review of the operations research literature in the context of public transport see Desaulniers and Hickman (2007). At the strategic level of network design, the main goal is to satisfy passenger demand with a certain system performance based on route choice and assignment models. At the operational level, frequencies and timetables are constructed using mathematical programming and optimization methods. Similar methods are applied for allocating vehicles and drivers, which are typically solved sequentially and may feedback to timetable planning.

Implementation problems are concerned with issues such as TSP at a specific intersection, stop or bus way capacity or the layout of a transit facility. Some of these issues can be studied by conducting field studies that can be used for estimating a mathematical formulation. Some common implementation decisions can be supported by designated tools, software and manuals that are available (e.g. TCRP 2003a). These decisions require a detailed level of modeling and empirical data measurements for limited scale and aspects. The analysis can help fine-tuning or refining implementation details based on location-specific conditions.

The focus of this thesis is mainly at the transit operation and management - the intermediate level of transit modeling. This domain includes a large variety of problems and applications, among them: transit performance and level-of-service analysis; evaluation of service reliability and control strategies; impacts of transit priority; assessment of real-time information (RTI) provision; restoration from service disruptions; layover and recovery time assessment; impacts of temporal or permanent route changes; timetable optimization and intermodal coordination.

Static assignment models are not suitable for this analysis as they cannot capture the time-dependent variation in transit supply and demand. Therefore, it cannot replicate the inter-related dynamic processes that drive service unreliability, crowding conditions, the generation of RTI and both operators and passengers reaction to system conditions. Furthermore, due to the nature of transit systems in terms of size, complexity and dynamics – in particular with the implementation of APTS - it is
unrealistic to apply global analytical models. Computer simulation models may offer a feasible, flexible and attractive tool for analyzing transit performance.

1.3 TRANSIT SIMULATION MODELS
In the context of general traffic operations, simulation models have been established as the primary tool for evaluation at the operational level. Recently, they have also been extensively used to represent and evaluate various ITS applications. Traffic simulations are classified into three classes, according to their level of detail and aggregation: Macroscopic, Microscopic and Mesoscopic. Macroscopic models represent traffic as a continuous flow based on flow-density functions without the explicit modeling of lanes or vehicles. At the other extreme, microscopic models represent traffic at the most detailed level: individual vehicles are represented and their behavior depends on their interactions with other vehicles, geometry, lane assignments etc. As a result of computational constraints, there is an inverse proportionality between the level of details and the possible size of networks under study. A third group of models exists on this scale, mesoscopic models, which represent individual vehicles but avoid detailed modeling of their second-by-second movement.

Transit simulations may serve several interests (Meignan et al., 2007): observation of network dynamics and design; evaluation and control of dynamic processes (e.g. transfers coordination); evaluation of network performance under alternative designs (e.g. routes or frequencies). Although simulation models can have many advantages for transit research, there has not been much effort in the development of transit simulation models.

Surveys of traffic simulations found that users’ perceive transit among the most important features to be included in the simulation (Algers et al., 1997, Boxill and Yu, 2000). However, only 52% of the micro-simulation models that were reviewed modeled transit, 26% produced transit related outputs and merely 6% modeled transit-related information. The researchers concluded that micro-simulations are not effective for large-scale networks because of the unnecessary level of details and the lack of transit modeling and none of them posses all the requirements for APTS representation. However, it should be noted that microscopic simulation had been improved significantly in recent years. At the time, none of the mesoscopic models reviewed, had neither a transit simulation component nor suitability to simulate APTS.
A simulator capable of representing transit systems and APTS applications requires several, possibly contradicting capabilities: on one hand a detailed representation is needed to capture the complex interactions among vehicles and passengers, on the other hand it is essential that the simulation model would be able to represent large scale metropolitan networks, in order to evaluate the performance of transit at a system level. Given these requirements, a mesoscopic traffic simulation seems the suitable platform for transit operation and APTS evaluation.

1.4 RESEARCH OBJECTIVES AND APPROACH
This study is aimed to develop a tool for analyzing transit performance under various operational conditions and APTS. The model is designed to enable the analysis of transit operations, passenger path choice decisions and their interactions with traffic dynamics at a network-wide level. The integration of these components enables a joint car and transit loading tool. It would be realized by designing a framework for representing the processes that determine how the system evolves.

Transit supply would be represented with the intention to capture the main sources of uncertainty and their inter-related dynamics. This is necessary in order to reproduce the variation in service conditions which is an important determinant of the experienced level-of-service. The simulation model will facilitate the evaluation of real-time operations strategies.

Transit demand would be modeled as an adaptive process where individual travelers make decisions based on their preferences and perceptions. A behavior model would imitate how travelers compose their choice-set and choose between travel alternatives along their trip from a given origin to their final destination. The specification of model parameters would be based on a survey that would be conducted as part of this study. In the design of the questionnaire, a special attention would be given to how travelers compose their choice-set.

The model will be implemented as a transit simulation model that will enable to analyze and evaluate alternative strategies at the system level. The simulation model would be developed within a mesoscopic traffic simulation model. In this research, Mezzo, a mesoscopic traffic simulation model (Burghout, 2004) is used as the development platform. As part of my master thesis, the basic entities of transit operations and their mechanisms were integrated into Mezzo, including timetables,
dwell times and passenger arrival processes (Cats, 2008; Toledo et al., 2008). The transit operations and control modeling capabilities are extended and elaborated in this thesis. An agent-based approach is adapted for the representation of individual travelers.

Examples of potential applications include: service reliability analysis, frequency determination; timetable optimization and service coordination; effects of temporal or constant transit route changes; impacts of transit priority measures; assessment of operational policies as control strategies or layover time at vehicle scheduling; passenger flows analysis and; the evaluation of real-time traveler information.

The objectives of this research are in line with a growing trend in transport research to move towards dynamic models. Hence, the development of a dynamic transit model contributes to narrowing down the existing modeling gap in the public transport research field. This development can stimulate the shift towards analysis tools that account for the variation in transit supply and demand. Moreover, it is consistent with the growing need for taking proactive and adaptive strategies towards transit management that are facilitated by recent APTS capabilities.

1.5 Thesis Outline
The remainder of this thesis consists of the following: a literature review where previous studies and the state-of-the-art of transit assignment models and route choice models are discussed (Chapter 2). Chapter 3 presents the framework of the dynamic transit model and describes the transit operations modeling components. Demand modeling is presented in Chapter 4— including the two-stage modeling approach, the choice-set generation model and the details of the dynamic path choice process. Chapter 5 describes the design and results of a survey that was conducted along with the estimation of the two-parts of the path choice model. Applications of the model are presented in Chapters 6 and 7. The model was used for the analysis of transit performance and the evaluation of control strategies (Chapter 6) as well as for the assessing the impacts of real-time information provision under various operations conditions (Chapter 7). Finally, this thesis concludes with a discussion of its contribution and an outline of potential future directions of research (Chapter 8).
2. LITERATURE REVIEW OF TRANSIT ASSIGNMENT AND SIMULATION MODELS

A review of the transit modeling literature reveals a detached portrait of the transit system with a pronounced division between the research on transit operations and transit assignment. The following review covers previous studies in transit modeling that focus on how transit travelers move in the transit system and how they are affected by transit system conditions. Aspects related to transit operations per se are discussed in the respective sections in the thesis.

The literature review is organized as follows: The transit assignment problem is presented, followed by a review of conventional approaches for transit assignment models (TAM). Section 2.2 discusses recent developments of transit simulation models and their potential to capture the dynamics of transit operations and individual travelers. A dynamic representation involves the modeling of traveler’s path choice decisions. Alternative approaches towards discrete choice models (DCM) are reviewed in Section 2.3. Methods for generating choice-sets and path choice models are discussed in the context of transit networks. This chapter concludes with a synthesis of the state-of-the-art and points out modeling issues that needs to be addressed (Section 2.4).

2.1 CONVENTIONAL TRANSIT ASSIGNMENT MODELS

2.1.1 THE TRANSIT ASSIGNMENT PROBLEM
Traffic assignment models constitute the forth class of models in the classic four-step transport forecasting process (Ortuzar and Willumsen, 2001). The assignment follows the phases of trip generation, trip distribution and mode choice. Traffic assignment models take the mode-specific travel demand OD matrix and distribute it over the transport network by assigning trips to routes. Similarly, the transit assignment problem is concerned with how flows are distributed over transit paths on a given transit network for a given OD travel demand. The interaction between travel demand and transit network supply determines the transit system performance. Therefore, the core of any assignment model is a route choice model. The route choice model links passenger decisions with network conditions based on user preferences and service
characteristics. The process of assigning passengers to transit paths requires the modeling of passenger perceptions and travel behavior.

TAM load transit passengers on a given transit network to obtain passenger loads and the level-of-service. Hence it is a fundamental analysis and evaluation tool at both planning and operational levels. Subsequently, much research effort was devoted to the development of TAM in the last few decades. Many of those modeling attempts adopted ideas from general traffic assignment models and tried to adjust them to transit network conditions. However, several characteristics of transit systems introduce additional complexities to the car traffic assignment problem. The main reason for greater complexity is the discontinuous availability of transit supply both in space and time. This is especially evident in the case of transfer connections with temporal and spatial constraints. Hence, the importance of modeling walking and waiting times. An additional complexity arises from the relationship between service uncertainty, passenger loads, comfort, travel times and capacity constraints. Furthermore, most transit networks consist of several modes with distinguished sub-networks. These networks exercise different levels of interaction with car traffic (Nielsen, 2000; Wahba and Shalby, 2005).

Traffic assignment models are commonly classified based on their deterministic or stochastic equilibrium conditions and their static or dynamic loading procedure. Likewise, these classifications also apply to transit assignment models. A static representation and loading process of the transit system could be justified in case of long-term planning applications. However, static assignment models neglect the evolvement of network conditions, time-dependent interactions and en-route user decisions.

Conventional TAM are static equilibrium assignment models which are insensitive to service disturbances, the effects of information and incidents. The following presents the two classes of conventional transit assignment models: frequency-based TAM (FB-TAM) and schedule-based TAM (SB-TAM). This classification is based on the representation of the transit network as it has substantial impacts on the passenger loading procedure. FB-TAM represents of the transit network at the line-level with the corresponding frequencies, while SB-TAM includes a more detailed representation of the time-dependent specific vehicle-runs (Lam and Bell, 2003; Ceder,
A review of the state-of-the-art FB- and SB-TAM developments is given in the following sections.

2.1.2 Frequency-based Assignment Models

Early attempts to propose TAM were based on applying user equilibrium (UE) conditions to transit networks (Dial, 1967; Le Clercq, 1972). These algorithms did not consider the common lines problem – how passengers are distributed between several lines that compose the trunk-line link. In a review of operations research methods applied to public transport problems, Desaulniers and Hickman (2007) list three main challenges in the determination of the minimum cost path: time-dependent stochastic attributes; path definition and its compatibility with the common lines problem and; impacts of capacity and discomfort.

A probabilistic framework for this problem was presented by Chirqui and Robillard (1975) assuming that passengers board the first arriving vehicle that belongs to a set of attractive lines. Marguier and Ceder (1984) extended the analysis of the common lines problem by considering the influence of bus regularity and passenger arrival process.

An important advancement in the field of transit path choice was the result of studies by Nguyen and Pallottino (1988) and Spiess and Florian (1989). Spiess and Florian defined travel strategy as a set of rules that when applied allows the traveler to reach his or her destination. Their optimal strategy model minimized the total travel time which is composed of access, waiting and in-vehicle time. It is still assumed that passengers board the first arriving bus from the attractive set of transit lines. The attractive set includes all the lines that their riding time is not longer than the expected total travel time of the remaining lines in the set. The latter is calculated as a weighted average by considering the line-probabilities to split proportionally to the frequencies, regardless of their riding time. The transit equilibrium model was formulated as a mixed integer program with an objective function of total travel time. The problem included flow conversation and non-negativity constraints. They were the first to transform the problem into a linear programming problem. Nguyen and Pallottino presented a graphic representation for the transit loading procedure. A hyperpath was defined as an acyclic directed graph from origin to destination which results from performing a strategy. The share of passenger flow using each outgoing transit link is
proportional to the corresponding frequencies on the hyperpaths so that flows can be calculated backwards, starting from the destination.

The notions proposed by the above studies – strategy and hyperpath - provided the foundations for many of the FB-TAM developed ever since. Passenger loads obtained by FB-TAM depend on model assumptions on passenger arrival process and service headway distribution. Note that these assumptions determine both waiting times and the line-specific probabilities calculations. Spiess and Florian (1989) suggested that the expected waiting time equals to half the joint headway calculated as the inverse of the simple sum of all common lines frequencies. This calculation of the expected waiting time is in line with the line-specific probabilities suggested by Nguyen and Pallottino (1988) – with each line attracting the ratio of its frequency to the joint frequency.

The underlying assumptions of these calculations are that passengers arrive randomly at stops and service headways are deterministic. Constant service headways can be obtained only under perfectly regular service. Moreover, the assumption that headways of common lines are independent implies that all headways are perfectly coordinated in the sense that arrivals from different lines that share the same segment are equally spaced. Such a perfect coordination is even theoretically possible only in the case of identical headways on each common lines corridor.

Unrealistic assumptions about service regularity and coordination result ultimately in an underestimation of the expected waiting time. Furthermore, the calculation of waiting times at transfer location is based on the same assumptions as for an origin stop, hence neglecting the case of timetable coordination. These assumptions of the FB-TAM are inconsistent with neither analytical models nor statistical analysis of real-world data (Chen et al., 2009; Bellei and Gkoumas, 2010). Even though these set of assumptions is unrealistic, as was pointed out even by the original contributions, they are widely applied, including by commercial static TAM as EMME/2 and TRANSCAD. Many of later developments in the domain of FB-TAM were directed to refine or relax some of the above assumptions:

- Service regularity – the assumption of perfectly even headways was first revised by Marguier and Ceder (1984) who considered the case of perfectly irregular service. This extreme memory-less arrival process implies an exponential headway
distribution. Jayakrishnan et al. (1994) discussed the implications of various assumptions on service regularity. Gentile et al. (2005) generalized the waiting time function by formulating it as an Erlang distribution which can accommodate the entire range of headway distribution from perfect regularity to perfect uncertainty. Shimamoto et al. (2010) developed a TAM that takes into account the correlation between successive vehicle arrivals.

- Common lines coordination - the assumption of perfect coordination of arrivals from different lines was replaced with the assumption that common lines arrive independently by Jansson and Ridderstolpe (1992). Hence, arrival times are uniformly distributed over their inter-arrival times. The static transit assignment tool VIPS (2000) allows the specification of different assumptions on service coordination including the case of timetable coordination at transfer hubs. Hsu (2010) estimated transfer waiting times under different headway variability of the feeding and connecting services.

- Behavioral rules – Previous studies have proposed various refinements to the assumption that passengers get on the first arriving vehicle which belongs to the set of attractive lines. Andreasson (1977) proposed a heuristic to remove from the choice-set paths that their in-vehicle time (IVT) is longer than the waiting time plus the IVT of one of the other alternatives. This method was further extended by Jansson and Ridderstolpe (1992). Jansson (2003) reviewed and compared the principles used in two well-known static transit assignment tools – EMME/2 and VIPS. The TAM of VIPS was embedded in VISUM (PTV, 2009). Billi et al. (2004) proposed the dynamic composition of the set of attractive lines so that each attractive line is associated with a certain waiting period. Nökel and Wekeck (2007, 2009) investigated the various behavioral assumptions and found significant difference in their choice-set composition and line-probability computations.

- Capacity constraints – None of the above studies enforced binding capacity constraints. Capacity effects were considered only implicitly in the optimal strategy/hyperpath approach. Spiess and Florian (1989) approximated the discomfort effect of capacity constraints by defining passenger travel times as an increasing function of passenger flow. Along the same line, the effective service frequency approach attempts to account for congestion effects by associating a higher probability to denied boarding with a lower effective frequency (De Cea and
An alternative approach is to associate a route segment in the transit network with a certain capacity and adding a discomfort component based on the flow-capacity ratio (Lam et al., 1999; Hamdouch et al., 2004; Leurent and Askoura, 2010). The network representation of Nguyen and Pallotino (1988) can be extended by introducing failure-to-board arcs that are used in case transit line capacity is exceeded (Kurauchi et al., 2003; Schmöcker et al., 2008). A detailed modeling of seat priority and allocation was developed by Schmöcker et al. (2011).

- Traveler heterogeneity and information - TAMs based on the UE formulation assume that travelers are homogenous and have perfect information on system conditions. Jayakrishnan et al. (1996) calculated the expected waiting times and line probabilities under various information scenarios. Larsen and Sunde (2008) discussed the different behavioral assumptions on passenger waiting time and highlight the importance of heterogeneity in travelers’ decisions. Following the developments in the field of traffic assignment models, Lam et al. (1999) formulated the stochastic user equilibrium (SUE) conditions for the transit assignment problem based on the formulation of Spiess and Florian (1989). This formulation considers the heterogeneity in passengers’ perceptions and network knowledge. Nielsen (2000) and Sumalee et al. (2011) developed a Probit-based FB-TAM in order to capture the perceived correlation between alternative paths due to overlapping.

FB-TAM are typically static as they consider average supply and demand conditions rather than the variations in service conditions and demand characteristics. The importance of such variations and their implications on individual runs are the motivation behind the development of SB-TAM.

2.1.3 SCHEDULE-BASED ASSIGNMENT MODELS
SB-TAMs represent both the supply and demand sides of the transit system as time-dependent. Transit service is represented in terms of individual vehicle runs following a given timetable. Passenger demand is segmented to time intervals associated with desired departure or arrival times. Time-dependent passenger demand is loaded on specific transit vehicles following a path choice model that takes into account the time-dependent properties (Nuzzolo and Crisalli, 2004). The schedule-based approach enables the consideration of timetable coordination and low-frequency services. These
functionalities are implemented in the SB-TAM available in VISUM (PTV, 2009) which is based on VIPS (2000).

Since total travel time in a schedule-based network becomes time-dependent, the concept of accumulating shortest-path widely used in shortest-path search methods is not valid anymore. Hall (1986) presented a method to find the shortest path for random and time-dependent travel times. Furthermore, this study introduced the concept of time adaptive path choice – path choice is not static but depends on arrival times.

Time-dependent transit networks can be represented using a time-space graph as proposed by Nuzzolo and Russo (1996). Any change in vehicle state – arrival, departure, dwelling - is represented in the diachronic graph by an arc. Nguyen et al. (2001) proposed an extension to the graphic framework of the transit assignment problem by including detailed boarding, alighting, access, egress, transfer and walking links. This representation enables to represent graphically departure and route choice decisions. These decisions are treated as a single simultaneous decision that depends on the expected travel conditions.

The developments in the domain of SB-TAM correspond to the modeling concerns of the FB-TAM research:

- Service regularity - the graphic representation of deterministic individual vehicle run arrival/departure times implies the assumption of perfect punctuality. Service irregularity can be captured either implicitly by adding a random term to the perceived utility function (Nielsen, 2004) or explicitly by simulating vehicle runs and dwell time as inter-dependent random variables (Nuzzolo et al., 2001; Huang and Peng, 2002).

- Capacity constraints – as in the case of FB-TAM, can be modeled implicitly or explicitly. However, capacity constraints can be potentially handled in a more delicate way in SB-TAM due to the representation of individual vehicle runs. Nuzzolo et al. (2001) used an asymmetric penalty cost function to approximate the impacts of capacity constraints. Hamdouch and Lawphongpanich (2008) developed a SB-TAM version of their user preferences over hyperpath set model (Hamdouch et al., 2004). Rochau et al. (2010) extended the model by incorporating the effect of unreliability that can arise from capacity constraints. The alternative approach towards modeling
capacity constraints by assigning passengers with a failure-to-board probability based on the ratio between the residual capacity and the number of passengers trying to board was applied by Zhang et al. (2010). A more elaborated priority scheme was developed by Sumalee et al. (2009) which was later contested by Schmöcker et al. (2011) on the ground of its feasibility for large scale networks.

- Traveler information – the dynamic properties of SB-TAM make it suitable for studying the effects of real-time information (RTI) provision. Hickman and Wilson (1995) developed an analytical framework for adaptive path choice model that enables to evaluate the influence of RTI regarding the next arrival of each bus line. They considered different levels of information accuracy, assuming that passengers make the best use from the available information through a deterministic network loading model. The effect of anticipated travel conditions based on traveler experience and RTI was included in the stochastic path choice model specified in Nuzzolo et al. (2011).

- Traveler heterogeneity - variations among travelers were studied also within the sphere of SB-TAM by introducing random components into the utility function. Tong and Wong (1998) investigated the impacts of time-dependent supply and demand conditions by including a sensitivity measure that varies among passengers. Nielsen and Frederiksen (2006) developed and estimated a SB-TAM where the utility function of an individual vehicle run included error components to account for stochastic service delays, variation in transit network knowledge and individual preferences. Modeling supply variations in the SB-TAM framework adds substantial complexity to the graph representation adapted from Nuzzolo et al. (2001).

As for car traffic, TAM can be based on equilibrium conditions or dynamic loading process. The SUE-TAM is formulated as a fixed-point problem typically solved by the heuristic method of successive averages technique (Nielsen, 2000; Nuzzolo et al., 2001; Zhou et al., 2008; Sumalee et al., 2009; Zhang et al., 2010). As pointed out by Nuzzolo and Crisalli (2009), the theoretical ground for equilibrium assignment is well-funded while the research on theoretic aspects of dynamic loading is still underway. In their review of SB-TAM, Nuzzolo and Crisalli further argued that in operational context mode choice should be analyzed based on schedule-based models. This implies the consideration of individual vehicle-run alternatives in the mode choice phase rather
than the generic modes – either by a joint departure time-mode choice or by joint mode-run choice. Recently, this approach was applied in the context of multi-modal corridors (Zhou et al., 2008; Cascetta et al., 2011).

2.2 Transit simulation models

2.2.1 Background

Computer simulation models prevailed as a prime analysis tool in the context of car traffic. A substantial research effort was devoted in the last three decades to the development of dynamic traffic assignment (DTA) models. Peeta and Ziliaskopoulos (2001) provide a comprehensive review of these developments. They highlighted the limitations involved with analytical approaches for developing a DTA model for general networks and the unrealistic representation of traffic dynamics that they imply. In contrast, the simulation-based approach has substantial advantages in the development of DTA models that are practical for realistic networks. Moreover, simulation models enable to incorporate multi-user classes and their respective interactions in the transport network, information provision and decision processes. They concluded that simulation models are more suitable for studying system robustness and for incorporating sources of randomness that yields the stochastic DTA problem. The main drawback of simulation models is the inability to form mathematical functions that describe the system properties in order to get some insights. De Palma and Marchal (2002) discussed modeling issues related to DTA simulation models. They concluded that the combination of event-based mesoscopic modeling of the supply side along with a disaggregate demand modeling of individual decision makers yields the best conditions for analyzing large-scale systems. This is particularly true when considering advanced traffic management systems (ATMS) and advanced traveler information systems (ATIS) applications. An important advantage of this approach is that it enables the behavioral modeling of decision makers based on time-dependent OD matrices (Nagel and Marchal, 2003).

The developments in the field of traffic assignment models point to the potential role that simulation models can play in the context of TAM. Simulation models provide an appropriate platform to enhance the realization of transit system modeling. These includes the capabilities to reproduce the dynamic nature of trip generation; the
dynamic evolution of network conditions; the interaction between supply and demand; the variation among travelers and their adaptive behavior and; to emulate the generation of passenger information services.

2.2.2 Transit Operations Simulation Models

Transit simulation models vary in the level of integration that they exercise with a general traffic simulation model. Few transit-related simulations were conducted through either manipulations or adjustments of traffic simulation models that do not represent transit operations (Khasnabis et al., 1997; Chang et al., 2003). Many others involved limited enhancements of existing simulation models that extended their capabilities for specific purposes (Liu et al., 1998, 2006; Ding et al., 2001; Werf, 2005; Cortes et al., 2005; Abdelghany et al., 2006; Wahba and Shalaby 2006a; Cortes et al., 2007). This intermediate approach includes a wide spectrum of integration levels: from completely external and separated transit sub-models in the form of an API (Application Programming Interface) to internal partial modifications. These simulations are useful for specific applications but lack a comprehensive transit modeling framework.

Several transit simulation models were developed in the last decade to enable a more realistic representation of transit operations characteristics. Fernandez and Planzer (2002) proposed a simulation model called PASSION, parallel stop simulation, for the analysis of stop design and performance. The model simulates the operations of the immediate stop area under different vehicle and passenger arrival patterns. Passenger and bus vehicle arrival times can be specified or drawn from a distribution. The model was calibrated and validated by comparing video recordings of stop operations with simulated outputs (Fernandez, 2010). Morgan (2002) integrated the transit representation into MITSIMLab, a microscopic traffic simulation that was designed for the design and evaluation of ATMS and ATIS (Yang and Koutsopoulo, 1996). As a microscopic simulation model, transit vehicle movements around stops were represented in great details, albeit the representation of trip chaining is limited due to network size constraints. The general traffic management simulator component was enhanced to evaluate TSP strategies. However, ATIS were not modeled in the transit context. TSPs were also the subject of a microscopic simulation analysis by Lee et al. (2005). Vehicle movements were modeled in detail in order to predict adequately the travel time between detection and arrival time at the intersection. Milkovits and Wilson (2010) represented transit routes as a sequence of running and dwelling times with the
latter calculated explicitly only at key stops. The simulation accounted for vehicle scheduling and the variability in dispatching times from the terminal. As is the case for all the studies mentioned above, passengers were merely represented in terms of volumes that are derived from a certain distribution.

The applicability of microscopic transit models to large-scale applications is limited because of their high level of network detail. This consideration led Meignan et al. (2007) to propose a transit simulation model that models traffic conditions at a macroscopic level except of the representation of individual bus vehicles. Mesoscopic simulation models, which provide modeling of individual vehicles but avoid detailed modeling of their movement, may be useful for system-wide evaluation of transit operations and APTS. The desired mesoscopic transit simulation model has to fulfill the requirements identified by Morgan (2002) for an APTS simulator: transit system representation, transit vehicle movement and interaction, transit demand representation, APTS representation and the generation of measures of effectiveness.

2.2.3 Transit Assignment Simulation Models
In a review of transit assignment models, Liu et al. (2010) compared the evolution of transit passengers’ route choice behavior modeling with that of road users. They concluded that transit modeling is consistently lagged behind the developments in traffic modeling. Based on the developments in the latter they expect multi-agent non-equilibrium models to emerge in the transit domain as well. The main modeling issues are supply uncertainties and adaptive user decisions. They identified dynamic loading process and multi-agent-based simulation as two potential approaches for modeling complex transit systems.

Following developments in the sphere of traffic assignment models, there are few recent efforts in the transit domain. The evolution of transit simulation models into dynamic transit assignment tools is at its early stages. This evolution is coupled with the microscopic simulation of individual travelers that has recently emerged as the new approach for modeling traffic dynamics and forecasting traffic conditions. The so-called agent-based approach used in a range of sciences is aimed at modeling complex systems by representing the strategies of individual agents and the dynamics between an agent and the environment and interactions between agents. Nagel and Marchal (2003) provided an interesting discussion on modeling and computational issues of multi-agent simulation models in the traffic sphere. These models are intended to mimic the
adaptive response of travelers to changing system conditions and the incorporation of information. Salvini and Miller (2005) and Ettema et al. (2007) reported the developments of ambitious urban models called ILUTE and PUMA, respectively. Wahba and Shalaby (2006b) discussed the potential advantages of a multi-agent simulation framework for modeling the transit assignment problem, in particular in the context of APTS and the evaluation of RTI (Wahba and Shalaby, 2006a).

A multi-agent simulation model of transit passengers was proposed by Meignan et al. (2007). Traveler behavior was modeled as a single pre-trip mode choice decision. This decision considered three alternatives – the shortest path by car, walk and transit alternatives. Waiting times for the transit alternative is calculated as half the planned headway. This implies that travelers always take the shortest path for a given travel mode, hence lacking path choice modeling framework.

Rieser et al. (2009) presented the extension of MATSIM, an agent-based transport simulation model, to transit trips. The activity-based framework of MATSIM was extended to include transit trips in the plan generation process that assigns a travel plan to an agent in the simulation model. A traffic simulation model executes the selected travel plans with an iterative learning process based on travel plans scoring function. However, there is no modeling of transit network supply. Total transit journey times are assumed to be deterministically twice the corresponding car free speed times. There is no consideration of transit paths composition and time components or capacity constraints. The trip-related component in the scoring function is based solely on this assumed total journey time, hence disregarding any difference between various transit paths. Mode choice is modeled as part of the re-planning learning process by treating mode and route choice jointly. Neumann and Nagel (2010) have used MATSIM for analyzing bunching and holding strategies. They introduced a disturbance in passenger demand at a certain stop in order to trigger bus bunching. This trigger was needed because of the inadequate modeling of the sources of transit supply uncertainties and transit dynamics that are in the core of the bunching phenomenon. Passenger volumes, travel times between stops and dispatching from terminals are all modeled as deterministic and in agreement with the timetable in MATSIM.

MILATRAS, a micro-simulation learning-based approach for transit assignment, is the most significant attempt to assign transit passengers in a transit simulation framework (Wahba and Shalaby, 2006a; Wahba, 2008). The model was implemented as
an API enhancement to PARAMICS, a mesoscopic traffic simulation model. However, the model represents the movement of transit vehicles between stops as a function of the link speed extracted from PARAMICS with exogenous travel time variability; hence buses and cars do not interact. The enhancement enables the representation of individual passengers conducting repetitive within-day and day-to-day decisions in a non-equilibrium process. It considers departure time, stop and individual vehicle-run choices, following the conceptual framework of SB-TAM and with the purpose to model medium-size networks with medium to low frequencies. Passengers in MILATRAS have no prior-knowledge; the joint decision depends on the experience accumulated by each traveler on transit performance based on recurring loading processes. The generalized cost is a deterministic function with passengers following the alternative that yields the lowest score function. The transit path learning process is modeled as a Markov decision process (Wahba and Shalaby, 2009a). A stationary state may be reached when passengers choose their optimal strategy after the exploitation behavior dominates the learning process. It should be noted that the learning-based assignment approach of MILATRAS yields in convergence the same general-cost function for all passengers having the same OD and departure time-period for any given decision alternative. Therefore, each group of passengers will follow the same path, since the choice model is deterministic with passengers choosing the alternative with the lowest cost function. Furthermore, the representation of transit operations in MILATRAS is rather limited; most noticeably vehicle scheduling and control operations are not modeled.

The iterative adaptation of traveler strategy is sometimes regarded as a learning process that carries out a non-equilibrium assignment of a system that may correspond on certain conditions to UE or SUE conditions (Nagel and Marchal, 2003). This process depends on the learning mechanism and the weight that is given to the most recent experience compared with past experience. Such a trade-off was exercised in the context of a microscopic dynamic traffic simulation by Liu et al. (2006), in SB-TAM by Nuzzolo et al. (2011) as well as in the context of simulation-TAM in MILATRAS (Wahba and Shalaby, 2009a). As pointed out by Nagel and Marchal (2003) there are many cognitive aspects that still need to be explored and embedded in models of adaptive behavior as risk evaluation, the importance of recent experiences and memory restrictions. Wang et al. (2010) investigated the performance of MILATRAS with that of two static FB-TAM that do not account for capacity constraints. Passengers boarding
and alighting flows at stops as well as on-board loads obtained by the three models were compared with passenger counts data from Toronto, Canada. The authors concluded that the disaggregate models (MILATRAS and MADITUC) outperformed the aggregate model (EMME/2).

It is evident that transit simulation studies vary considerably in the level of detail with respect to supply and demand aspects. The models can be classified as macroscopic, mesoscopic and microscopic with respect to transit modeling. The classification with regards to the supply side is straightforward as it is derived from the general classification used for traffic simulation models based on the level of traffic flow modeling. However, it is very common that the level of representation of different aspects of transit operations (e.g. dwell time, vehicle scheduling, control strategies, stop capacity) is inconsistent because of specific model interests. A microscopic representation of transit vehicle movements is potentially counterproductive to the modeling of higher-level transit operation processes as control strategies and fleet considerations.

A parallel classification of the demand side implies that macroscopic models consider passengers in terms of flows that are teleported from one location to the other based on high-level pre-trip routes while microscopic models represent individual travelers with their preferences and model their perception, behavior and detailed interactions at stops and on-board. A mesoscopic demand modeling corresponds then to an intermediate level that models individual travelers and their en-route path decisions without modeling their second-by-second movement.

2.3 TRANSIT PATH CHOICE MODELING
Route choice models describe how traffic demand is distributed over the traffic network. The complexity of transit path choice models emerges from the following reasons: non-continuous service availability; limited connectivity and the importance of

---

1 Throughout this thesis, the term route choice refers to car traffic/road network, while path choice is used within the context of transit networks. The term transit route is reserved for the list of links and stops that constitute a transit service and correspond to a certain transit line. Transit path consists of all the physical and conceptual elements that their ordered combination can describe a travel plan between two points in the network.
transfers; complex correlations between transit paths and; multi-modality and the intermediate walking links.

The next section reviews the development of choice modeling in the wider sense and four alternative modeling schemes. This is followed by a review of the two choice processes – choice-set generation and discrete choice in the context of route choice and transit paths in particular.

2.3.1 Choice Modeling

2.3.1.1 The Role of Choice-Set and Choice Model Classification

Decision making involves choosing an alternative from a set of alternatives. This set has to be mutually exclusive and collectively exhaustive (Ben-Akiva and Lerman, 1985). There are extensive research efforts concerned with the role of choice-sets in decision making. Although there is no agreed taxonomy for choice models, an investigation of this extensive body of knowledge suggests that it is useful to classify choice models according to two aspects:

1. Explicit modeling of the choice-set – whether the decision model consists of a single stage or two distinguished stages.
2. Type of decision protocol – The type of rules – compensatory or non-compensatory - used for carrying out the choice decisions. The former allows one attribute to compensate for another based on some trade-off relations. In contrast, non-compensatory decision protocols may not consider an alternative because it does not satisfy a certain attribute, regardless of other attributes’ values.

The combinations of these aspects define four classes of choice models (Figure 2.1). One-stage models obtain the final choice from a universal set by either applying a compensatory choice model that consists of a random utility model (RUM) or some kind of sequential screening process. In contrast, two-stage models compose an intermediate choice-set by applying either a compensatory or a non-compensatory choice-set generation model (CSGM). The exact rules vary between models within the same class. The following sections describe the four choice model classes and review the main model characteristics of each class. The purpose of this review is to highlight the various ways that the concept of choice-set may be incorporated within the framework of the choice model.
2.3.1.2 Non-compensatory one-stage models

The choice process can be considered as sequential screening of alternatives until a final choice decision is obtained. The underlying assumption is that decision makers tend to simplify complex choice situations through cognitive processes such as screening, prioritizing and satisfying. Models that belong to this class do not model the choice-set explicitly but rather consider it as an additional outcome of the choice process, sometimes referred to as endogenous choice-set generation (Richardson, 1982).

The theory of elimination by aspects (EBA) was developed by Tversky (1972). It represents the decision process as a successive elimination of alternatives based on a set of aspects or criteria and the corresponding weights. A procedure of sampling and enforcing aspects is iterated until only one alternative remains. Tversky suggested that most typically an aspect has the form of a critical threshold. Klein and Bither (1987) proposed to determine the threshold levels so that they will maximize the difference in utilities between alternatives. Hence, this method includes a compensatory component – determining threshold levels and ranking them by using utility function values – for undertaking a non-compensatory process. The researchers tested these hypothesizes in
various marketing decision context and the results supported the hypothesis that
decision makers use different threshold values in correspondence with utility function
differences. However, the influence of utility on threshold levels and aspects importance
was not captured correctly across different choice situations.

A simplified version of the EBA theory is undertaken by the lexicographic
strategy. It also implies that individuals choose an alternative by following a sequential
elimination process, but it does not require to rate alternatives according to their
relative importance but rather to rank them. Alternatives are sequentially eliminated by
following a hierarchical list of desired attributes. In the context of transit route choice,
Shih et al. (1997) applied the lexicographic strategy by using the number of transfers,
total trip time and service frequency as the list of aspects in a descending order.

EBA and lexicographic strategies involve the dynamic enforcement of criteria on
a set of alternatives. In contrast, Richardson (1982) proposed a decision protocol that
considers the alternative sequentially with a given set of criteria. The decision process
continues to screen alternatives until a satisfying alternative is chosen. The decision
maker may stop the search because the cost of an extra search is higher than the
expected gain from continuing the search. The author investigated the probability of
searching under different levels of prior information on the universal set and various
searching costs.

2.3.1.3 COMPENSATORY ONE-STAGE MODELS
This class implies RUM-DCM on the universal set of alternatives. The probability of an
alternative to be chosen depends on the RUM structure. It has the following general
form:

\[ P_n(i|\beta, X) = P_n(i|U, \beta, X) \quad \forall i \in U \]  \hspace{1cm} (2.1)

Where \( i \) is an alternative, \( \beta \) is a vector of coefficients, \( X \) is matrix of individual and
alternative attributes and \( U \) is the universal-set of alternatives.

DCM usually assumes that there is a pre-defined choice-set that is appropriate
for the considered decision situation. However, in many choice situations, it is
unreasonable to assume that the decision maker considers, evaluates and compares all
existing alternatives. Nevertheless, this assumption can be justified by arguing that if an
alternative has a low probability to be included in the choice-set, these attributes are
incorporated into its perceived utility and therefore results in a corresponding low
probability to be chosen.

25
The existence of a primary choice-set composition phase does not necessarily imply that these two phases cannot be integrated into a compensatory single-phase model. A laboratory experiment by Johnson and Meyer (1984) supported this notion. The prediction of the compensatory model remained robust across varying decision contexts and choice-set sizes, even though participants reported changes in the underlying screening techniques. Horowitz and Louviere (1995) argued that the utility function may contain all the relevant information about individual preferences, thus making the initial choice-set composition unnecessary. They tested their hypothesis by estimating models with different relations between the choice-set and the final choice phases using data on shopping choices. Their results suggested that indicators for inclusion in the choice-set reflect the same preferences that are included in the utility function. However, the researchers presumed that this may vary between choice’s contexts, with the two-phase choice process being more relevant for contexts with large number of alternatives that may be unknown to the decision maker. Both qualities are characteristics of transit route choice.

2.3.1.4 Semi-compensatory two-stage models

This class of models consists of two distinguished stages. The CSGM reduces the universal set to a compact choice-set that contains only the alternatives that are actually considered by the decision maker by applying non-compensatory methods. Then, a DCM assigns probabilities to each alternative in the set and concludes with the final choice (Figure 2.1). The role of the choice-set in these models is therefore to exclude alternatives from the choice model:

\[ P_n(i|\beta, X) = \begin{cases} P_n(i|C, \beta, X) & i \in C \\ 0 & i \notin C \end{cases} \]  (2.2)

Where \( C \subseteq U \) is the set of alternatives that resulted from the choice-set generation stage.

The screening rules have to dismiss alternatives that are irrelevant to the decision maker in the specific choice context. Recker and Golob (1979) emphasized the need to distinguish between the tolerance of an attribute and its importance. While the tolerance level is a requirement on the range of values that has to be fulfilled in the non-compensatory stage, the importance of an attribute is a relative weight that has to be incorporated within a compensatory decision protocol. They concluded that it is more
plausible that decision makers perform non-compensatory procedures at an initial phase in order to reduce their choice set followed by a compensatory choice phase.

The conclusions of Recker and Golob (1979) were reinforced by more recent studies of Gilbride and Allenby (2004), Elrod et al. (2004) and Cantillo and Ortuzar (2005). These empirical studies compared semi-compensatory two-stage models with compensatory one-phase model. In both cases the semi-compensatory model performed significantly better. Moreover, both studies found that some of the attributes were important only at one of the phases – playing a significant rule only at the screening phase or only at the compensatory phase. These findings support the argument for modeling the two phases separately.

Cascetta et al. (2002) proposed a two-stage model with a probabilistic choice-set. The so-called implicit availability/perception (IAP) model generates a master-set that is used for sampling a choice-set. The probability of an alternative in the master-set to be included in the choice-set is determined by a binary Logit model that is function of path attributes. The membership probability can be interpreted as a measure of perception that indicates the availability of the alternative from the decision maker perspective. Previous findings suggest that the screening rules exercise a conjunctive relationship (Gilbride and Allenby, 2004). A conjunctive choice process was also proposed by Cantillo and Ortuzar (2005) and Kaplan et al. (2011).

Dominancy rules make up a distinguished class of non-compensatory screening rules. Dominated alternatives are worse than another alternative in at least one aspect without been better than it in any other aspect. In this case, the dominated alternative (also known as Pareto non-optimal) is inferior to the dominating alternative, regardless of the individual preferences. For example, Androutsopoulos and Zografos (2009) incorporated dominancy rules in a transit route choice model by excluding all the alternatives that were dominated by other alternatives. In contrast, the dominancy attribute can be used in a compensatory model as a component in the utility function as was proposed by Cascetta and Papola (2009).

2.3.1.5 Compensatory two-stage models
Heterogeneity in preferences can be reflected at the choice-set generation stage as well as at the choice stage. Walker (2001) provides a comprehensive review of the various approaches for capturing heterogeneity among individuals in DCM. Andrews and Manrai (1998) argued that heterogeneity at the choice-set composition stage may be as
important as variation in the choice stage. Therefore, the CSGM does not have necessarily to result with a single deterministic choice-set. Instead, various choice-sets can be associated with corresponding probabilities. This class of models takes the following general form:

$$P_n(i|\beta, \gamma, X) = P_n(i|C, \beta, X) \cdot P_n(C|U, \gamma, X) \quad \forall i \in C, C \subseteq U$$  \hspace{1cm} (2.3)

Where $\gamma$ is a vector of parameters that is involved with assigning probabilities to choice-sets.

The number of possible choice-sets is theoretically defined by the combinatorial number of subsets that can be formed from the universal set. This yields the following expression (Manski, 1977):

$$P_n(i|\beta, \gamma, X) = \sum_{C \in M_n} P_n(i|C, \beta, X) \cdot P_n(C|M_n, \gamma, X)$$  \hspace{1cm} (2.4)

Where $M_n$ is the set of all non-empty subsets of the set, $U_n$ and $U_n \subseteq U$ is the set of all feasible alternatives at the specific decision context, sometimes known as a master-set. This formulation incorporates deterministic inclusion criteria (denoted by $U_n$) as well as a stochastic choice-set generation component. Swait and Ben-Akiva (1987) developed a probabilistic CSGM that incorporates both stochastic and deterministic constraints in terms of their critical value. Swait (1984) discussed the case of a maximum choice-set size but point out that there are no guidelines on how to determine the case-specific choice-set size.

Roberts and Lattin (1991) developed a compensatory model for the choice-set phase, although they proposed it as an approximation for the non-compensatory screening process. This analysis takes into account the trade-off between the costs associated with the information processing and the benefit in terms of potential perceived utility. The model is not applicable to choice contexts where different alternatives involve various complexity levels (e.g., as function of the number of transfers or timetables).

The probabilistic representation of choice-sets can account also for situational and latent factors at the individual level as well as random constraints at the choice-set generation process (Boccara, 1989). Ben-Akiva and Boccara (1995) developed a choice model with latent choice-sets. This integrative model implies that a change in an alternative's attribute will result in two effects: the availability of the alternative in the first phase and an impact on the attractiveness of the alternative in the second phase, in case it is available.
Following Horowitz and Louviere (1995), Swait (2001) suggested that choice-sets are only an expression of the corresponding alternative utilities, implying that individual preferences rather than constraints are the main factor in composing choice-sets. He developed a DCM that integrates choice-set generation into model specification.

2.3.2 Choice-set generation models

2.3.2.1 Choice-set generation model properties

Two-stage choice models involve the construction of a choice-set that includes only the alternatives that are actually considered by the decision maker. It can be conceptualized as a gradual process of obtaining the final choice-set from a universal set. Bovy and Fiorenzo-Catalano (2007) highlighted the theoretical and computational advantages in explicit generating of choice-sets prior to the route choice phase. Among their arguments they mentioned the ability to control the size and composition of the choice-set, and the flexibility in adopting various choice models. In addition, it has advantages at the implementation phase, in particular in the case that it is part of a dynamic assignment model. Bovy (2009) discussed modeling and implementation issues related to CSGM for transport networks. He argues for a non-compensatory CSGM followed by a compensatory choice from the choice-set. The CSGM is driven by constraints as well as preferences. Bovy concluded that it is unrealistic to apply a fully probabilistic approach for the CSGM due to the size of the master-set.

Several studies have used various notions of intermediate choice-sets (Bovy and Stern, 1990; Shocker et al., 1991; Hoogendoorn-Lanser and Van Nes, 2004). The intermediate levels may represent hierarchal levels of awareness or cognitive availability. Alternatively, they can be regarded as useful simplification that illustrates cut-off rules that can be imitated by the researcher when trying to reproduce the target choice-set. This functionalistic approach perceives the CSGM as an algorithm aimed to reproduce the choice-set of an individual or a group of individuals without attempting to imitate the cognitive process involved with the choice-set composition.

There are various methods for generating the choice-set in a transport network. Path search methods for general traffic networks have been studied extensively because of their importance to traffic assignment models. The most common methods are: K-shortest path algorithm, methods based on the labeling approach, link elimination, link
penalty, simulation methods and branch and bound procedures. For a comprehensive review of these methods, see Ramming (2002).

A comparison of CSGM was conducted by Bekhor et al. (2006) and Prato and Bekhor (2007). They compared the prediction accuracy across combinations of CSGM and route choice model. Their findings suggest that the heterogeneity of the choice-set is important for improving predictions. Moreover, the number of alternatives has a greater influence than the generation method on the estimation of route choice model parameters. Bekhor and Prato (2009) followed up on these results by developing a method to test the transferability of path generation techniques in the road network context. They concluded that path generation models are transferable at the model specification level and under certain conditions transferable at the parameters level, while path choice model were shown to be less transferable. Bekhor et al. (2008) analyzed the impact of choice-set size on the flow patterns obtained from a stochastic traffic assignment model. They concluded that larger choice-sets are associated with a lower value of the objective function at convergence. However, an increase in path-set size caused a linear increase in the computational time per iteration.

2.3.2.2 TRANSIT CHOICE-SET GENERATION MODELS

Generation methods

The main difference between road and transit networks is that the latter has limited connectivity and availability. These characteristics introduce a hindrance for applying the K-shortest path, link elimination and link penalty techniques. In addition, it requires taking into consideration transfer configuration and timetables. Previous studies adopted general path search methods and modified them for transit network applications. Branch and bound methods to construct a time-dependent choice-set with forward or backward search algorithms according to the planned departure time or arrival time, were developed by several authors (Tong and Wong, 1998; Friedrich et al., 2001; Huang, 2007). Tan et al. (2007) used a recursive search method which continuously searches for paths that involve more transfers. A dynamic programming method was applied by Androutsopoulos and Zografos (2009) to backtrack all possible paths between each pair of stops in the network. A connectivity matrix which represents network configuration by indicating if there is a direct path between each pair of nodes was used in several path search studies (Koncz et al., 1996; Su et al., 2008). Assuming that all transfers are taking place in major stations, Liu and Pai (2006)
developed a hub-based planning strategy that used the connectivity matrix to introduce transition matrices that yield the minimum number of transfers required between each pair of nodes and the number of possible ways to travel with a given number of transfers.

A simulation-based technique for generating a choice-set for a multi-modal network was applied by Carlier et al. (2003), based on the method proposed by Fiorenzo-Catalano and van der Zijpp (2001). The simulation-based method samples attribute coefficients and values from truncated normal distributions. The path that obtains the minimum generalized cost in each simulation draw is added to the choice-set. Bovy and Fiorenzo-Catalano (2007) applied a combination of the labeling approach and a simulation-based technique for generating a choice-set for a multi-modal corridor network. The method randomized link attributes and travelers’ preferences simultaneously.

Hoogendoorn-Lanser and van Nes (2004) applied the branch and bound method that was developed by Prato and Bekhor (2006) in the context of multi-modal transit networks. The method included logical and behavioral rules for determining the CSGM process. A constrained enumeration method was also applied by Friedrich et al. (2001) for a transit network. Guo and Wilson (2011) applied a labeling approach for generating a path set for the London underground network. The labels were chosen by the authors based on their ability to reproduce chosen paths.

Assessment of generated choice-sets

Only limited research has been devoted to compare those analytical procedures against transit path choice data (Bovy, 2009). The main impediment in evaluating CSGM lies in the unobservable and fuzzy nature of choice-sets and the difficulties in designing a suitable data collection method. Hoogendoorn-Lanser and van Nes (2004) conducted an extensive trip survey and interviews designed to collect data on consideration choice-sets on a multi-modal corridor. They compared the reported choice-sets with an analytical method. This method was later on extended to complete door-to-door trips (Hoogendoorn-Lanser et al. 2007). Van Nes et al. (2008) evaluated the capability of objective choice sets to replicate subjective choice-sets.

Guo and Wilson (2011) calculated the degree of overlap between the set of chosen paths and the set of generated paths obtained by a combination of labels. An
increase in the number of labels increased the probability that a chosen path is covered by the generated choice-set (effectiveness) as well as the probability that a path that was not chosen will be generated (efficiency). The application of stochastic simulation-based technique for the CSGM by Fiorenzo-Catalano et al. (2004) resulted in choice-sets that were larger than the set of chosen paths and reported choice-set by an order of magnitude. Bovy and Fiorenzo-Catalano (2007) emphasized the importance of applying a filtering process following their stochastic CSGM in order to remove largely overlapping paths and reduce the choice-sets size.

2.3.3 Random utility discrete choice models

2.3.3.1 Discrete choice models properties

Three of the four choice modeling schemes presented in Section 2.3.1 include the application of a RUM-DCM. A single phase of sequential screening is not feasible in the case of a large universal-set as is the case in the route choice context. RUM-DCM provides a rigorous theoretical background for representing traveler behavior. The stochastic representation of traveler perceptions is extensively used in the domain of route choice modeling - for a comprehensive review see Ramming (2002) and Prato (2009).

Different assumptions on the distribution of the random component in the utility function yield alternative model formulations. The widely-used Multinomial Logit (MNL) is derived from the assumption that the random components of the utility function are independent and identically Gumbel distributed (IID) (Ben-Akiva and Lerman, 1985). The violation of the assumption regarding independency between the random components of utility functions has long been studied. It is known as the red bus-blue bus problem or the independence of irrelevant alternatives (IIA) property. It is an important issue in the route choice context as path alternatives are not independent of each other due to overlapping.

Several models have been suggested in order to incorporate the dependency between alternatives into the choice model. Among these models there is the Nested Logit (NL) model which divides alternatives to groups and allows correlation between groups of alternatives but remains with the IIA property between alternatives in the same nest. Other models were developed in the particular context of capturing overlapping in road route choice context. The Path-Size Logit (PSL) model incorporates
into the systematic component of the utility function a correction term that represent
the degree to which it overlaps with other routes (Ben-Akiva and Bierlaire, 1999). The
path size factor can be defined in various ways to represent overlapping degree as
percentage of shared links by number, length or travel time (e.g. Menghini et al., 2009).
Ramming (2002) suggested a slightly different path size factor definition aimed to
decrease the effect of long paths included in the choice set. However, Frejinger and
Bierlaire (2007) showed that this definition is sensitive to minor changes in path
lengths and may lead to counter-intuitive results and therefore recommend the original
definition. Frejinger and Bierlaire addressed it by using the Error Component (EC)
model to capture explicitly correlations between sub-networks, a sequence of links that
compose part of the network.

The Cross Nested Model (CNL) developed by Vovsha and Bekhor (1998) revises
the NL structure to overcome the overlapping problem by allowing alternatives to
belong to more than one nest. Vovsha and Bekhor proposed to define the degree of
membership as the share of each link from the total route length or travel time. Swait
(2001) suggested the use of CNL for one-stage choice set generation and choice model
(termed GenL) by defining the choice sets as the nests so each alternative can belong to
several choice sets. However, Swait found it is unrealistic to estimate a model with more
than 5 or 6 alternatives as the model requires the estimation of a scaling parameter for
each possible choice-set. In contrast, Bekhor et al. (2008) showed that a complex route
choice model as CNL can be applied with reasonable efforts on a realistic-size network.
Bliemer and Bovy (2008) found that the size and the composition of the choice-set have
a large impact on route choice probabilities. The inclusion of irrelevant paths biased the
probabilities, in particular in case that the irrelevant paths have a large overlap with the
relevant paths. Even though they have tested RUM that account for overlapping such as
PSL and CNL, none of the models was found to be robust with respect to choice-set size.

2.3.3.2 Transit path choice models
Path choice factors

A large body of research in transport economics and transit demand modeling provides
estimations of parameter values that can be used in transit path choice modeling.
Wardman (2004) carried out a meta-analysis of the values of time in transit systems
based on empirical data from the UK and previous studies. The values of in-vehicle,
walking and waiting times can be used in the specification of a path choice model.
Horowitz (1981) investigated the perceived impediment involved with transfer and concluded that transfer penalties have to be taken into account on top of the respective time components. Since then, a large number of studies examined the determinants of transfer penalty. Iseki and Taylor (2009) provided a comprehensive review of these studies and their findings with respect to travel time components, transfer penalty and characteristics of the interchange facility. Guo and Wilson (2004) analyzed the determinants of transfer penalties by estimating various path choice models with interchange attributes and stop-specific constants. Raveau et al. (2011) and Guo (2011) enhanced transit path choice model by incorporating attributes of the transit map as additional explanatory variables. The topological distance and the representation of transfer locations were found to have a significant impact on path choice decisions. Comfort conditions were recently acknowledged as a decision factor (Hensher et al., 2003; Nuzzolo et al., 2011). In contrast, the importance of service reliability was rarely studied in the context of path choice models.

**Overlapping**

Only little attention had been given to the overlapping problem in the transit path choice context, as Schuessler and Axhausen (2007) point out in their review. Such an attempt to account for dependency between transit path alternatives had been done by Friedrich et al. (2001). They incorporated an independence factor into the probability function which was a function of departure and arrival times, total perceived journey time and trip fare. Such a dependence factor was introduced into the path utility function in VISUM (PTV, 2009, p. 430).

Another study of overlapping in transit networks was conducted by Hoogendoorn-Lanser et al. (2005) who applied PSL model on multi-modal trips. They used the path size factor definition of Ramming (2002) and compared three alternative overlapping definitions based on distance, expected travel time and number of legs by estimating the corresponding PSL models. They found that defining overlap in terms of number of shared legs yielded the best fit, suggesting that this definition best resembles passengers’ perceptions. Hoogendoorn-Lanser (2005) investigated the correlations between estimated transfer-related parameters. She found that the correlations between transit attributes cause problems in identifying and estimating the values of the corresponding parameters. Bovy et al. (2008) presented how PSL size factor can be
defined in case of multiple common links. Nevertheless, as they stress out, the PSL model assumes that all the independencies and correlations result from the geographical overlap.

**Dynamic path choice modeling**

The increasing use of traffic information systems in recent years have led to a growing interest in modeling route choice as a dynamic discrete choice process (Gao et al., 2008, 2009; Pel et al., 2009). Fosgerau et al. (2009) proposed an estimation procedure for a dynamic route choice model. Travelers are assumed to make a new route decision at each node and consider the utility associated with the remaining downstream route. This allows the consideration of relatively small choice-sets and the estimation of parameters at the link-level rather than the path-level.

Few studies have represented transit path choice as a sequential decision making process. Wahba (2008) modeled path choice decisions as a Markov process. Passengers shift between states by choosing the action that minimizes their generalized travel cost based on a deterministic utility function with the downstream trip attributes being determined by past experiences through a learning process. Nökel and Wekeck (2008) considered passenger alighting decisions under different levels of RTI. They discussed the impact of RTI on the choice-set structure and possible nesting structures for the alighting decision. They derived the share of alighting passengers at a certain stop in the context of a FB-TAM. A bounded rationality assignment model was proposed by Fonzone and Bell (2010) where travelers make a sequence of myopic decisions. Travelers’ expectations are assumed to be based on perfect information regarding the shortest path up to an intermediate location and a-priori expectations for the remaining path (based on average values or without any expectations). Hence, the model does not reflect the actual availability of information along passengers’ trip. Instead, the path choice is modeled as a rolling horizon with the choice of intermediate nodes based on stochastically minimizing the expected utility within the pre-defined planning horizon which was defined in terms of number of links.

2.4 Synthesis and Modeling Issues

Transit choice-set generation methods as well as route choice models often evolved by adapting models that were developed in the domain of road/traffic networks. Liu et al.
(2010) reviewed how the developments in transit path choice modeling are lagged behind the counterpart developments in road networks. Indeed, the two problems have important similarities and modeling issues as choice-set size, accounting for overlapping and estimation methods. However, there are also important differences which limit the transferability of developments in the road network sphere to the transit network sphere. The fundamentally different definition of path alternative has implications on the design of CSGM, correlation structure and the sequence of decisions. The transit path choice process has a distinctive discrete pattern which can be represented as a sequence of DCM. Very few studies have reflected this in their approach towards transit assignment modeling.

The centre of gravity changed during the evolution of TAM from the frequency-based approach to the schedule-based approach. The conventional static TAM cannot capture how the model evolves over time and the complex time-dependent interactions between system components. The modeling of variations in transit demand and supply conditions is essential for the analysis of transit operations and performance. Moreover, a dynamic representation is necessary in order to model the adaptive strategies exercised by both operators and travelers. The incapability of the conventional approaches to capture these dynamics may lead to unrealistic passenger loads.

A new generation of TAM represents individual travelers and is concerned with their dynamic path choice process rather than equilibrium conditions. Simulation models can facilitate the dynamic loading of travelers over a dynamic representation of the transit system conditions. The division in the transit modeling sphere between transit operations and transit assignment models is clearly reflected in previous developments of transit simulation models. None of the existing models represents both supply and dynamic aspects dynamically.

The following is a non-exhaustive list of the modeling issues that have to be considered in the process of developing a transit simulation model:

- Supply uncertainty – time-dependent variation and uncertainty on the supply side may be the result of exogenous processes or endogenous processes. Exogenous sources include traffic congestion, traffic incidents, weather conditions, a vehicle that breaks down, defected infrastructure or events that attract large crowds. Processes that are endogenous to the transit system are terminal operations and dispatching,
the bunching phenomenon, the implementation of control strategies such as holding, transit signal priority or short turning. The explicit modeling of these processes within a joint network framework will contribute to a more realistic reproduction of supply uncertainty rather than generating it based on statistical distributions and assembling independent models of separate system components. Emulating the dynamics of these sources will enable to analyze their impacts and potential methods to prevent them.

- Demand uncertainty – traveler decision making process is based on experience, information and anticipation. The findings of Fonzone et al. (2010) from an international survey on public transport suggested that the vast majority of transit travelers reported variations in their typical trips in at least one of the following aspects: departure stop, deviating from preferred line, transfer stop and changing line once on board. Travelers were more inclined to change stops than lines. An additional modeling issue is the portrait of an economic rational decision maker. This portrait could be enhanced by considering computational limitations, network-related knowledge and limited adaptation (e.g. Prato et al., 2011). Furthermore, passenger decisions are to a certain extent interdependent due to discomfort and capacity considerations.

- System capacity – transit capacity has to be analyzed within a wider perspective of various system components. The transit capacity and quality of service manual (TCQSM) (TCRP, 2003a) presents person capacity as the result of three factors: vehicle capacity, passenger demand characteristics and agent policies. Leurent (2011) identified seven classes of capacity issues: vehicle capacity of an infrastructure; vehicle fleet; passenger capacity of a vehicle; passenger capacity of a route; passenger capacity of a station; vehicle storage and movement capacity of a station and; capacity of a station with respect to access modes. TAMs that do not account for the dynamic effects of crowding conditions may result in unrealistic waiting times and passenger loads – both over- and under-utilized transit lines (Schmöcker et al., 2008).

These three modeling issues are obviously related since system capacities both depend and cause supply uncertainty. Passenger decisions are highly sensitive to service uncertainty and are also subject to capacity constraints. Moreover, the dynamics of the
interaction between travelers’ decisions and the transit system supply are an endogenous source for deteriorating service reliability.

The approach taken in this thesis is to model path choice decisions dynamically within the framework of an agent-based simulation model. The transit simulation model developed in this thesis brings together demand and supply sides and models their interactions dynamically. Passenger path choice process is modeled as a two-stage semi-compensatory model. Following previous studies in transit path choice modeling the large universal-set would be first reduced to a consideration-set. Traveler decisions would be modeled in the probabilistic framework of RUM-DCM. The following chapter provides the framework for the transit simulation model.
3. DYNAMIC TRANSIT MODEL FRAMEWORK

The performance of transit systems is a result of complex interactions between various system components. A dynamic perspective allows analyzing the way that system components evolve over time and their interactions under various conditions. This chapter is organized as follows: First, the components that have to be represented in order to capture transit system dynamics are discussed. Sections 3.2 and 3.3 elaborate on the modeling of traffic dynamics and transit operations, respectively. The different levels of demand representation are outlined in Section 3.4 as chapter 4 is devoted to the development of a detailed dynamic passenger model. Section 3.5 presents how these model components are integrated within the simulation model framework.

3.1 MODEL COMPONENTS
Transit systems consist of various components that interact through several processes. Following the objectives of this study, the core of the model consists of three main components: traffic dynamics, transit operations and traveler decisions (Figure 3.1). Modeling these components involves the dynamic movement of cars, transit vehicles in mixed-networks (e.g. bus, light rail, metro) and transit users. Each agent in the system carries out decisions, interacts with other agents and so affects the way the system evolves over time. This modeling approach is sometimes referred to as agent-based modeling (e.g. Salvini and Miller, 2005; Ettema et al., 2007). For example, traffic conditions that result in high travel time variability is associated with a reduction in transit service reliability. This in turn will affect transit users’ waiting times and crowding levels at stations and on-board. While the interactions between system agents are specific in time and space, the accumulated impacts of those interactions along transit lines and service times are not. Therefore, the modeling of transit systems also has to consider the evolution of the system over time at the network level. Hence, the developed model aims at analyzing transit systems at the network level.
APTS applications can affect the performance of transit system components. Control and information technologies are utilized for improving transit performance and the dissemination of real-time information (RTI). Real-time control strategies may regulate the service by enforcing transit signal priority (TSP) or holding strategies. Travelers may make different decisions based on their experience and the information available to them at various stages along their journey. The following sections describe how each of the above components is represented in the model.

### 3.2 Traffic Dynamics

Most transit services travel on mixed-traffic networks where traffic dynamics influence their performance. The interaction occurs both along road segments as well as at intersections. The level of traffic dynamics representation has to be on one hand detailed enough to allow the modeling of local interactions (e.g. at stops or

---

Figure 3.1: Transit model components
intersections) and on the other hand general enough to enable large-scale applications. Therefore, an intermediate ('mesoscopic') level of representation with regards to traffic dynamics was adopted in this model.

3.2.1 Mezzo
Mesoscopic traffic simulation models represent traffic dynamics at an intermediate level between microscopic models and macroscopic models. Macroscopic models represent traffic at an aggregated level based on flow-density functions without representing lanes or vehicles. In contrast, microscopic models represent traffic at a detailed level with explicit driving behavior characteristics of individual vehicles, such as lane changing, acceleration and gap acceptance. There is an inverse relationship between the level of detail and the computational effort that the model requires and hence its applicability to large-scale networks. Mesoscopic models offer a compromise between these two aspects by providing a useful trade-off between the level of detail on one hand and the ability to analyze at the system-wide level on the other hand. Mezzo represents individual vehicles, but models their progress in the network through speed-density relationships. This level of representation allows modeling the propagation of congestion dynamically as well as route choice decisions while avoiding the detailed modeling of vehicles' second-by-second movements.

The transit simulation model is built within the platform of Mezzo, a mesoscopic traffic simulation model (Burghout et al., 2006). Mezzo is an event-based simulation model, which incorporates an iterative dynamic traffic loading procedure. De Palme and Marchal (2002) argued that a mesoscopic simulation with an event-based architecture can outperform time-based microscopic models by one or two order of magnitude in terms of computational time. An overview of the traffic modeling in Mezzo is presented next. A complete description of the structure of Mezzo and its implementation details are presented in Burghout (2004).

Mezzo is an open-source program that has a modular structure due to an object oriented programming (OOP) approach designed to enable further enhancements and developments. Each entity in the simulation model (e.g. node, queue, vehicle, OD pair) is represented as an object with its related variables and functions. As an event-based simulation model, a list of events triggers actions according to the chronological order of the booked events. The model is programmed in C++.
The traffic simulation model requires input on network configuration (nodes, links, turning movements) and their corresponding characteristics (server capacities, speed-density functions, and historical travel times) as well as demand data (OD matrices and vehicle mix). The model generates output of time-dependent measures of performance at the link, trip and OD pair level. The output includes measures of effectiveness for each link and each time period (average speed, density, inflow, outflow and queue length), link travel times and results at the trip level (total travel time for each vehicle trip and aggregated at the OD pair level). A graphical user interface (GUI) animates the changes in selected performance measures for each link (Figure 3.2).

Figure 3.2: Mezzo GUI screen

3.2.2 Speed-Density Relationship

Links in Mezzo are divided into two parts: a running part, which contains vehicles that are not delayed by the downstream capacity limit; and a queuing part, which extends upstream from the end of the link when capacity is exceeded. The boundaries between the running and queuing parts are dynamic and depend on the extent of the queue. Vehicles enter the exit queue in the order they complete their travel in the running part. The earliest exit time is calculated based on the speed which is a function of the density in the running part only. Travel times on the running part are determined by a speed-density function. The default function is:

\[
V(k) = \begin{cases} 
V_{\text{free}} & k < k_{\text{min}} \\
V_{\text{min}} + (V_{\text{free}} - V_{\text{min}}) \cdot \left[1 - \left(\frac{k - k_{\text{min}}}{k_{\text{max}} - k_{\text{min}}}\right)^a\right]^b & k \in [k_{\text{min}}, k_{\text{max}}] \\
V_{\text{min}} & k > k_{\text{max}}
\end{cases}
\]

(3.1)

Where:

- \(V_{\text{free}}\) is the free-flow speed
- \(V_{\text{min}}\) is the minimum speed
- \(k\) is the density
- \(a\) and \(b\) are parameters
- \(k_{\text{min}}\) is the minimum density
- \(k_{\text{max}}\) is the maximum density

\[a = \frac{(V_{\text{free}} - V_{\text{min}})}{V_{\text{free}} - (V_{\text{free}} - V_{\text{min}}) \cdot \left[1 - \left(\frac{k_{\text{min}}}{k_{\text{max}} - k_{\text{min}}}\right)^a\right]^b}
\]

\[b = \frac{(V_{\text{free}} - V_{\text{min}})}{V_{\text{free}} - (V_{\text{free}} - V_{\text{min}}) \cdot \left[1 - \left(\frac{k_{\text{min}}}{k_{\text{max}} - k_{\text{min}}}\right)^a\right]^b}
\]
This speed-density function guarantees that vehicles move at free flow speed when the density is less than a given minimum threshold and have a minimal speed in case the density is above a certain boundary.

In the case of mixed-traffic, travel times of transit vehicles are calculated based on traffic conditions as for any other vehicle, albeit for the partial distance until the next stop or intersection. The explicit representation of background traffic allows capturing the impacts of congestion on transit operations. However, this representation requires the preparation of the relevant input as OD demand matrices, vehicle mix, route choice and preferences as well as additional information on network geometry. Alternatively, there is an option in the model to capture the effects of background traffic implicitly by representing link travel times as random variables with distributions that are derived from empirical travel times of transit vehicles. This is done by appropriately adjusting the values of link capacities and parameters of the speed-density function. A potential drawback of this simplified approach is that travel times on the various links are independent, and so do not capture the correlations among the travel times on neighboring links. Therefore, depending on available data and resources, and the scope of the application, the most appropriate method of representing background traffic can be used. While the explicit approach is clearly more realistic, the second is computationally more efficient and less demanding in terms of data requirements and input (only the distributions of travel times experienced by transit vehicles are needed).

The level of interaction with traffic conditions may vary depending on the existence of dedicated lanes and the associated usage rule (e.g. bus lane, bus way, tram tracks). The effect of a dedicated lane could be captured implicitly as empirical travel time distribution will reflect local conditions. An explicit representation of dedicated lanes calls for the introduction of restricted lanes and turning movements in Mezzo, which is an ongoing development.
As mentioned, Mezzo does not represent the second-by-second vehicle movements and hence cannot model the delicate interactions between individual cars and transit vehicles which typically occur at stop surroundings. In case the user is interested in representing different parts of the network in different levels of detail, the hybrid capabilities of Mezzo enables a microscopic-mesoscopic simulation. Previous studies have validated the model and demonstrated its capabilities in the context of incident conditions on a mixed freeway-urban network (Burghout et al., 2005).

3.2.3 Delays at Intersections
Travel times on the queuing part are determined by individual stochastic queue servers. At the downstream part of the link, the vehicles join a single queue of vehicles waiting to move out of the link. Queue servers process the vehicles in this queue and pass them on to the next link if it is not full. Each turning movement server searches backwards from the head of the queue for vehicles that intend to use the turn movement it regulates and processes them sequentially. A maximum queue look-back limit may be defined for each turning movement in order to represent dependencies among turning movements (e.g. when a queue in one movement blocks access to the lanes used by another turning movement). Service times generated by queue servers can follow truncated normal, lognormal or log-logistic distributions with the parameters corresponding to the saturation flow rate and the capacity of the respective turning movement. Give way is modeled on a turn-by-turn basis with each turning movement waiting for higher priority turning movements restricted by a maximum waiting time.

Mezzo is able to simulate fixed traffic signal plans. The signal plan input includes the cycle time, the starting green time of each phase and the corresponding duration and turning movements controlled by it. While the simulation model captures the crossing capacity constraint implied by the ratio of green time to cycle time, it does not prevent the simultaneous operations of conflicting movements.

3.2.3 Traffic Assignment
Vehicles in Mezzo are generated at mean rates specified by time-dependent Origin-Destination (OD) flow matrices. Vehicle generation follows a Poisson process - the interval between vehicle arrivals follows a negative exponential distribution. Vehicles are generated at their origin with a pre-defined destination according to the respective demand rate. Vehicle type is sampled based on a user-specified vehicle mix.
Stochastic dynamic user equilibrium (SDUE) assignment model loads vehicles to routes as shown in Figure 3.3. The procedure uses the shortest path algorithm to generate new route alternatives at the first loop which takes place at the initialization phase based on a random noise generator. The route search loop contains a SDUE loop so that new routes are generated only based on converged travel times.

![Diagram](https://via.placeholder.com/150)

**Figure 3.3: Dynamic assignment procedure in Mezzo**

Pre-trip route choices follow the MNL model with a set of pre-defined routes and historical link travel times. According to the MNL model, the probability that a driver will choose route $r$ is as follows:

$$P_r(t) = \frac{e^{V_r(t)}}{\sum_{j \in S} e^{V_j(t)}}$$ (3.2)

Where $V_r(t)$ is the systematic utility of route $r$ at decision time $t$, and $S$ is the set of feasible routes between the relevant OD pair.

Drivers can adapt their route choices en-route in response to information they receive. The en-route switching model is based on comparing the expected travel time on the shortest route alternative and the expected travel time on the remainder of the current route, both based on the updated information available on network conditions. Again, the MNL model defines the probability for each possible decision.
3.3 Transit Operations

The Transit model, BusMezzo, is completely integrated within Mezzo, the traffic simulation model. Transit-related components were developed and introduced into the simulation model in order to enable the dynamic representation of transit operations. This section describes the key concepts involved with the supply side of the transit system: timetables, dwell times, and vehicle scheduling.

3.3.1 Transit Network

The transit network layer lies on top of the fundamental layer of the relevant physical networks of roads and railways. The transit network layer includes routes, lines, stops and the corresponding timetables. Transit routes are defined by a list of links, while transit lines are defined by their origin and destination terminals and the sequence of stops that are served in between. Stops are located on links and may be assigned to several lines. The walking distances between stops may be given to enable stop choice and transfers, without the explicit representation of the pedestrian network.

The timetable plays the role of a reference point and may be used for control (e.g. dispatching, holding at time points) and the calculation of measures of performance (e.g. on-time performance, deviation from schedule). The model aims at accommodating timetables of different formats – the complete timetable with the scheduled departure time from each stop along the route for each transit trip; fixed intervals between stops with time-dependent headways or; fixed intervals between stops with fixed headways.

3.3.2 Dwell Time

Transit trip travel times consist of two parts: riding times (travel time between departure time from stop \(s\) and arrival time at stop \(s+1\)) and dwell times. Dwell time is the additional travel time caused by the need to serve a stop. It includes the time needed to get off traffic and enter the stop, opening the doors, boarding and alighting of passengers, closing the doors and getting back to traffic. Note that this definition does not include the time spent at stops due to operator regulations or control.

It is important to model the dwell time in detail as it accounts for a large share of transit travel time. Based on field surveys that were conducted in several US cities, Levinson (1983) concluded that dwell times contribute 9-26% of the total travel time, compared with 12-26% spent in traffic delays. In a recent study, Bertini and El-Geneidy (2004) analyzed disaggregate AVL and Automatic Passenger Counts (APC) data from
the Portland metropolitan area. They found that 16% of the total bus travel time is spent with open doors at stops. The total time lost due to serving stops was estimated at 33% of the total travel time. Furthermore, dwell time is an important source of unreliability as it causes high variability with a coefficient of variation in the range of 0.6 to 0.8 (TCRP, 2003a). Thus, an accurate model for estimating dwell times is important in order to capture their impact on service reliability (i.e. bunching), level of service, and operational efficiency.

The importance of dwell times to transit operations has led to an extensive research of its factors. Many dwell time studies used regression models to estimate the independent variables. The main factors influencing the dwell time can be grouped into stop characteristics, vehicle characteristics and passenger volumes. In addition, the payment method and the boarding regime are important determinants of the passenger service process. Figure 3.4 presents the main determinants of dwell time reported in the literature and the relations among them. Arrows indicate an effect relation. For example, boarding regime and door configuration determine the delay caused by a boarding passenger. Some of these factors can be accounted for by adding constant components (e.g. bay/in-lane stop; lost time related to vehicle characteristics), while others affect the fundamental structure of the dwell time function (e.g. door configuration and boarding regime which determine the marginal contribution of an additional boarding or alighting passenger).

Figure 3.4: Dwell time determinants
The transit operations model in BusMezzo enables the user to define various dwell time function structures and specify the respective coefficient values. All of the determinants mentioned in Figure 3.4 are addressed explicitly in the following discussion. The general form of the dwell time function of vehicle type $f$ on trip $k$ of line $l$ at stop $s$ is:

$$D^{k,f}_{s,l} = \text{lost}_{s}^{f} + \text{PST}_{s,l}^{k,f} + v_{s,l}^{k,f} \quad (3.3)$$

Where:

- $\text{lost}_{s}^{f}$ - a constant delay associated with stop $s$ and vehicle type $f$
- $\text{PST}_{s,l}^{k,f}$ - total passenger service time
- $v_{s,l}^{k,f}$ - error term

The error term captures the stochastic nature of passenger and driver behavior. The fixed delay that is independent of passenger activities can be attributed to various factors related to vehicle type and stop characteristics. The current implementation of the dwell time function includes the following factors:

$$\text{lost}_{s}^{f} = \beta_0 + \sum_{s,\text{type}} \beta_t \cdot \delta_t^s + \sum_{v,\text{type}} \beta_f \cdot \delta_f + \beta_c \cdot \delta_c^s \quad (3.4)$$

$\beta$'s are parameters and $\delta$'s are indicators. $\beta_0$ is the fixed delay that is not attributed to any of the other factors. $\beta_t$ and $\beta_f$ are the delays associated with stop type $t$ and vehicle type $f$, respectively. For example, a bay stop imposes a longer entering and getting back to traffic times. In addition, transit vehicle types vary in their fixed dwell time component due to their acceleration and declaration speeds and their doors mechanics. $\delta_t^s$ and $\delta_f$ are the corresponding indicators that take the value one if the stop or the vehicle belong to the relevant type and zero otherwise. $\beta_c$ is the delay caused by stop capacity limitations (e.g., vehicle stopping outside the stop area or waiting for a vacant space) where $\delta_c^s$ indicates whether stop $s$ is fully occupied when the vehicle on trip $k$ approaches to enter the stop. This incorporation of stop capacity impacts avoids the micro-modeling at the stop area level. If there is no data to support the inclusion of some of the above factors, the respective coefficient can be set to zero.

The passenger service time is the main component of the dwell time function as it is subject to high variations in space and time. Deuker et al. (2004) concluded, based on a regression analysis of a large AVL and APC database, that the number of boarding and alighting passengers was the most significant determinant of dwell time. Passenger
service time can be simply modeled as the sum of processing times of the two passenger streams:

$$PS_{s,l}^{k,f} = \beta_a^f \cdot A_{s,l}^k + \beta_b^f \cdot B_{s,l}^k$$  \hspace{1cm} (3.5)

Where:

- $A_{s,l}^k$ - number of alighting passengers at stop $s$ on trip $k$ of line $l$
- $B_{s,l}^k$ - number of boarding passengers at stop $s$ on trip $k$ of line $l$
- $\beta_a^f$ - service time per alighting passenger from vehicle type $f$
- $\beta_b^f$ - service time per boarding passenger on vehicle type $f$

The linear form of expression 3.5 is appropriate when the two processes – boarding and alighting – take place one after the other. Moreover, each additional passenger has the same marginal contribution to the dwell time. In practice, boarding and alighting are a parallel rather than a sequential process. Hence, the process that takes longer time determines the resulting passenger service time, as follows:

$$PS_{s,l}^{k,f} = \max \{\beta_a^f \cdot A_{s,l}^k, \beta_b^f \cdot B_{s,l}^k\}$$  \hspace{1cm} (3.6)

This calculation assumes that boarding and alighting passengers are distributed evenly between an equal number of doors. More generally, the door that dominates the service time determines the total passenger service time (Lin and Wilson, 1992):

$$PS_{s,l}^{k,f} = \max_d\{PS_{s,l}^{k,f,d}\}$$  \hspace{1cm} (3.7)

With this function, the time needed for passengers to board and alight is calculated separately for each door, where $PS_{s,l}^{k,f,d}$ is the passenger service time at door $d$. The door-specific terms depend on the exact door configuration (number and location of doors) and boarding regimes (e.g. boarding only from the front door) as these factors influence how boarding and alighting passengers are distributed between doors. Hence, the expression above can accommodate various conditions. Note that expression 3.6 is a special case of expression 3.7 with two doors that are used exclusively for either boarding or alighting streams.

An additional common case that is implemented in the simulation model is based on the dwell time function adopted in the Transit Capacity and Quality of Service Manual (TCRP, 2003a). The function refers to the case of two doors ($d \in \{\text{front, rear}\}$) where boarding is allowed only from the front door and alighting passengers are distributed between the doors. The door-specific terms are:

$$PS_{s,l}^{k,v,\text{front}} = \beta_a^f \cdot p_{\text{front}}^v \cdot A_{s,l}^k + \beta_b^f \cdot B_{s,l}^k + \beta_c \cdot \delta_{s,k,v}^c \cdot B_{s,l}^k$$  \hspace{1cm} (3.8)
\[
PS_{s,l}^{k,v,\text{rear}} = p_a^f \cdot (1 - p_a^{\text{front}}) \cdot A_{s,l}^k
\]  
(3.9)

Where:
- \( p_a^{\text{front}} \) - fraction of passengers that alight from the front door
- \( \beta_c \) - extra delay per boarding passenger due to crowding conditions on-board
- \( \delta_{s,k,v}^c \) - crowding indicator that takes the value of one if the number of passengers on-board exceeds the number of seats and zero otherwise, or formally:
  \[
  \delta_{s,k,v}^c = \begin{cases} 
    1 & \text{if } L_{s,l}^k > \text{seats}^f \\
    0 & \text{otherwise}
  \end{cases}
  \]

Where:
- \( L_{s,l}^k \) - passenger load on trip \( k \) of line \( l \) when approaching stop \( s \)
- \( \text{seats}^f \) - number of seats on vehicle type \( f \)

The TCRP function is consistent with the conclusions of Puong (2000) that the crowding factor – delay caused by standees - affects only boarding passengers. The importance of incorporating the effect of crowding was emphasized by the analysis of Puong that the crowding factor explains 90% of the dwell time variation. The effect of crowding in the TCRP function is zero as long as passenger load is smaller than the number of seats and increase linearly with the number of boarding passengers when it is above the number of seats. In addition, this effect is independent of vehicle capacity which implies that the delay imposed by an additional standee is the same regardless of the remaining capacity on-board. In contrast, Weidmann (1994) suggested a non-linear crowding factor that is based on the remaining capacity for standees. Unlike the dwell time function of the TCRP, he claims that the crowding factor impacts both boarding and alighting passengers. Weidmann’s formulation is in line with the findings of Lin and Wilson (1992) that the crowding effect is non-linear and that crowding affects boarding and alighting rates identically. The total passenger service time is given by:
\[
PST_{s,l}^{k,f} = \left[ \beta_a^f \cdot A_{s,l}^k + \beta_b^f \cdot B_{s,l}^k \right] \cdot \left[ 1 + \frac{3}{4} \left( \max \left\{ 0, \frac{l_{s,l}^k - \text{seats}^f}{\text{cap}^f - \text{seats}^f} \right\} \right)^2 \right] 
\]  
(3.10)

Where \( \text{cap}^f \) is the passenger capacity on vehicle type \( f \). Note that this expression implies that the impact of crowding on passenger service time is in the range of \([0,0.75] \) with an increasing marginal affect. This crowding effect can be added to either expression 3.6 or 3.7.

An additional complication involved with the calculation of the dwell time is the inter-dependence between the number of arriving passengers and the dwell time as
passengers arriving during the dwell time may board the vehicle. Therefore, the dwell time model includes an internal feedback loop that re-evaluates the number of arriving passengers during the previous dwell time. This process continues until no passenger boards the vehicle during the marginal dwell time – because no additional passenger arrived or due to capacity constraints. This is equivalent to the formulation of Bellei and Gkoumas (2010).

All the above dwell time structures are implemented in BusMezzo. The user can specify the values of the relevant coefficients. These values can be obtained from manuals, the literature or calibrated based on local measurements. The model enables to define several dwell time functions and assign them to different vehicle types and stops.

Furthermore, the delay at certain stops can depend on the possibility to overtake other vehicles that use the stop. When a transit vehicle is ready to depart from a stop it may be subject to delays caused by the inability to maneuver and overtake the vehicle in front that is not ready to depart yet. This blocking phenomenon is typical to tramlines and bus ways. The user can specify at which stops vehicles may be unable to overtake at stops. At these stops, the departure time of trip $k$ from stop $s$ is as follows:

$$ET_{s,l}^k = \max (AT_{s,l}^k + DT_{s,l}^k, ET_s^k)$$

(3.11)

Where:

$AT_{s,l}^k$ - arrival time of trip $k$ on line $l$ at stop $s$

$ET_s^k$ - exit time of the trip that stands at stop $s$ and blocks the vehicle that carries out trip $k$

Hence, the exit time is determined by the later between two expressions: the time the vehicle is ready to depart and the time that the vehicle in front departs. It is assumed that only the vehicle that is directly adjacent may prevent a vehicle from exiting. Note that the exit time may be subject to control strategies that may hold the vehicle at the stop until a certain criteria have been fulfilled, as discussed in Section 3.3.4.

3.3.3 VEHICLE SCHEDULES

Transit vehicles follow a schedule that consists of a sequence of trips. This schedule may include also deadheading, which are non-service trips between terminals. Each trip has origin and destination stops. In the case of transit vehicles, the route between origin and
destination is pre-defined. The destination of a certain trip is the origin of the next trip on the vehicle schedule.

The ability to simulate the chain of trips that a vehicle undertakes allows modeling the propagation of delays. It is an important property of the model as it enables to capture the dependency between successive trips through the propagation of delays from trip to trip. At the system-level, the model can be used to analyze recovery and layover time policies and the relationship of fleet size and assignment to the resulting level of service. Most transit operations models reported in previous literature do not model vehicle schedules (e.g. Lee et al., 2005; Fernandez, 2010), albeit MITSIMLab does (Morgan, 2002). It should be noted that crew scheduling is not modeled.

The construction of schedules involves the determination of layover and recovery time factors. Recovery time serves as a buffer between successive parts of the ride, in order to avoid the propagation of delay along the trip. The allocation of the recovery time can be at the end of the trip, distributed along the route or a combination of both (TCRP, 2000). The distribution of the recovery time along the route based on a certain percentile of the trip travel times is a useful way to distribute the unreliability risk both from operator and passenger perspectives. Moreover, whenever there is a control mechanism along the route in place, the recovery time has to be incorporated into the timetable. This is indeed the common practice among bus operators (Ceder, 2007). The commonly-used value is the 85th percentile of trip travel time distribution (TCRP, 2000). Therefore, it is assumed that the recovery time is incorporated into the timetable. In addition, the Layover time allows drivers to rest between successive trips.

The layover and the recovery times are typically combined and have to balance between two contradictory objectives: high reliability that requires long recovery times and high efficiency that requires the smallest possible margins between successive trips. According to Strathman et al. (2002) there are three common criteria for determining the recovery time: (1) Levinson's rule – the difference between the mean or the median and the 95th percentile running time; (2) the rule of thumb – 18% of the median running time or; (3) based on labor rules or contract conditions.

The departure time of trip $k$ of line $l$ from the origin terminal (also known as dispatching time) is calculated as the latter between the scheduled exit time and the
time the vehicle is available to depart after it completed its previous trip and some recovery time:

\[ ET_{s,l}^k = \max (SET_{s,l}^k, AT_{s,l}^{k-1} + RT_{\min} + \varepsilon_{s,l}^k) \]  \hspace{1cm} (3.12)

Where:
- \( RT_{\min} \) - pre-defined minimum recovery time
- \( \varepsilon_{s,l}^k \) - error term that represents the stochastic part of the recovery time

The error term captures factors as driver adherence to the timetable and supervision errors.

### 3.3.4 Control and Management Strategies

Transit control strategies consist of a wide variety of operational methods aimed to improve transit performance and level of service. APTS are increasingly integrated into transit systems, enabling improved management and operation strategies that incorporate real-time information (FTA, 2000 and 2006). BusMezzo was designed to support the implementation of control and management strategies.

Holding control strategies are among the most widely used transit control methods aimed to improve service regularity and service coordination by regulating departure time from stops according to pre-defined criteria. The strategy contains a set of rules that determine at which stops along the route departure times will be subject to regulation (those stops are known as time points) and which criteria are used for determining the departure time. The criteria may depend on the timetable or real-time location and occupancy of transit vehicles. Evaluating the effects of holding strategies and assessing different holding designs requires a dynamic representation of complex interactions between stochastic processes, in particular when considering holding strategies that are based on real-time information. BusMezzo supports the implementation of control strategies at stops. Every time that an activity of an entity that is potentially subject to control (e.g. time point stop) is taking place in the simulation model, the model checks whether the control strategy is valid and enforces it. In the presence of holding strategies, the departure time from the stop is given by:

\[ ET_{s,l}^k = \max (AT_{s,l}^k + DT_{s,l}^k, CT_{s,l}^k) \]  \hspace{1cm} (3.13)

Where \( CT_{s,l}^k \) is the time resulting from the control strategy implemented for line \( l \) on trip \( k \) at stop \( s \). Various holding control strategies were applied in the simulation model. Chapter 6 is dedicated for discussing their details and related case studies.
Another set of control and management strategies include fleet management schemes as expressing, short turning and deadheading (Ceder, 2007). Express services stop only at a subset of the stops along the route, vehicles that are subject to short turning run only on a partial route and deadheading vehicles run empty to increase the capacity of a certain route, as illustrated in Figure 3.5. The interactions between fleet management schemes and the level of service can be directly assessed in BusMezzo by running the simulation with different pre-defined lines and vehicle schedules.

Figure 3.5: Illustration of fleet management schemes: (A) Normal operations; (B) Expressing; (C) Short turning; (D) Deadheading;

An additional class of control strategies is TSP. Signal plans can give priority to transit vehicles by static allocation of longer green times to transit corridors. Active signal plans can provide priority by adjusting the signal plans when a transit vehicle approaches the intersection. An active approach may be unconditional – carrying out the same adjustment for each approaching transit vehicle – or conditional – different adjustments depending on real-time conditions as delay, headway or on-board occupancy. The evaluation of various TSP schemes requires a detailed representation of traffic dynamics. Morgan (2002) studied the effects of conditional TSP on a single corridor in Stockholm by integrating a microscopic bus operations model within a microscopic traffic simulation model, MITSIMLab. The hybrid version of Mezzo was used for the evaluation of adaptive transit signal control schemes in a microscopic scale.
(Burghout and Wahlstedt, 2007). Currently, BusMezzo does not support the implementation of TSP.

3.4 PASSENGER DEMAND
Passenger demand can be represented at various levels of detail depending on the application of interest and data availability. Regardless of the level of demand modeling, passenger demand is time-dependent and varies in space. All of the levels of demand representation discussed below assume that passengers arrive randomly at stops. This assumption is valid only in the context of urban transit systems with high-frequency service. However, it is violated in case of a low-frequency service as previous studies have demonstrated (e.g. Abkowitz and Tozzi, 1987; Seneviratne, 1990). The threshold between a random arrival pattern and a non-random pattern is around 10-12 minutes headway. Longer headways will cause an increasing share of the passengers to follow the timetable.

In a review of dynamic traffic assignment (DTA) models, Peeta and Ziliaskopoulos (2001) pointed out that due to computational tractability concerns, an approach that combines a microscopic representation of individual trip makers with a macroscopic representation of some traffic dynamics has an advantage in DTA modeling. The aim of this study is to undertake a similar approach and enable this combination in the context of transit modeling. The dynamic path choice model (DPCM) that is in the core of the detailed demand modeling is discussed in the following chapter.

3.4.1 LEVELS OF DEMAND REPRESENTATION
The level of demand representation in BusMezzo can be classified as follows:

- Boarding and alighting rates per stop and line
- OD matrix per line
- OD matrix at the stop level
- OD matrix at the spatial unit level (e.g. activity anchors, zones)

The following describes each level of demand representation and the corresponding modeling capabilities and input requirements.

1. **Boarding and alighting rates** – passengers are modeled in terms of volumes and have an origin stop and a target line. The simulation model keeps track of the number of
passengers at stops and on-board. The number of passengers arriving at each stop is modeled as a stochastic process that follows a pre-defined distribution (e.g. Poisson process) with line-specific parameters. The number of alighting passengers is calculated as a function of the number of passengers on-board and the probability to get off the vehicle at a given stop. This level of modeling requires two input components: average passenger generation rates during time period $\tau (\lambda_{s,l}^{T})$ and the probability to alight at stop $s$ from line $l$ during time period $\tau (p_{s,l}^{T})$. The model adopts the approach of most studies of these processes that assume that passenger arrivals follow the Poisson distribution (e.g. Fu and Yung, 2002; Dessouky et al., 2003) as was confirmed by empirical studies (e.g. Bowman and Turnquist, 1981; Rajbhandari et al., 2003). Therefore, the number of arriving passengers is the product of the passenger generation rate and the headway from the leading vehicle:

$$c_{s,l}^{k} \sim \text{Poisson}(\lambda_{s,l}^{T} \cdot h_{s,l}^{k})$$

(3.14)

Where $c_{s,l}^{k}$ is the number of arriving passengers and $h_{s,l}^{k}$ is the time headway between the preceding bus (on trip $k - 1$) and the bus on trip $k$. Trip $k$ arrives at stop $s$ during time period $\tau$. In case that the passenger arrival process takes place over periods with different generation rates (i.e. the relevant headway is spread over two or more time periods), the number of passengers wishing to board is calculated as the sum of the generations in these time periods. Since passengers have pre-defined transit lines, the number of arriving passengers for each line is calculated independently of other lines.

The alighting probability determines the expected number of passengers that will take off at a given stop. Assuming that each passenger has probability $p_{s,l}^{T}$ to alight, the distribution of the number of alighting passengers is Binomial. This representation of the alighting process was used in previous studies (e.g. Liu and Wirasinghe, 2001; Morgan, 2002; Bellei and Gkoumas, 2010). The number of alighting passengers is proportional to the number of passengers on-board when approaching stop $s$, $l_{s,l}^{k}$:

$$A_{s,l}^{k} \sim \text{Binomial}(l_{s,l}^{k}, p_{s,l}^{T})$$

(3.15)

The number of passengers boarding the arriving vehicle, $B_{s,l}^{k}$, is restricted by the remaining capacity on-board, after the number of alighting passengers is taken into consideration:

$$B_{s,l}^{k} = \min (c_{s,l}^{k}, cap_{l}^{f} - l_{s,l}^{k} + A_{s,l}^{k})$$

(3.16)
Passenger boarding process is always subject to capacity constraints in BusMezzo, while enforcing the first-in-first-out (FIFO) regime. Hence, passengers that cannot board a vehicle are left behind and wait for the next vehicle. Note that both \( A^k_{s,l} \) and \( B^k_{s,l} \) are a function of \( L^k_{s,l} \). The number of passengers on board is a state variable that is updated by subtracting the number of alighting and adding the number of boarding:

\[
\mu^k_{s+1,l} = L^k_{s,l} - A^k_{s,l} + B^k_{s,l}
\]  

(3.17)

Where the boundary value is typically set to zero.

2. **OD matrix at the line level** – passengers have a pre-defined destination stop on a specific line. The model keeps track of the number of passengers on-board destined to each stop along the line. The number of passengers arriving at a stop is the sum of stochastic processes that corresponds to generation rates per line and downstream stop. Therefore, the demand is given by time-dependent matrices of average generation rates. Each arrival process is determined by the generation rate at stop \( s_o \) that is destined to downstream stop \( s_d \) on line \( l \) during time period \( \tau (\lambda^l_{s_o,s_d}) \). The number of passengers that get off the vehicle at every stop is simply the number of passengers on-board that have this stop defined as their destination. By segmenting passengers to stop destinations, passenger loads can be replicated more accurately as well as it implications on dwell times and capacity.

3. **OD matrix at the stop level** – trips are initiated at an origin stop and passengers have to choose their path to a pre-defined destination stop. Demand is given as time-dependent generation rates at stop \( s_o \) that is destined to downstream stop \( s_d \) during time period \( \tau (\lambda^l_{s_o,s_d}) \). The total number of waiting passengers is defined by the sum of stochastic arrival processes to various destination stops.

This level of representation requires DPCM and the modeling of individual passengers that undertake successive decisions. Passengers may have heterogeneous preferences over modes, transfers, comfort and the trade-off between various paths’ attributes. This level of demand modeling enables to study the interaction of passenger decisions and transit performance. Modeling individual passengers allows to analyze and to evaluate in detail the level of service and the impacts of various transit service conditions and information provision.

4. **OD matrix at the spatial unit level** – passenger demand is originated and destined at the production and attraction sources of activities. Therefore trips are
initiated at a certain spatial unit (e.g. zone, block, dwelling unit) and have a corresponding destination unit. This level of representation is consistent with activity-based modeling as trips are associated with the actual locations where individual performs their activities. It could be potentially enhanced by assigning to individuals a set of trips that correspond to a sequence of activities that need to be carried out sequentially. Passengers can make rescheduling decisions at every node of the transit system or when acquiring new information, in a similar fashion to the decision trees proposed by Sun et al. (2009).

This level of demand representation models explicitly the origin and destination stop decisions - passengers choose at which stop to initiate their transit trip and at which stop to finish it before reaching their final destination. Note that this can also be modeled with the previous level of modeling for a limited set of origins and destinations as centroids of traffic analysis zones (TAZ). This can be achieved by defining the walking distances between artificial stops and transit service stops.

3.4.2 DATA REQUIREMENTS
Transit demand data collection methods are undergoing fundamental changes with the rapid development and implementation of APTS. Traditional collection methods as travel habit survey, manual passenger counts and on-board surveys are integrated or replaced with location-based automated data collection methods. An increasing number of transit systems are deployed with APC and Automatic Fare Collection (AFC) as described in the state-of-the-art reviews of the FTA (2000, 2006). These systems enable the automatic collection of large data volumes with an extensive spatial coverage and continuous samples which will be practically impossible to collect through surveys due to the costs involved. The specification of an OD matrix in terms of stops requires detailed demand data that is only seldom available to transit authorities, although the increasing use of automatic data collection methods may change it in the near future.

In an overview paper of the state-of-the-art of smart card data use, Pelletier at al. (2011) highlight the mass data collection and the need to accommodate it in a disaggregate modeling approach. Detailed ridership statistics and their time and spatial distribution may facilitate the generation of detailed OD matrices. Furthermore, smart card data may provide transfer data that is otherwise available only by surveying passengers. An additional advantage of smart card data is the possibility to analyze transit user segments based on card holder information. Uniman et al. (2010) illustrated
how smart card data facilitates the analysis of reliability measures that are based on OD passenger flows at the single-line level.

The four different levels of passenger demand modeling listed above impose different degrees of data requirements. When passenger demand is represented in terms of boarding and alighting rates, input data can be obtained from passenger counts or APC measurements, without linking the origin and destination stops at the single traveler level. In the case that an OD matrix at the line level is required, boarding and alighting data have to be linked. This information may be collected via an on-board survey or in the rare case that the AFC system requires tapping both when getting on and off a vehicle as is the case in Seoul (Jang, 2010). However, in many cases only incomplete boarding and alighting data or passenger load data is available. Kikuchi and Kronprasert (2010) proposed a method for constructing OD tables at the line-level based on fragmented passenger counts and subjective knowledge on ridership patterns along the line. The method exploits the relationship between transit line OD parameters to formulate flow conversation equations.

Modeling the path choice decisions of individual passengers requires a corresponding higher level of demand data. Transit agencies typically construct OD matrices at the zonal level (TAZ) based on travel habit surveys and the four-steps planning model. Assignment tools as TRANSCAD and EMME allow the extraction of an OD matrix at the stop-level from the assignment results. The obtained matrix incorporates the static model results with respect to origin and destination stop choices. Furthermore, GIS-based applications that integrate a layer of transit network and a layer of land uses can be used to analyze transit demand in high resolution (e.g. Benenson et al. 2010).

Considerable research is currently given to the potential utilization of automated data collection methods for the estimation of transit OD matrices. In the common case that AFC data contains only boarding data, it is necessary to apply a method for estimating the destination location. Several papers discussed methods to interpret the boarding and alighting data in order to compose transit stop-level OD. Boarding stops are inferred by either matching the APC data with AVL data (Wang et al., 2011; Munizaga et al., 2011) or searching for the closest stop distance-wise on the reported route (Nassir et al., 2011). The inference of alighting stops is far from straight-forward. The common approach is based on a set of assumptions with regards to trip-chaining.
patterns. It is assumed that a daily trip sequence is made up of transit trip segments only, with no other modes in between. All trip segments that belong to the same card holder are stored chronologically. The alighting stop is then inferred from the boarding location of the next trip segment by searching for the closest stop on the specific route. Closest here could be in terms of distance (Trepanier et al., 2007) or a generalized time expression (Munizaga et al., 2011). Transfers are identified by setting a certain constraint on the time difference between consecutive alighting and boarding records. Trepanier et al. (2007) proposed that in case that the alighting location cannot be inferred from the directly adjacent trip segments, day-to-day patterns can be used.

Previous studies vary in the additional constraints or filtering rules and the identification heuristics that they embedded for composing the complete trip-chain. In particular, various upper bounds on the distance and time proximity allowed between consecutive trip segments in order to distinguish between transfers and activities. The OD matrices obtained from these methods were compared with the corresponding survey matrix or APC data (Munizaga et al., 2011; Wang et al., 2011; Nassir et al., 2011). The methods for composing an OD matrix based on AFC systems cannot handle single trips – trip segments that have no consecutive trip – and are of course limited to users that use a smart travel card. Moreover, the assumptions involved with the interpretation of boarding and alighting locations may introduce inaccuracies in the obtained OD matrix.

Personal mobile devices may enable to estimate OD matrices, especially with the rapidly increasing market share of smart phones. Kostakos et al. (2010) demonstrated how Bluetooth technology can be used for detecting the boarding and alighting stops of single passengers by placing a Bluetooth sensor on-board a transit vehicle. They matched the Bluetooth records with the AVL data so they could identify the boarding and alighting stops. Personal mobile devices could also be used for estimating the OD matrix at the zonal level. Chung and Kuwahara (2007) compared an OD matrix estimated based on mobile phone and GPS probes to the one obtained by Tokyo metropolitan from a travel habit survey. They concluded that with the existing measurement error, OD matrix at medium-size zones with a radius of 5 kilometers - can be estimated with a high level of confidence. Hence, the technology is not yet mature enough to support the estimation of OD matrices at a high resolution. However, with the
current pace of technological developments it is presumed that detailed estimation of transit demand will become more accessible in the foreseen future.

3.5 SIMULATION MODEL ARCHITECTURE
The transit simulation model is completely integrated within the mesoscopic traffic simulation, Mezzo. The model is developed within an object-oriented framework. Transit modeling implies the introduction of several new entities (also known as classes) to the simulation model such as transit line, route, trip, stop, vehicle and vehicle type. Furthermore, the modeling of passengers' decision making involves the introduction of additional entities such as passenger, OD in terms of stops and path alternative. The attributes associated with the various entities as well as the interrelations between them can be seen in the simplified object-oriented framework available at Appendix A.

Figure 3.6 shows a flowchart of the transit simulation process. As an event-based simulation model, the time clock of the simulation progresses from one event to the other according to a chronological list of events that refers to the relevant objects. At the start of the simulation, all objects are initialized and some of them register an event. The execution of most events triggers the generation of new subsequent events. Three modules are highlighted in the figure as they are the main focus of the following chapter: The choice-set generation model and passenger decisions.
On initialization of the simulation run, the list of transit lines is read and the corresponding LINE and transit route (ROUTE) objects are created. The simulation books events for the scheduled departure time of the first trip from each line. In addition, the model goes over all OD pairs of stops and registers an event for the generation of the first passenger from each OD pair. As part of the initialization phase, all transit vehicle types objects (TVTYPE) are initialized and the simulation checks whether path-sets were given as an input or a background-set has to be generated. The main simulation loop progresses through the execution of registered events. The transit-related events are: enter link; exit link; enter stop; (vehicle) trip departure; (vehicle) trip end; passenger arrival at stop and; passenger generation. The simulation
clock advances to the earliest registered event. The model identifies the type of event and carries out the respective actions. At the beginning of the simulation, two event types are booked – the departure time of the first trip for each line and the generation of the first passenger for each OD pair. When a passenger generation event is activated the PASSENGER object is initialized with the respective OD pair and path-set. The passenger makes a connection decision to determine at which stop to initiate the trip. In case the passenger decides to walk to another stop then the simulation books an event to passenger arrival time at the chosen stop. If the passenger stays at the current stop then the number of waiting passengers is updated accordingly. Regardless of passenger's decision, the simulation model books an event for the arrival time of the next passenger with the same OD pair.

The arrival of a passenger at a stop is followed by checking whether this is passenger’s destination. If it is not yet the final destination, then the passenger may reconsider whether to stay and wait at this stop. A connection decision is then performed by passengers that arrived at this stop after alighting from a transit vehicle. In addition, reconsideration may take place due to additional information that becomes available (e.g. vehicle arrival time information display). The model can be modified to address various triggers for reconsideration such as elapsed waiting time or overcrowded on-board conditions. Passengers that stay at the current stop are added to the list of waiting passengers.

When a scheduled trip departure event is activated the TRIP object is generated. A transit vehicle (TVEHICLE) is assigned to this trip. If the assigned vehicle is not yet in service (in case that this trip is the first on its trip chain) then a transit vehicle object is generated and assigned the properties of the required vehicle type. It then enters the first link on its route. This is also the case if the TVEHICLE object already exists and is available to depart. In case that the vehicle is not yet available to depart (i.e. has not completed the recovery time from its previous trip), the trip departure is deferred until the vehicle becomes available.

A transit vehicle that enters a link on its route checks whether or not there are stops to be serviced on this link. If there are no stops on the link, the link exit time is calculated and an event to enter the next link is added to the event list. If there is a stop on the link, the travel time to the stop is calculated and an event to enter the stop is
generated with the appropriate arrival time. The driving time to the stop is calculated as a proportion of the link travel time, depending on the location of the stop.

The arrival of a transit vehicle at stop triggers a series of passenger decisions. Passengers waiting at the stop decide whether to board the vehicle or keep waiting for another vehicle. Passengers on-board decide whether to alight at the stop or stay on-board. If the decision implies that the passenger has a new stop on his/her travel path (e.g. alighting at the stop), then the simulation registers an event for passenger arrival time at that stop. The simulation also checks whether the current location is passenger’s final destination.

After all passenger decisions are made, the dwell time is calculated. Based on the dwell time and control strategies that may be implemented, the timing for a new event to exit the stop is determined. When the vehicle exits the stop, similarly to the event of entering a link, the simulation checks if there are any more stops on the link and calculates the driving time to the next stop or to the end of the link based on the current traffic conditions and the respective distance. An event to enter the next stop or to exit the link is generated. It should be noted that the simulation model is able to process multiple trips and lines simultaneously.

Finally, when the transit vehicle arrives at the end terminal, BusMezzo checks whether or not there is an additional trip for this vehicle. If so, and the next trip has already been activated (i.e. the trip scheduled departure time has already passed), the vehicle is assigned to the next trip and enters its first link. In the case that the next trip is not activated, the vehicle waits until the scheduled departure time. The transit vehicle is removed if there are no more trips assigned to it.

The main simulation loop is designed to support the implementation of control strategies, which requires additional steps. Each object that is a potential subject for control strategy is indicated by a flag. Every time that an event is executed, the model checks whether a control strategy is defined for this type of event, and if so, executing the control logic to determine the appropriate action.

The simulation model requires a set of input files and generates a set of output files that can be used for further analysis at various levels of aggregation. A complete list of the inputs and outputs of the simulation model is given at Appendix B. Figure 3.7 lists the main input and output data, as well as the processes involved with the simulation model. Some of these processes have been discussed above, while others – namely
passenger decisions, real time information and control strategies - are presented in detail in the following chapters. The simulation model can be regarded as a laboratory tool that can be further enriched to enable a broader range of applications and studies.

Figure 3.7: Transit simulation model inputs and outputs
4. **Dynamic Path Choice Model**

The previous chapter presented the framework of the transit simulation model. It provided an overview of the mesoscopic traffic dynamics modeling and described the main modeling components involved with the representation of transit operations. The core of the transit operations modeling capabilities was developed and discussed in my master thesis (Cats, 2008) and related papers. The supply side of the transit system will be further discussed in the context of transit performance analysis and the evaluation of control strategies in Chapter 6. The following chapters discuss the modeling of the demand side – the dynamic loading of travelers in the transit system.

4.1 **Two-Stage Modeling**

The modeling approach adopted in this thesis is to represent the transit path choice as a semi-compensatory two-stage choice process. Figure 4.1 illustrates the two-stage approach that is applied in this study. The first phase is a non-compensatory rule-based choice-set generation model (CSGM). The deterministic generation process is based on network configuration (lines, stops) and the corresponding timetables which specify trip departure times and expected travel times. It results in a path-set for a given OD pair of locations in the network. The path-set is given as an input to the probabilistic dynamic path-choice model (DPCM). The simulation model, BusMezzo, generates individual travelers which undertake successive path choice decisions that are triggered by the evolving transit system conditions (e.g. vehicle arrival). The evaluation of alternative actions (e.g. board vs. stay) depends on traveler’s preferences and traveler’s expectations. The latter are determined by prior-knowledge and the availability of real-time information (RTI). Traveler’s ability to carry out his/her decision is subject to vehicle capacity constraints.

---

The CSGM could be applied dynamically by generating the time-dependent path-set upon making a path decision. This allows a consistently adaptive approach of the path decision process. Alternatively, Bovy (2009) highlighted the theoretical and practical advantages of specifying the choice-set as a preliminary phase. Among those advantages he listed the higher adequacy when dealing with overlapping as well as the large savings in computational effort as the exhaustive choice-set generation phase is performed once. The generation of an intermediate choice-set avoids the enumeration of all path alternatives for each traveler choice. The generated choice-set is aimed to reproduce the set of alternatives that are considered by travelers when carrying out their trip.

The CSGM is currently implemented in BusMezzo as an initial phase. It results with a path-set for each pair of locations in the network that is given as input to the DPCM. However, there is no limitation to modify the simulation model so that the CSGM would be applied dynamically.

The remainder of this chapter consists of the following: Section 4.2 elaborates on the OD composition and demand generation; Section 4.3 presents the DPCM process followed by the definition of a path alternative in the context of this thesis; Section 4.5 presents in detail the deterministic non-compensatory CSGM along with a corresponding estimation method. Section 4.6 discusses the successive travel decisions and their respective choice structures; The choice model structure and the assumptions
used in the current implementation with respect to travelers’ anticipations are presented in Section 4.7. Chapter 5 presents a survey that was conducted in order to estimate the proposed models.

4.2 DEMAND GENERATION

The current implementation of BusMezzo considers transit demand as a Point-to-Point (P2P) demand that is originated and destined at spatial points in the network. This corresponds to an intermediate level between the third and the fourth level of passenger demand representation discussed in the previous chapter (Section 3.4.1). The population of individuals is generated based on a time-dependent OD demand matrix. The values of the OD matrix correspond to travelers’ generation rates. Each generated traveler is assigned with the respective OD pair. Hence, the number of arriving travelers at a certain origin is defined by the sum of stochastic arrival processes to various destinations. Since the model is developed with the purpose of high-frequency urban transit services, it is assumed that travelers’ departure time is distributed randomly without the consideration of trip departure choice. Each arrival process is assumed to follow the Poisson distribution independently. The P2P demand may refer to one of the followings:

- Transit stops – well-defined geographical locations where boarding and alighting from transit services takes place (bus stops and rail platforms).
- Activity anchors – a geographical site (e.g. urban square, campus, shopping mall) that can be represented as a spatial point.
- Centers of gravity of spatial zones – a spatial point which represents the centre of a spatial area. Conventional transport planning tools use traffic area zones (TAZ) as the basic unit of demand for travel with the centre of each TAZ defined as centroid.

The OD matrix can be composed of elements of different nature. Distances between geographical locations can be specified or obtained from a GIS tool. Travelers can initiate their transit trip at various stops subject to their path choice decisions by travelling between connected spatial points (e.g. stop-stop, anchor/centroid-stop). Note that the CSGM takes into account the connections between the origin and the first stop and between the last stop and the destination. This modeling approach allows having an OD matrix that refers to key locations such as transit hubs or urban landmarks without
having to specify generation rates at the stop level. It has the advantage of modeling travelers’ distribution over the relevant stops through the execution of the DPCM in addition to its advantage in terms of data requirements. The model allows the specification of a connection between each pair of spatial points in the network, regardless of their characteristics. This property avoids the limitation induced by the hierarchal structure used in existing transit assignment tools (e.g. nearby stops that belong to a neighboring TAZ are inaccessible due to arbitrary cut offs).

The current implementation of the P2P OD matrix in terms of centroids assigns all travelers with the same distances to alternative transit stops. However, in reality these distances vary among individuals based on the exact trip origin and destination. The approximation of these distances is especially important in modeling the first stop decision which obtains the market share of each stop. The variation in distances between stops and among individuals that have the same origin centroid could be potentially captured by sampling techniques.

4.3 PATH DECISION MAKING PROCESS
The DPCM considers traveler trips as a sequence of adaptive travel decisions that are associated with traveler progress in the network. This stands in contrast to static assignment models which consider path choice as a single decision of the complete path to be followed. Furthermore, at no point the traveler has to choose between paths. Instead, each travel decision considers alternative actions that are relevant for a given trip stage.

Figure 4.2 presents the dynamic path choice process in BusMezzo. Passengers are first generated at their origin based on a random arrival process with the generation rates specified at the OD matrix. Passenger perception is shaped by the individual’s prior-knowledge and the information that is available to each individual at a given trip stage. The first decision, a connection decision, is concerned with choosing at which stop to initiate the trip. Passengers may also choose to stay at the current location (in case their origin is a transit stop). While passengers wait at a stop they may receive new information that will deviate from their expectations and will lead to the reconsideration of their current location by performing a new connection decision. Such information could arise from passenger experience (e.g. elapsed waiting time) or RTI display.
Each time that a transit vehicle arrives at the stop, a *boarding decision* is invoked as long as the passenger remains at the stop. The passenger has to choose whether to
board this vehicle or stay at the stop and wait for another vehicle. Passenger decision is then also subject to capacity constraints. Upon boarding, the passenger updates his/her perceptions on downstream alternatives based on the information available at this stage. Then an initial alighting decision is performed by assessing all potential alighting stops and the onward paths associated with them. While the passenger is on-board, new information which deviates from passenger’s expectations may cause the reconsideration of his/her alighting decision. As in the case of waiting at stop, this information could be the experienced travel times or crowding conditions or RTI that became available. Each time the transit vehicle approaches a transit stop, the simulation checks which of the passengers on-board are designated to alight at this stop.

Following passenger alighting, a new connection decision is performed. This decision considers whether to stay at the current stop and wait for another line, walk to a nearby stop or walk to the destination if feasible. In case the passenger had reached his/her final destination then the trip is concluded. Otherwise this intermediate stop is regarded as the new origin stop and the corresponding background path-set is retrieved. The DPCM continues with a chain of boarding, alighting and connection decisions. Before presenting the choice structure and the respective choice-set of actions for each travel decision in section 4.6, it is necessary to provide the definition of a path alternative and the details of the CSGM.

4.4 PATH ALTERNATIVE DEFINITION
Each transit trip has an origin and a destination. However, there may be numerous alternative transit paths that connect between a single OD pair. While there is no obvious way to define a transit path, the definition should be unambiguous. The following presents the definition of a transit path alternative that is used throughout this thesis.

Each transit path consists of three types of elements: transit stops, transit lines and connection links. The term connection links includes access, egress and transfer links that can be carried on by various non-transit travel modes (e.g. walking, cycling, park and ride). Each combination of these three components defines a unique path. A path alternative $a_{i}^{od}$ which connects an origin ($o$) to a destination ($d$) is defined by an ordered set of transit stops ($S_{a_{i}^{od}} \subseteq S$), transit lines ($L_{a_{i}^{od}} \subseteq L$) and connection links
\{(C_a^{od} \subseteq C)\}. S, L and C are the sets of all the transit stops, transit lines and connection links in the network, respectively. \(a_i^{od}\) is an element in the set of paths \(A^{od}\).

The complete path can be composed by chaining the connection links, transit stops and transit lines. First, the passenger has to get from the actual origin to an origin transit stop. For example consider a traveler that starts a trip at origin \(o\) and has to get to destination \(d\), as illustrated in Figure 4.3. The traveler has a set of alternative origin stops that are accessible from their origin by non-transit modes \((s_2, s_3)\). Connection links connect each origin location with the corresponding accessible transit stops \((c_1, c_2)\). In the case of walking, this set of stops includes all the stops which are within a walking distance. Transit stops may be served by several transit lines, some of them may provide a direct connection to passenger’s final destination \((l_1)\) while others may require to transfer to another line \((l_2, l_3)\). The transfer may take place at the stop where the passenger went off the ingoing transit line \((s_4, s_5)\) or from another stop requiring a connection link \((s_7)\). Similarly, the destination can be reached directly by a transit stop or require a connection link \((c_4)\) between the last transit stop and the final destination.

![Figure 4.3: Illustration of transit path alternative components](image)

There are various ways to enumerate and define the path alternatives in Figure 4.3. The approach adopted in this thesis is to define as a single path all the alternatives that imply the same chain of stops with equivalent link attributes. Hence, assuming that
travelers are indifferent between parallel lines the network in Figure 4.3 includes three path alternatives:

1. Go to stop $s_2$ using connection $c_1$; board line $l_1$; alight at $d$.
2. Go to stop $s_3$ using connection $c_2$; board either line $l_2$ or $l_3$; alight at either stop $s_4$ or $s_5$; board line $l_4$; alight at $s_6$; Go to $d$ using connection $c_4$.
3. Go to stop $s_3$ using connection $c_2$; board either line $l_2$ or $l_3$; alight at stop $s_4$; Go to stop $s_7$ using connection $c_3$; board line $l_5$; alight at $d$.

The proposed path alternative definition is constructed by listing connection links and transit lines alternately between consecutive transit stops, with the first and last links reserved for connection links. The graph that represents a transit path consists of nodes and links. Nodes correspond to transit stops and origin and destination locations. Links are either segments of transit lines or connection links. In order to maintain a uniform and standard path definition, an artificial looping connection link is introduced for each node (stop). These artificial connection links correspond to the alternative of staying at the stop, have a null travel time and hence have no influence on the actual decision process. This is necessary in order to have a model that is flexible enough to accommodate various forms of passenger demand data.

The following general definition can be applied to construct any path alternative given the ordered sets of $S_{a_i}^{qd}$, $L_{a_i}^{qd}$ and $C_{a_i}^{qd}$. For clarity reasons, the path alternative and the OD pair notations are sometimes omitted in the following discussion, keeping in mind that we consider a specific path alternative that belongs to a path-set of a specific OD pair. The index $i$ always refers to path alternative $a_i$. The path definition is:

$$ a = \{o, c_1, S_2, S_3, \ldots, S_{2j-1}, c_j, S_{2j}, L_{i}, S_{2j+1}, \ldots, S_{2m}, L_{m}, S_{2m+1}, c_{m+1}, d\} \quad (4.1) $$

This is a conceptual definition as it includes different types of elements and it is aimed only to present the sequence of events.

Each element in the path alternative is a set. This definition enables to group several transit lines that connect between a pair of transit stops or several transit stops which are connected by the same transit lines. In the case that the OD pair corresponds to transit stops they can be replaced by $S_1$ and $S_{2m+2}$.

Since the path definition is composed of chain of stops with connection and transit links alternately in between, the stops index is in the order of two compared with the connection and line indexes. Therefore, the set of transfer stops $S_{2j-1}$ is
followed by connection set $C_j$ that brings the traveler to $S_{2j}$ where they may board transit lines that belong to set $L_j$. In what follows, index $j$ refers to the respective element in path definition 4.1. Note that odd indexes refer to connection stops while even indexes refer to stops where boarding may take place.

The number of transit lines included in path alternative $(m)$ determines the number of transit stops $(2m + 2$, including OD) and connection links $(m + 1)$ included in the path definition. The number of transfers is $m - 1$.

Each of the elements in $a_i$ has to fulfill certain topological conditions that are necessary in order to construct a meaningful path which can be formalized as follows:

$$s_1, s_2 \in S^l; s_1 <_X s_2 \quad \forall (s_1 \in S_{2j}, s_2 \in S_{2j+1}); \forall l \in L_j \quad (4.2)$$

$$s_1^c \in S_{2j-1}, s_2^c \in S_{2j} \quad \forall c \in C_j \quad (4.3)$$

$$s = s_2^{C_{j/2}}, s \in S^{l_{j/2}} \quad \forall s \in S_j, j \text{ is even and } j \neq 2m + 2; \forall c_{j/2} \in C_{j/2}; \forall l_{j/2} \in L_{j/2} \quad (4.4)$$

$$s = s_2^{C_{(j+1)/2}}, s \in S^{l_{(j+1)/2}} \quad \forall s \in S_j, j \text{ is odd and } j \neq 1; \forall c_{(j+1)/2} \in C_{(j+1)/2}; \forall l_{(j+1)/2} \in L_{(j+1)/2} \quad (4.5)$$

Where:

- $S^l$ - the ordered set of transit stops along the route of transit line $l$
- $<_X$ - an operator that checks whether the term on the left comes before the term on the right on the ordered set of $X$
- $s_1^c$ - the origin stop of connection link $c$
- $s_2^c$ - the destination stop of connection link $c$

These conditions implies that every transit line and connection link have to connect two consecutive stops on the path definition. In the case of transit lines, the order of the stops along the transit route has to be taken into account (as each route direction is regarded as a separate transit line). In the common case of several transit lines using the same corridor (known as ‘the common lines problem’ in the literature; Chirqui and Robillard, 1975), expression 4.5 implies that all of them will be elements of $S_j$. Similarly, all the lines connecting two consecutive sets of transfer stops as elements of $L_j$. This property of the proposed path alternative definition has an important role in minimizing the potential overlap between considered paths. However, when grouping lines it is important to include only elements that passengers are indifferent between them. Hence, the above conditions may be necessary but not sufficient.
4.5 Choice-set Generation Model
The dynamic path choice model requires a background path-set for each potential OD pair. This section presents the method for generating such sets.

4.5.1 Choice-set Generation Process

4.5.1.1 Generation Process
A CSGM in the context of route choice and transit path choice in particular is non-trivial. It is aimed to generate a set of reasonable path alternatives for each OD pair. Note that it is necessary to compose a choice-set also for OD pairs that have null travel demand as they make become relevant during the dynamic loading process.

A general scheme for the CSGM includes the implementation of a path search method (e.g. dynamic programming, branch and bound, recursive search) with the exclusion of unreasonable path alternatives based on a set of filtering rules. This process has to be on one hand limited enough to dismiss irrelevant paths and on the other hand to be flexible enough to include the major paths that are considered by travelers. This choice-set may be referred to as the collective consideration choice-set (Bovy, 2009).

Figure 4.4 presents the scheme of the proposed CSGM. The process contains four modules. The first three could be either static or dynamic. In the case of an initial CSGM phase, it results in a master-set that is used throughout the probabilistic DPCM. The simulation model can then retrieves the static master-set and applies dynamic filtering rules before carrying out the choice process.
The CSGM is a generic process which may accommodate various generation methods, constraints and filtering rules. Note that the purpose of CSGM is to generate all the paths that may be attractive under some circumstances – for example an infrequent line may happen to arrive at the stop and become relevant. The following presents one possible formulation of the CSGM by discussing the four modeling modules that are presented in Figure 4.4.
4.5.1.2 Path Generator
The path generator module generates all direct paths which do not involve transferring between legs and indirect paths that satisfy a set of logical constraints. A recursive search method was adopted in this study as the path generation method. First, it constructs a data structure for all the direct paths connecting each OD pair in the network. Second, indirect paths are constructed by examining paths between all stops within an increasing distance in terms of number of transfers from the origin stop and all stops that are within one connection link distance from the destination stop. Thus it introduces in each stage an additional stop between the origin and the destination by setting the most recently added stop as the new intermediate origin. This method is known as forward search (Tan et al., 2007). The process looks first for all direct paths between the origin stop and each of the stops connected directly to the destination stop – layer \( (d - 1) \) in Figure 4.5.

Figure 4.5: Recursive search illustration
(dashed lines – connection links; solid lines – transit links)
At stage \( y \), the method searches for direct connections between the \( y-th \) layer and the \( (d-1) \) layer. Each stage extends the depth of the tree by another connection or transit link, alternately. The recursive search continues as long as the depth of the tree does not exceed a pre-defined criterion of the maximum number of transfers (\( \tau_{\text{trans}}^{\text{max}} \)). Moreover, the number of loops in this recursive process is constrained by a relative definition of the maximum allowable number of extra transfers (\( \tau_{\text{ext}}^{\text{trans}} \)). This constraint implies that the number of transfers cannot exceed by more than \( \tau_{\text{ext}}^{\text{trans}} \) the minimum number of transfers that was obtained among previously generated paths for this specific OD pair.

The generated paths are also subject to two additional **logical constraints**:

- **No loops** – paths can not include the same stop twice, unless there is a single walking link in between (recall the path definition, Section 4.4). Let us define an indicator which takes ‘1’ if stop \( s_k \) belongs to a set of stop \( S_j \) on a path alternative definition, or formally: 
\[
\delta_{\text{stop}}_{k,j} = \begin{cases} 
1 & s_k \in S_j \\
0 & \text{otherwise}
\end{cases}
\]

The following constraint ensures that the same stop will not repeat later on the path definition further downstream with the exception of the immediate successor:
\[
\sum_{j=1}^{2m+1} \sum_{h=j+2}^{2m+2} \delta_{\text{stop}}_{k,j} \cdot \delta_{\text{stop}}_{k,h} \neq 1 \quad (4.6)
\]

In the case of a circular line, it may be possible to induce a loop by boarding and alighting at the same stop. Hence, the following additional constraint is included in the CSGM:
\[
\sum_{j=1}^{m} \delta_{\text{stop}}_{k,2j} \cdot \delta_{\text{stop}}_{k,2j+1} \neq 1 \quad (4.7)
\]

By refereeing only to pairs of \((S_{2j}, S_{2j+1})\), staying at the same stop by using a transit line is prohibited while enabling to stay at the same stop due to artificial connection link.

- **No abrupt transit legs** – paths can not imply that passengers get off a transit line only to wait to board the same line again. The indicator \( \delta_{\text{line}}_{k,j} \) takes ‘1’ if line \( l_k \) belongs to the set of lines \( L_j \) on a path alternative definition, as follows:
\[
\delta_{\text{line}}_{k,j} = \begin{cases} 
1 & l_k \in L_j \\
0 & \text{otherwise}
\end{cases}
\]

The following constrain ensures that the same line does not repeat on the immediate successor:
\[
\sum_{j=1}^{m-1} \delta_{\text{line}_{k,j}} \cdot \delta_{\text{line}_{k,j+1}} \neq 1
\] (4.8)

All the paths that were generated form an initial-set. The purpose of the next module is to exclude unreasonable paths that are not actually considered by travelers.

4.5.1.3 Filtering
The recursive search method may generate a large number of alternative paths. Previous studies suggested that the CSGM has to include non-compensatory rules aimed to reduce the choice-set size before applying the compensatory choice phase (Recker and Golob, 1979; Cantillo and Ortuzar 2005). Therefore, an alternative that under-performs compared with a certain threshold is excluded from further consideration, regardless of its other attributes. Following previous studies (Swait and Ben-Akiva, 1987; Gilbride and Allenby, 2004; Gantillo and Ortuzar, 2005; Kaplan et al., 2011) the filtering rules exercise a conjunctive relation.

The filtering rules enforce behavioral constraints at the individual-path level as well as at the choice-set level. The following rules assess each of the paths in the initial-set independently of other path alternatives:

- **Maximum walking distance** – excludes paths with a total walking distance that is greater than a given threshold, or formally:

\[
\sum_{j=1}^{m+1} d_{c_j} \leq \tau_{\text{walk}}^{\text{max}}
\] (4.9)

Where \(d_{c_j}\) is the distance of connection link \(c_j \in C_j\) and \(\tau_{\text{walk}}^{\text{max}}\) is the upper bound walking distance threshold.

- **No opposing lines** – most transit lines have a corresponding opposing route direction. Adding a constraint that path alternatives cannot include two opposing line directions will prevent the generation of paths that imply alighting further downstream just to take back the opposing direction. Let \(\delta_{\text{opp}_{v,w}}\) be an indicator that takes '1' if \(l_v\) and \(l_w\) are opposing transit links and '0' otherwise (as pre-defined in the transit network input). Then the following constraint is introduces:

\[
\sum_{j=1}^{m-1} \sum_{h=j+1}^{m} \delta_{\text{line}_{v,j}} \cdot \delta_{\text{line}_{w,h}} \cdot \delta_{\text{opp}_{v,w}} \neq 1
\] (4.10)

This constraint ensures that two opposing line directions would not be included on two different legs of the path alternative. Hence, the definition of opposing direction implies that they are mutually exclusive.
The alternatives that have retained in the choice-set are further screened to check if they fulfill requirements at the choice-set level. The following rules were introduced:

- **Maximum number of extra transfers** – even though this rule is applied throughout the generation process, it has to be reassessed as it may be violated by paths that were generated later on by the recursive search process:

  \[ \delta_{\text{trans}} = \begin{cases} 1 & \min_{a_i \in A} |L_i| + \tau_{\text{ext}}^\text{trans} < |L| \\ 0 & \text{otherwise} \end{cases} \quad (4.11) \]

- **Maximum IVT** – checks that the total in-vehicle time (IVT) of each path alternative does not exceed a pre-defined ratio threshold from the alternative with the minimum total IVT. This constraint assures that unreasonably long paths will not be included in the choice-set. This constraint is formulated as follows:

  \[ \delta_{\text{IVT ratio}} = \begin{cases} 1 & \min_{a_i \in A} \left\{ \sum_{l_j \in L_i} EIVT_{l_j} \right\} \cdot \tau_{\text{max ratio}}^{IVT} < \sum_{l_j \in L_i} EIVT_{l_j} \\ 0 & \text{otherwise} \end{cases} \quad (4.12) \]

  Where \( \tau_{\text{max ratio}}^{IVT} \) is the maximum total in-vehicle time ratio and \( EIVT_{l_j} \) is the expected in-vehicle time involved with transit line \( l_j \) calculated as:

  \[ EIVT_{l_j}(t) = ST_{s_2j+1}^{l_j}(t) - ST_{s_2j}^{l_j}(t) \quad (4.13) \]

  \( ST_{s_2j}^{l_j}(t) \) is the scheduled time of transit line \( l_j \) at stop \( s_j \) on time period \( t \). Please note that path alternative definition implies that it does not matter which \( l_j \in L_j \) is used for the calculation as only transit lines with the same riding time between consecutive stops can be grouped into a single alternative.

- **Dominancy rules** - An alternative that is not better than another alternative in the choice-set in any aspect and is worse than this alternative in at least on aspect, is regarded as dominated (or ‘Pareto non-optimal’ in economics terminology). Only paths that are not dominated by other path alternatives are retained in the choice-set. It is assumed that the inferiority of dominated alternatives excludes them from the choice-set considered by the decision maker (e.g. Androuetsopoulou and Zografos, 2009).
The aspects considered for determining dominancy are the number of transfers, total IVT and total walking time. A dominancy rule that considers these three aspects can be formulated as:

$$\delta_i = \begin{cases} 1 & \forall a \in A: \left( \left( |I_i| \leq |I_a| AND \sum_{t \in T_i} EIVT_i \cdot \gamma_{IVT} \leq \sum_{t \in T_a} EIVT_a AND \sum_{c \in C_i} d_{c,i} \cdot \gamma_c \leq \sum_{c \in C_a} d_{c,a} \right) \text{ OR } \left( |I_i| \leq |I_a| AND \sum_{t \in T_i} EIVT_i \cdot \gamma_{IVT} < \sum_{t \in T_a} EIVT_a AND \sum_{c \in C_i} d_{c,i} \cdot \gamma_c < \sum_{c \in C_a} d_{c,a} \right) \right) \text{ OR } \\ 0 & \text{otherwise} \end{cases}$$

Where $\delta_i$ indicates if alternative $i$ is dominated in the relevant choice-set, $\gamma_{IVT}$ and $\gamma_c$ are the dominancy perception threshold for the IVT and connection distance, respectively. The perception thresholds are incorporated to determine if a path alternative exceeds the maximum allowable digression relative to other path alternatives. These thresholds account for cognitive process limitations and a behavioral assumption that the difference has to be noticeable enough to cause the exclusion of a dominated alternative. Various dominancy rules can be formulated in a similar fashion.

In order to improve the efficiency of the path generation method the dominancy rules were enforced not only at the filtering stage but also throughout the generation process. It is presumed that the immediate exclusion of unattractive alternatives saves unnecessary comparisons and computational efforts.

The path alternatives that satisfied all the filtering rules compose the unconsolidated-set. The set obtained at this stage may include several path alternatives that are virtually identical besides containing a different line or transfer stop on the same corridor. The next phase screens the choice-set of each OD pair to examine whether some path alternatives can be merged.

4.5.1.4 Merging
The way in which paths are defined (Section 4.4), enables to accommodate numerous paths that could be merged into sets of common stops and lines. The merging of path alternatives constructs hyperpath-like alternatives. This helps to avoid some of the potential overlapping that may introduce bias in the random utility choice model. Alternatives that have the same sequence of lines and vary in their intermediate stops can be merged to form a single alternative as long as the routes between these intermediate stops are identical for both the incoming and outgoing transit legs.
Similarly, alternatives that have the same sequence of stops and vary in their transit lines in between can be merged if both lines follow the same route between the relevant pair of stops. The obtained choice-set is referred to as master-set.

The master-set is an outcome of a rule-based static and deterministic CSGM. The process so far results in a single time-independent choice-set for each OD pair in the network. This process is performed at the initialization of the simulation model. It is a computationally expensive process but it does not have to be performed more than once for a given network. The master-set could be given as an input to BusMezzo for various scenarios of the same network.
The following pseudo-code summarizes the list of actions involved with the generation of a master-set:

<table>
<thead>
<tr>
<th>Step</th>
</tr>
</thead>
</table>
| **Step 1** | *Direct paths* - For every pair of stops \((\forall s_o, s_d \in S)\), search for a transit line or a connection link that connects these stops. Add path alternative \(a_i\) to the respective initial set \(A\).
|     | Denote by \(S_1\) and \(S_{d-1}\) the set of all stops that have a direct path to \(s_o\) and \(s_d\), respectively. Set \(y = 1\).
| **Step 2** | *Indirect paths* – Take the next pair of stops and search for an indirect path by repeating step 3. When the search has been exhausted go to step 4.
| **Step 3** | *Recursive search* –
|     | (A) \(S_{y+1}\) is the set of stops that have a direct connection with \(s \in S_y\). Search for a direct path \(a\) (using a connection or a transit link) between stop \(s \in S_y\) and \(s \in S_d\) subject to logical and dominancy constraints. Add path alternative \(a^{od}\) to the initial set \(A\).
|     | (B) If \(\frac{y}{2} - 2 \leq \tau_{max}^{trans} \text{ AND } \frac{y}{2} - 2 \leq arg\left(\min_{a_i \in A}|L_i|\right) + \tau^{trans}_{ext}\) then \(y = y + 1\) and go back to (A). Otherwise go back to step 2.
| **Step 4** | *Individual-path assessment* – Remove all the alternatives that do not fulfill each of the individual-level behavioral rules.
| **Step 5** | *Choice-set assessment* – Remove all the alternatives that do not fulfill the behavioral rules that are defined relative to other alternatives (including dominancy rules).
| **Step 6** | *Merging* – Compress alternatives by merging based on common lines and then stops.

### 4.5.1.5 Dynamic Filtering

During the dynamic loading, the master-set is further examined each time a traveler takes a decision. The dynamic filtering rules aim to exclude paths that are unreasonable under the time-dependent conditions. The following couple of dynamic filtering rules are applied upon each path choice decision:
Availability (or maximum allowable expected waiting time) – It is assumed that a considered alternative path cannot contain a service that is not available in a reasonable time horizon. This condition can be formulated as:

\[ \delta_{\text{wait}_i} = \begin{cases} 
0 & \text{min}_{l_j \in L_i} \{ EAT_{s_{2j}}^{l_j} : EAT_{s_{2j}}^{l_j} > t \} - t \leq \tau_{\text{wait}}^\text{max} \\
1 & \text{otherwise} 
\end{cases} \forall l_j \in L_i (4.15) \]

Where \( EAT_{s_{2j}}^{l_j} \) is the expected arrival time of line \( l_j \) at stop \( s_{2j} \) (based on the information available to the traveler), \( t \) is the current time and \( \tau_{\text{wait}}^\text{max} \) is the maximum acceptable waiting time.

‘Only if it may be worthwhile to wait’ – exclude transit lines that their IVT is longer than the IVT plus the worst-case perceived waiting time (the longest scheduled headway) for alternative lines. Hence, this filtering rule may result in the exclusion of some transit lines from the set of lines on a given path leg if the following rule is violated for some line \( l_j \in L_i \):

\[ EIVT_{ij} < EIVT_{i_d} + \max_{k \in K_t} \{ ET_{s_{k+1}}^{l_{d}} - ET_{s_{k}}^{l_{d}} \} \forall l_j, l_d \in L_j, \forall L_j \in L_i (4.16) \]

\( ET_{s_{k+1}}^{l_{d}} \) is the scheduled departure time of transit line \( l_d \) from stop \( s \) on the \( k \)-th trip. \( K_t \) is the set of trips that are scheduled to take place during time period \( t \). This resembles the reasoning that is sometimes incorporated into the calculation of passenger waiting times in static assignment models (Jansson and Ridderstolpe, 1992).

The path alternatives that remain in the choice-set compose the collective consideration-set. This is the temporal set of attractive alternatives for some individuals travelling between a given OD. The proposed CSGM could be further enriched by considering a specific individual or groups of travelers as they may vary in their choice-set composition processes and preferences. The heterogeneity among passengers can be incorporated into the CSGM, the choice model or both in the form of latent choice-sets or random coefficients (Boccara, 1989; Gopinath, 1995; Walker, 2001).

### 4.5.2 Choice-set estimation

The CSGM described above can be specified and implemented. A methodology for estimating the CSGM is needed. Since choice-set composition is not an observable behavior, it is not common to have data that will enable its estimation. A stated-
preferences survey can be used to collect data on how individuals compose their choice-set by asking explicitly which alternatives were considered as part of the choice process. Such data was collected through a survey that was conducted as part of this study and is described in details in chapter 5. The estimation of the CSGM parameters can then be based on searching for parameter values that best reproduce the reported choice-sets. The following formalizes the problem and discusses the approach taken towards it. The generalized problem is then simplified in order to deduct a feasible estimation method for the CSGM parameters. Note that the estimation method discussed below is not restricted to transit route choice. In fact, it is not limited to transport-related decisions and can be applied within any discrete choice model context (e.g. location choice, purchasing decisions).

4.5.2.1 The General Model
Choice-set estimation can refer to the choice-set as the entity to be estimated. This implies assigning each possible choice-set a probability to be chosen, in line with the probabilistic perspective of choice sets that is formulated as follows (Manski, 1977):

\[ P_n(i|V_i, Q_n) = \sum_{C \in M_n} P_n(i|C, V_i, Q_n) \cdot P_n(C|M_n, V_i, Q_n) \quad \forall i \in C, C \in M_n \]  

(4.17)

Where \( V_i \) is a set of alternative \( i \) characteristics and \( Q_n \) is the set of individual \( n \) attributes (e.g. socio-demographic and travel habits characteristics). \( C \) is a consideration choice-set and \( M_n \) is the set of all non-empty subsets of the set. \( U_n \) is the set of all feasible alternatives at the specific decision context from a universal set \( U (U_n \subseteq U) \).

This formulation is usually presented in terms of summing the probability shares of a certain alternative over all possible choice-sets. An alternative approach may consider the second term as the probability that a certain choice-set will become the consideration set that is used by the decision maker. Hence, the probability of a potential choice-set to be the consideration set is the probability that all the alternatives that define this set would be considered and all the others are not part of the set:

\[ P_n(C|M_n, V_i, Q_n) = \prod_{i \in C} P_n(i|V_i, Q_n) \prod_{i \notin C} [1 - P_n(i \in C|V_i, Q_n)] \quad C \subseteq M_n \]  

(4.18)

This approach is in line with our semi-compensatory two-stage choice model. Note that this formulation does not require composing and enumerating all possible choice-sets. Each alternative has a probability to be included in the considered choice set which can be referred to as ‘inclusion probability’ or ‘membership probability’. 

85
Reported choice-sets from a stated-preference survey are the data required for estimating a CSGM. Each alternative is either included or excluded from the choice-set of a given respondent. Let $y_{in}^{od}$ be an indicator that equals one if alternative $i$ is included in the stated choice-set of individual $n$ that traveled between a given OD pair and equals zero otherwise. The inclusion rate of alternative $i$ is therefore denoted by $\bar{y}_i^{od} = \frac{\sum_n y_{in}^{od}}{N^{od}}$, where $N^{od}$ is the number of respondents for a given OD pair. The inclusion rate can be interpreted as the probability of an alternative to be selected to the choice-set.

The estimation of the non-compensatory CSGM involves two aspects and two corresponding sets of parameters for each individual $n$:

- **Activation** - which filtering rules to include in the CSGM. An activation parameter $\beta_{rn}$ indicates the probability that individual $n$ applies filtering rule $r$ as part of the choice-set composition process. However, an individual either enforces or not a given filtering rule as indicated by a binary variable $\theta_{rn}$ which is the result of a Bernoulli trial with a probability of $\beta_{rn}$.

- **Thresholds** – the respective threshold parameter values of various filtering rules. $\alpha_{rn}$ is the threshold level that individual $n$ applies when enforcing filtering rule $r$. In general, a filtering rule may have more than one threshold parameter or no thresholds (e.g. strict dominance rule). Nevertheless, we use the same index notation in order to indicate to which filtering rule a threshold parameter refers, without loss of generality.

Since individuals have heterogeneous preferences, they may vary in the set of filtering rules that they apply. For example, some individuals may apply a filtering rule with a strict limit on the number of transfers, while others may be willing to transfers as much as necessary to obtain a minimum door-to-door time. Moreover, the threshold levels can also vary considerably within the population. For example, consider a filtering rule that limits the maximum walking distance involved with a path alternative. The threshold level is expected to vary among individuals. Some of the variation could be explained by socio-demographic attributes, such as age or gender.

Filtering rules and threshold levels are to be determined based on their capability to reproduce reported choice-sets. The corresponding inclusion decision obtained from the model is represented by $z_{in}^{od}$. The estimation method aims to
minimize the difference between the generated and the reported vectors, \( Z \) and \( Y \), respectively.

Each filtering rule can be formulated as a mathematical condition that determines whether alternative \( i \) satisfies filtering rule \( r \) or not for individual \( n \), which results with the indicator \( \delta_{irn} \) value. The satisfaction indicators of individual \( n \) can be summarized in a matrix with dimensions of the number of filtering rules (\(|R|\)) and the number of alternatives for the relevant OD pair (\(|A^{od}|\)).

A filtering rule is a strict requirement that is either satisfied or failed. \( \beta_{rn} \) is assumed to be fixed at the individual-level as it is presumed that individuals’ assessment procedure is consistent over alternatives. Whether a filtering rule is satisfied or not depends on both path and individual attributes as well as on a stochastic component for the unexplained variation. Hence, the probability to satisfy a filtering rule depends on the distribution of \( \epsilon_{rn} \), the error component:

\[
p(\delta_{irn} = 1) = p(f(V_i, Q_n) + \epsilon_{rn} \leq \alpha_{rn}) = F_{\epsilon_{rn}}(f(V_i, Q_n) - \alpha_{rn}) \tag{4.19}
\]

Where \( F_{\epsilon_{rn}} \) is the cumulative density function of \( \epsilon_{rn} \). The error component accounts for two sorts of disturbances: perception error which arises from uncertainty associated with the individual and threshold variation which is caused by model uncertainty. Perception errors can result from unfamiliarity, misinformation and cognitive processing disturbances. For example, a filtering rule that is based on the ratio to the shortest path alternative can be based on mistaken travel time estimate, inaccurate information or calculation errors. Although the two disturbances are derived from distinguishable theoretical concepts, the resulting mathematical expression that incorporates either one of them is equivalent. Hence, it is assumed that \( \epsilon_{rn} \) accounts for both perception errors and unexplained threshold heterogeneity.

Activation parameters may be inter-dependent as the probability of imposing a certain filtering rule may be correlated with the probability to enforce related rules. Moreover, activation parameters can be correlated with the respective threshold parameters. For example, if an individual is highly concerned about travel time then we can expect that a related filtering rule will not only have a high activation probability but will also impose a strict threshold. The correlation matrix between all activation and threshold parameters denoted by \( \pi \) has the size of \( 2R \times 2R \), where \( R \) is the number of filtering rules.
The decision of individual \( n \) whether to include alternative \( i \) in the choice-set or not depends on the satisfaction of activated filtering rules. Following Gilbride and Allenby (2004), it is assumed that:

**Assumption 1:** The choice-set generation is composed based on a conjunctive relation between the filtering rules.

Therefore, an alternative has to satisfy each of the active filtering rules in order to be included in the choice-set. Hence, the CSGM consists of an AND logic condition. In order to define the joint probability function of rule satisfaction the following assumption is introduced.

**Assumption 2:** All activation and threshold parameters are independent

These two assumptions combined imply that one can define the inclusion probability as the product of not failing any of the effective filtering rules:

\[
p(i \in C_n) = \prod_{r=1}^{R} \left[ p(\delta_{irn} = 1, \theta_{rn} = 1) + p(\theta_{rn} = 0) \right] = \prod_{r=1}^{R} (\beta_{rn} F_{rn}(f(V_i, Q_n)) + 1 - \beta_{rn})
\]

(4.20)

Our purpose is to find satisfaction vector \( \beta \) and the distribution parameters for threshold error vector \( \epsilon \) that will result in inclusion rate vector \( Z \) that will maximize the likelihood of the sample probability vector \( Y \). Each observation in the sample is a combination of an individual and an alternative. The number of observations is therefore: \( \sum_{od} N_{od} A_{od} \), where \( N_{od} \) is the number of individuals with a specific OD pair and \( A_{od} \) is the number of alternatives that are relevant for this OD pair. Let us compose a vector of all observations \( y_{in}^{od} \) with index \( j \). In order to indicate the respective individual and alternative that construct observation \( j \), we will use the notation \( j_{in} \) (an individual is associated with a single OD pair).

The likelihood function has the following form:

\[
L^* = \prod_{j_{in}=1}^{\sum_{od} N_{od} A_{od}} \left[ \left\{ p(i \in C_n) \right\}^{y_{jin}} \cdot \left[ 1 - p(i \in C_n) \right]^{1-y_{jin}} \right]
\]

(4.21)

It follows that the log-likelihood function is:

\[
\ln L^* = \sum_{j_{in}=1}^{\sum_{od} N_{od} A_{od}} \left[ y_{jin} \cdot \ln(p(i \in C_n)) \right] + \sum_{j_{in}=1}^{\sum_{od} N_{od} A_{od}} \left[ (1 - y_{jin}) \cdot \ln(1 - p(i \in C_n)) \right]
\]

(4.22)
The optimal conditions are the derivatives of the log-likelihood function by the estimated parameters. Hence, the conditions consist of derivatives by the activation parameters $\beta$ and the derivatives by the parameters that determine the cumulative density function of the threshold parameters’ error component $\epsilon_{rn}$.

The estimation problem formulated involves the specification of individual-specific activation and threshold parameters. The number of observations per individual is determined by the number of potential alternatives. Hence, the estimation problem above cannot be solved as there are not enough observations to specify all the individual-specific parameters.

**Assumption 3:** The probability to apply a filtering rule comes from a distribution which characterizes the entire population (e.g. $\beta_{rn} \sim N(\beta_r, \sigma_r^2)$).

This implies that the probability to apply a filtering rule can vary in the population based on a certain distribution function rather than individual-specific activation probability distributions. This can be represented by a uniform activation parameter ($\beta_{rn} = \beta_r$) in Equation 4.20 while the maximum LL problem (Equation 4.22) remains the same. It reduces the number of activation parameters to be estimated to the order of $2R$ instead of $R \cdot N$.

**Assumption 4:** The probability of threshold levels comes from a distribution which characterizes the entire population (e.g. $\alpha_{rn} \sim N(\alpha_r, \sigma_r^2)$).

If both threshold and activation parameters are distributed normally, then the total number of estimated parameters is $4R$. The optimization problem is a nonlinear mixed-integer problem which is NP-hard. Furthermore, the integrals of parameters’ cumulative distribution function introduce computational complexities in the case of normal distributions.

**4.5.2.2 Simplified Models**

The following introduces additional assumptions with each one of them further simplifying the general model presented in the previous section.

The formulating of the estimation problem at the preceding section assumes heterogeneous preferences in applying filtering rules and the respective threshold levels. This formulation can be simplified by assuming that the probability to apply each
filtering rule is uniform across the population with a random draw for each individual determining whether a certain rule is applied.

Assumption 5: The probability to apply a filtering rule is universal and does not vary between individuals

Under this assumption, we only need to estimate the universal activation probability of each filtering rule and the respective threshold parameters. Then the inclusion probability is:

\[ p(i \in C_n) = \prod_{r=1}^{R} (\beta_r F_{\alpha_r} + 1 - \beta_r) \]  \hspace{1cm} (4.23)

The probabilistic assumption regarding the activation of filtering rules can be relaxed by considering their activation as a binary estimation problem with \( \beta_r \in \{0,1\} \) \( \forall r \).

Assumption 6: A filtering rule is either applied by all individuals or by none

The inclusion probability function defined by equation 4.23 consists of multiplying the ‘passing’ probability of each filtering rule (either by satisfying or ineffectiveness). Each component in the multiplying equals 1 if \( \beta_r = 0 \) and equals \( F(\alpha_r) \) if \( \beta_r = 1 \). Hence, the inclusion probability is the probability to satisfy all effective rules, as required. Therefore:

\[ p(i \in C_n) = \prod_{r \in R_e} F_{\alpha_r} \cdot \{ r \in R_e: \forall r \in R, \beta_r = 1 \} \]  \hspace{1cm} (4.24)

The log-likelihood definition remains as in equation 4.22.

As a final withdrawal from the general probabilistic model that was presented in the previous section towards the deterministic case, the threshold levels can also be regarded as deterministic and homogenous.

Assumption 7: The threshold level is constant among individuals for a given filtering rule

This assumption implies that a filtering rule is either satisfied or not, depending on whether the criterion is met or not, with a deterministic binary value. Therefore, the estimation problem is restricted now to find the set of filtering rules that should be applied and their corresponding threshold levels that will maximize the goodness-of-fit between the choice-sets generated by the model and the reported choice-sets. Hence the number of parameters to be estimated is \( 2R \).

The maximum likelihood method is not applicable for binary variables as the satisfaction of filtering rules does not follow a probability function anymore since the
threshold levels are deterministic. Hence, the optimization problem is formulated as a non-linear least squares objective function with non-linear constraints:

\[
\min \sum_{j=1}^{N_A} (y_{j} - z_{j})^2 \quad (4.25)
\]

s.t.

\[
z_j = \left[ \frac{\sum_{r \in R} \beta_r \delta_{irn}}{\sum_{r \in R} \beta_r} \right] \quad \forall i \in A \quad (4.26)
\]

\[
\delta_{irn} = q(\alpha_r, V_i)
\]

\[
\beta_r \in \{0,1\} \quad \forall r
\]

Where \( \alpha_r \) and \( \beta_r \) are to be estimated for every \( r \). The objective function minimizes the number of false identifications – generating paths that are not reported (false negative) or not generating a reported path (false positive). It may be argued that false positive should be assigned with a higher penalty than false negative in the context of path choice modeling (Bovy, 2009).

The rounding definition of \( z_{j} \) can be revised by replacing equation 4.26 with the following pair of constraints:

\[
\sum_{r \in R} \beta_r - \sum_{r \in R} \beta_r \delta_{irn} + z_{j} > 0 \quad \forall j \quad (4.27)
\]

\[
\sum_{r \in R} \beta_r \delta_{irn} - \sum_{r \in R} \beta_r \cdot z_{j} \geq 0 \quad \forall j \quad (4.28)
\]

The justification for the equivalence of equations (4.27) and (4.28) with equation (4.26) is as follows: The value of \( z_{j} \) can only take values in the range of \( [0,1] \). These two constraints enforce the binary conjunctive relation between the filtering rules and the satisfaction indicator. Their combination yields the following range for \( z_{j} \):

\[
\sum_{r \in R} \beta_r \delta_{irn} - \sum_{r \in R} \beta_r \cdot z_{j} \leq \frac{\sum_{r \in R} \beta_r \delta_{irn}}{\sum_{r \in R} \beta_r} \quad \forall j \quad (4.29)
\]

Since \( \delta_{irn} \) are binary variables, \( \sum_{r \in R} \beta_r - \sum_{r \in R} \beta_r \delta_{irn} \geq 0 \) as the first term is the number of effective filtering rules and the second term is the number of both effective and satisfied filtering rules. The latter is a subset of the former. If \( \sum_{r \in R} \beta_r - \sum_{r \in R} \beta_r \delta_{irn} = 0 \) for some alternative \( i \) then all activated rules are satisfied and equation 4.29 implies:

\[
0 < z_{j} \leq 1 \Rightarrow z_{j} = 1
\]

The second case is that \( \sum_{r \in R} \beta_r - \sum_{r \in R} \beta_r \delta_{irn} > 0 \) for some alternative \( i \). It follows that:

\[
0 \leq \sum_{r \in R} \beta_r \delta_{irn} - \sum_{r \in R} \beta_r \cdot z_{j} \leq \frac{\sum_{r \in R} \beta_r \delta_{irn}}{\sum_{r \in R} \beta_r} < 1 \Rightarrow z_{j} = 0 \quad (4.30)
\]

Therefore, the combination of equations 4.27 and 4.28 implies the binary definition of conjunctive rules. Similarly, the filtering rules can be reformulated in order to replace the binary satisfactory definition with a conventional equation.
This sequence of simplifications enables us to obtain a straightforward CSGM estimation method. It had been applied for estimating the CSGM based on data collected by a survey that was conducted as part of this study. The results of this estimation are presented and discussed in Section 5.3. The formulations presented in this section may facilitate a more elaborate CSGM estimation in the future.

4.6 PATH CHOICE DECISIONS
The DPCM includes three decision models - connection, boarding and alighting (Figure 4.2). Each of these decisions may first involve the generation or adjustment of the corresponding path-set. The following presents their respective choice trees and how they are linked to path alternatives' elements.

4.6.1 CONNECTION DECISION
The first type of decision process that a traveler experiences is the need to choose at which transit stop to initiate the trip by making a non-transit connection, typically walking. A connection decision also takes place each time the traveler alights from a transit vehicle. In both cases the traveler chooses between alternative stops by evaluating all the path alternatives which connect each of the candidate stops with the traveler's final destination. Hence, the decision structure can be represented as a choice between alternative stops (Figure 4.6).

![Figure 4.6: Stop choice decision structure](image)
The set of alternative connection stops from the traveler’s current location at time \( t \), \( CS(t) \), is composed of all the stops that fulfill the following condition:

\[
\{ s_k \in CS(t) : s_k \in S_{i,2} \text{ for some } a_i \in A(t) \}
\]  

(4.31)

Where \( S_{i,2} \) is the second set of stops on the definition of path alternative \( a_i \) that belongs to the current path set \( A(t) \). The maximum size of this set is the number of stops that are connected to the traveler’s current location.

The decision structure presented in Figure 4.6 is by no means the only possible way to represent the connection choice. Alternatively, the connection decision could be considered as a multi-stage choice process as presented in Figure 4.7. Both structures are based on the differentiation between staying at the same stop and waiting for another transit service, walking to a nearby stop in order to wait there for another transit service or walk directly to the final destination. The two structures differ in the way alternatives are bundled.

This illustrates that there are several ways to represent the choice process. Different structures have implications on the set of feasible actions, the number of choice levels and ultimately the probability to make a certain decision. The choice process structure could be determined by estimating alternative model structures. Due to lack of evidence on how travelers bundle alternatives that could support a certain structure, the current implementation is based on the flat choice structure shown in Figure 4.6. This structure avoids assumptions on how stop alternatives are grouped and is consistent with the alighting stop choice structure.
Traveler is at the origin/alighted at a stop

Apply adaptive CSGM

Apply decision rules

WALK

STAY

Walk to a connected stop

Walk to destination

Apply decision rules

GO TO STOP1

GO TO STOP2

GO TO STOP3

Figure 4.7: Alternative stop choice decision structures
4.6.2 Boarding decision

When the traveler waits at a stop, each arriving transit vehicle triggers a boarding decision process. However, some transit lines may be irrelevant for the traveler’s trip and hence can be immediately dismissed (Figure 4.8) if they don’t fulfill the following condition:

$$l_{sk}(t) \in L_{i,1} \text{ for some } a_i \in A(t) \quad (4.32)$$

Where $L_{i,1}$ is the first set of lines on the definition of path alternative $a_i$ that belongs to the current path set $A(t)$. In case the arriving transit line is included in the traveler’s path set, a decision whether to board it or stay at the stop and wait for another vehicle takes place. Moreover, in case the traveler chooses to board the vehicle the execution of this decision is subject to capacity constraints.

The binary boarding decision has a well-defined set of possible actions. The path set is divided between path alternatives according to their inclusion of the arriving transit line at their first leg. As can be described formally by:

$$BD_{sk}(t) = \{a_i \in A(t) : l_{sk}(t) \in L_{i,1}\} \quad (4.33)$$

$$AS_{sk}(t) = \{a_i \in A(t) : l_{sk}(t) \notin L_{i,1}\} \quad (4.34)$$
Where $BD_{s_k}(t)$ is the set of path alternatives that are relevant for the decision to board the transit line arriving at stop $s_k$ at time $t$, notated as $l_{s_k}(t)$. $AS_{s_k}(t)$ is the set of alternatives associated with the decision to stay at the stop. These sets are mutually exclusive and collectively exhaustive ($BD \cup AS = A$, $BD \cap AS = \emptyset$).

4.6.3 ALIGHTING DECISION

An alighting decision takes place during traveler’s ride. There are two possible extreme modeling approaches with respect to the exact timing of the alighting decision: choosing between the set of potential alighting stops immediately upon boarding a vehicle versus a process that is triggered each time the vehicle approaches a stop. In the latter case, the traveler needs to identify first whether this stop is relevant based on the path-set. If the next stop is the traveler’s final destination the traveler will alight there, otherwise a binary decision whether to alight at this stop or to stay on-board is performed (Figure 4.9).

Figure 4.9: Alighting decision structure – single (left) and repetitive (right)

The single-choice approach for alighting stop decision results in a multi-choice decision where the set of candidate alighting stops $ALS_l(t)$ from line $l$ at current time $t$ when the vehicle had departed from stop $s_k$ is defined as follows:
\[ ALS_l(t) = \{ s_h \in Z^l : h > k, A^{s_h} \neq \emptyset \} \]  

Where \( Z^l \) is the set of transit stops along the route of transit line \( l \). Note that even in the case that one of the stops in the choice set is the passenger’s final destination it does not necessarily imply that this stop will be chosen, as it may be more beneficial to transfer at another stop due to shorter travel time.

A repetitive approach implies that each decision is a binary choice between alighting and staying on-board. The set of path alternatives relevant for the two possible decisions when approaching stop \( s_k \) is therefore:

\[ ALS_k(t) = \{ a_i^{s_h} \in A^{s_h} \} \]  
\[ OB_{s_k}(t) = \{ a_i^{s_h} \in A^{s_h} : \forall s_h \in Z^l, h > k \} \]

Where \( AL_{s_k}(t) \) and \( OB_{s_k}(t) \) are the sets of alternatives relevant for alighting or staying on-board respectively. In contrast to the single-choice approach, if the next stop is traveler’s final destination then the traveler will alight at this stop unconditionally as there is no possible benefit in alighting at a further downstream stop.

These modeling approaches represent the two extreme cases of no adaptation or complete adaptation with respect to the alighting decision. However, it is presumed that travelers take an intermediate approach of a pre-defined set of alighting destinations that is subject to further consideration in case of exceptional conditions. Hence, travelers make an alighting decision immediately upon boarding and may reconsider this decision in case of exceptional conditions (including the availability of information), as shown in Figure 4.2. This implies that single choice structure shown on Figure 4.9 is performed upon boarding and repeated in case of reconsideration.

### 4.7 Evaluating Path Alternatives

#### 4.7.1 The Joint Utility of an Action

All passenger decisions are modeled within the framework of random utility discrete choice models. Random utility models provide a well-established theoretical foundation for modeling the compensatory relations between alternatives’ attributes. The structure of the choice tree is fundamentally the same for all traveler decisions – boarding, alighting and connection. Each travel decision can be represented as a choice tree with a set of actions as the first level and path alternatives at the second level (see Figure 4.10). The set of actions vary between decision contexts and may correspond to stops
(connection decision or single alighting decision) or binary decision between stay and move (boarding decision, repetitive alighting decision).

Let us consider the general decision case where individual decision maker \( n \) is at certain location \( o \) with path-set \( A^{od} \). The individual has to choose an action \( c \) from the set of alternative actions \( C \). The previous section defined how \( A^{od} \) is divided over the alternative actions for different decisions. The path-set associated with action \( c \) is donated by \( A^c \subset A^{od} \). The evaluation of alternative actions requires the assessment of all the path alternatives that are associated with the corresponding element.

The utility of path alternative \( i \) is determined by its expected attributes as anticipated by individual \( n \). The deterministic part of the utility function of individual \( n \) for a single path alternative \( i \) has the following form:

\[
V_{in} = \beta_{in} X_{in} \tag{4.38}
\]

Where \( \beta_{in} \) is a vector of attributes' coefficients and \( X_{in} \) is the vector of expected values of the corresponding attributes. The expected values depend on both system conditions and passenger's perspective.

The joint utility of \( A^c \) is given by the logsum term:

\[
v_{c,n} = \ln \sum_{i \in A^c} e^{v_{i,n}} \tag{4.39}
\]

Since the logsum can capture the joint utility of all path alternatives from a given location or by using a given service, it is widely-used as an accessibility measure (Sweet, 1997) and in transport economic evaluations (De Jong et al., 2005).

In the context of the proposed path choice model, the logsum term expresses the utility of an action as a function of the utilities of the respective path alternatives. Hence,
it reflects the joint utility for a bundle of alternatives. An important property of the logsum term is that aggregating additional elements has a non-decreasing effect on the aggregated utility:

\[ 0 < \sum_{i \in I} e^{V_{i,n}} < \sum_{k \in K} e^{V_{k,n}} \Rightarrow \ln \sum_{i \in I} e^{V_{i,n}} < \ln \sum_{k \in K} e^{V_{k,n}} \Rightarrow V_{i,n} < V_{k,n} \quad I \subseteq K \] (4.40)

This property is in line with the assumption that having an additional path alternative available cannot result in decreased attractiveness and is not the average utility that counts but rather based on utility maximization theory.

The current implementation of the path choice model in BusMezzo represents all traveler decisions as a MNL model. MNL is widely-used and is derived from the assumption that the random components of the utility function are independent and identically Gumbel distributed (IID) (Ben-Akiva and Lerman, 1985; Train, 2003). Hence, the probability to choose action \( c \) from the set of alternative actions \( C \) is:

\[ p_{c,n} = \frac{e^{\mu c_n}}{\sum_{c \in C} e^{\mu c_n}} \] (4.41)

Where \( \mu \) is a scaling parameter. Note that since \( A^{od} \) is divided into mutually exclusive and collectively exhaustive subsets for the respective actions, each alternative path in \( A^{od} \) is taken into consideration exactly once in the choice probability. A well-known property of the MNL model is the independence of irrelevant alternatives (IIA). More elaborative model structures could be embedded into BusMezzo in order to account for potential overlapping among alternatives.

The following section presents the attributes that are included in the path utility function and their expected values. Note that the choice model is not applied at the path level but rather at the action level. The chosen path is merely an outcome of individual's successive decisions but passengers do not choose a path per-se at any given point along their trip.

The estimation of the DPCM requires data of complete path records with the respective sequence of decisions and the detailed conditions on the chosen and alternative actions. Smart card and AVL data collection could facilitate a comprehensive estimation. The estimation of the DPCM would enable to investigate the interdependence between successive decisions and the induced level of adaptation. Furthermore, if path choice data is available for each individual for a number of days (panel data) then the importance of habitual behavior could be analyzed. The DPCM
could be estimated by examining the probability that individual \( n \) followed a certain path alternative \( i \) when travelling between a given \( od \) pair, \( p_n(a_{i}^{od}) \):

\[
p_n(a_{i}^{od}) = p_n(c_1) \cdot p_n(c_2 | c_1) \cdot p_n(c_3 | c_1, c_2) \cdot \ldots \cdot p_n(c_{m-1} | c_1, \ldots, c_{m-2})
\]  

(4.42)

Where \( c_1, c_2, \ldots, c_m \) denote the sequence of path decisions between the pre-defined origin \( o \) and destination \( d \). Equation 4.42 formulates the path probability as the joint conditional probabilities of intermediate decisions that lead to the composition of this specific path.

The corresponding likelihood function of the estimation problem is:

\[
L^* = \prod_{j=1}^{N_{od}^{A_{od}}} \left[ p_n(a_{i}^{od}) \delta_{i,n} \right]
\]  

(4.43)

Where \( \delta_{i,n} \) is an indicator that takes ‘1’ if path alternative \( i \) was chosen by individual \( n \).

Recall from Section 4.5.2 that \( j \) is an index of individual-path combinations.

A limited estimation of the path choice model was carried out as part of this thesis. The path utility function coefficients were estimated based on a survey. The survey referred to a single boarding decision with two alternative paths. The estimated values are reported in Section 5.4.

4.7.2 **Anticipated Values of Path Alternative Attributes**

Each travel decision depends on the current individual’s expectations with respect to future travel attributes. These expectations depend on traveler’s prior-knowledge, preferences as well as the level of information that is available when making the decision. The following attributes are considered as the most important explanatory factors in transit path choice decisions (Wardman, 2004): waiting time, in-vehicle time (IVT), access and egress times, transfers and the monetary cost. The impact of these factors may vary between stops, modes and trip stages. Additional attributes such as travel time reliability and crowding conditions may have a secondary influence on traveler decisions.

The inclusion of path attributes in the DPCM depends on two requirements. First, its explanatory power is well-established and significant so that the corresponding \( \beta \) coefficient can be specified. Second, there is sufficient ground to justify the modeling of traveler’s anticipations with respect to this attribute. This requires solid knowledge on how travelers perceive and quantify an attribute and the respective prior-knowledge. Based on these requirements the current path choice implementation in BusMezzo takes into account all travel time components – IVT, walking (access, egress and
transfer) and waiting times – as well as transfers. It is assumed that in the context of high-frequency urban transit systems, travelers have a prior-knowledge of network topology, timetable travel times and planned headways.

The following sections present how traveler’s anticipation on the various trip attributes are modeled to emulate traveler’s considerations when carrying out a travel decision. The assumptions used in the current implementation are by no means the sole way to consider travelers’ anticipation. Instead, they should be viewed as an attempt to build the foundations for further enrichments such as experience, service reliability and timetable coordination.

Traveler’s anticipations are then assigned to the respective utility function components. The term anticipated value is used rather than excepted value in order to avoid potential confusion with the meaning that the latter carries in the static TAM literature where it refers to the expected value over the travelers’ population based on path shares.

It is important to keep in mind that anticipation is made at a certain time $t$ on events that will take place at some later time $t + \vartheta$. The time difference between the current time and the time where the value of an attribute is of interest, $\vartheta$, depends on the trip legs in between the respective decisions. For example, the anticipated waiting time at a downstream transfer stop takes into account when does the traveler expect to arrive at this stop which depends on the anticipated waiting time at the current stop and the IVT from the current stop to the transfer stop.

4.7.2.1 Anticipated waiting time

The anticipated waiting time at any stop $s \in S_{2j}$ on a path alternative is determined by the joint frequency of the respective set of lines $L_j$:

$$AWT_{2j}(t) = \frac{1}{2\Sigma_{l \in L_j} H_l^P(t+\vartheta)}$$ (4.44)

Where $H_l^P(t+\vartheta)$ is the planned headway on line $l$ during the relevant time period. This expression is often used in static TAM to estimate the actual waiting time, while its use in this model is limited to the estimation of travelers’ anticipation. Actual waiting times are the outcome of the dynamic progress of agents in BusMezzo and are derived from the time difference between arrival time at stops and boarding times.

Note that the underlying assumption when using equation 4.44 to express travelers’ perceptions is that their expectations are based on perfect regularity and
coordination between alternative lines. The findings of Avineri (2004) suggested that travelers indeed underestimate the waiting time. He explained this so-called ‘waiting time paradox’ through the cumulative prospect theory. People tend to underestimate their anticipated waiting time because our intuition is based on arithmetic means based on uniformity and symmetry instead of a weighted mean, also known as size-biases sampling. In order to examine this explanation, Avineri conducted an experiment which revealed that even when people are provided with headway regularity information they have misperceptions about their anticipated waiting times, plausibly because of using a low reference point. Interestingly, there is evidence that travelers overestimate their experienced waiting time (Hall, 2001; TCRP, 2003b; Mishalani et al., 2006; Psarros et al., 2011; Watkins et al., 2011).

The information that is available to a traveler when making a certain decision is determined by the dissemination means and their locations, and by individual characteristics (i.e. prior-knowledge and experience, availability of personal mobile device). Following previous studies (Nuzzolo et al., 2001; Nökel and Wekeck, 2009; Coppola and Rosati, 2009), it is assumed that passengers perceive RTI as credible and incorporate it into their adaptive decision process. This assumption is supported by several empirical studies that reported that where RTI is displayed, 70-100% of passengers use it as a source of information (TCRP, 2003b). Hence, the value of anticipated waiting time is based on the highest level of information and equals to the remaining time till the arrival of the earliest transit lines from the relevant lines and stops sets:

$$AWT_{2j}(t) = \min_{s \in S_{2j}} \min_{l \in L_j} \{RT^l_{s}(t, \theta)\}$$  \hspace{1cm} (4.45)

Where $RT^l_s(t, \theta)$ is the remaining time until the arrival of transit line $l$ at stop $s$ provided at current time $t$. The RTI may refer to another downstream stop that the traveler expects to reach at time $t + \theta$. The RTI arrival time predictions depend on the time that the inquiry refers to as well as the time that the projection is made. The increasing availability of personal mobile devices such as smart phones provides travelers an access to RTI throughout their entire trip. Transit agencies apply various methods to predict real-time arrival information. Chapter 7 presents the framework for modeling RTI in BusMezzo including the emulation of RTI generation, different levels of RTI provision and its impacts on traveler decisions.
Finally, the total anticipated waiting time of path alternative $a_i$ at time $t$ is the weighted sum over waiting times at transit stops along the path:

$$ TAWT_i(t) = \sum_{j=1}^{m} w_{2j} \cdot AWT_{s2j}(t) \quad (4.46) $$

The different weights, $w_{2j}$, may account for increasing perceived waiting time at downstream legs or different waiting conditions (e.g. seats, shelter, crowding).

Elapsed waiting time, the time already spent at the stop may affect traveler’s expectations and decisions. However, the exact impact is not clear. On one hand, the elapsed waiting time may affect travelers to anticipate shorter remaining waiting time since vehicle arrival process is not ‘memory less’. On the other hand, longer elapsing times may make travelers more inclined to take alternative options. Nuzzollo et al. (2011) provide support for the latter case, as the elapsed waiting time was found to have a positive value in the utility function of the boarding decision indicating that the longer travelers have waited the higher the prospect that they will board the next arriving vehicle.

Elapsed waiting time is considered in BusMezzo as an additional source of information that may trigger the reconsideration of the last connection decision. In case a traveler experiences waiting time that exceeds the anticipated waiting time by a certain threshold, the connection decision is reconsidered and may result in the traveler choosing to walk to another stop. An additional reason for reconsideration due to violated expectations is the case of RTI that deviated from traveler’s prior-knowledge (e.g. RTI becomes available only when arriving at the stop).

### 4.7.2.2 Anticipated in-vehicle time

The assumption that travelers have a prior-knowledge of the timetable travel times can be justified by accumulated experience that allows them to have a realistic estimation regarding riding time between pairs of stops. Therefore, the anticipated IVT on leg $j$ (that is travelling on line $l_j$ from stop $s_{2j}$ to $s_{2j+1}$) at the relevant time period is:

$$ AIVT_j(t) = ST_{s2j+1}^k - ST_{s2j}^k \quad (4.47) $$

Where $ST_{s}^k$ is the scheduled departure time of trip $k$ which is the next scheduled trip to arrive at the stop $s$. The index $k$ can be formulated as follows:

$$ k = \arg \left( \min_{k \in K^l} \{ST_{s2j}^k : ST_{s2j}^k > t + \vartheta \} \right) \quad (4.48) $$

Where $K^l$ is the set of trips assigned to line $l$. Note that path alternative definition implies that it does not matter which $l_j \in L_j$ is used for the calculation as only transit
lines with the same riding time between consecutive stops can form a single alternative. This expression will take into account the case of timetables that reflect the varying transit conditions on different times-of-days.

In some cases passengers may anticipate IVT that differ from the timetable travel times. Such anticipation can be due to static information (e.g. constructions, events) or based on RTI provision on service disruption downstream. In these cases the anticipated IVT is directly derived from the information provided.

The total anticipated IVT for path alternative \( a_i \) at time \( t \) is the sum over the alternative's legs:

\[
TAIVT_i(t) = \sum_{j=1}^{m} w_j \cdot AIVT_j(t)
\]

Having a weighting factor \( w_j \) allows to incorporate a non-uniform value of time on different legs due to their position along the path or transit mode-specific coefficients (e.g. bus, tram, metro).

### 4.7.2.3 Anticipated connection time

Since connection legs refer to non-transit trip segments, their travel times do not depend on transit network configuration. The anticipated connection time depends on the speed of the relevant non-transit modes. In the case of walking connections, anticipated connection time depends on the walking distance and the assumed walking speed which varies among travelers. The total anticipated connection time on alternative \( a_i \) for individual traveler \( n \) is calculated in a similar fashion to the total IVT:

\[
TACT_{i,n} = \sum_{j=1}^{m+1} w_j \cdot \frac{d_{c_j}}{v_{c_j,n}} \quad c \in C_j
\]

Where \( d_{c_j} \) is the length of connection \( c_j \) (that is travelling from stop \( s_{2j-1} \) to \( s_{2j} \)) and \( v_{c_j,n} \) is the velocity of traveler \( n \) on connection \( c_j \) (velocity would become time-dependent if the model would be extended to accommodate car-based connections as park and ride).

The path alternative definition implies that the calculation is the same for every \( c \in C_j \). As with the IVT, the weighting factor can account for differences between legs due to their position along the path (access, transfer, egress).

The current implementation in BusMezzo considers only walking connections. Each traveler is assigned with a walking speed drawn from a truncated normal distribution: \( v_n \sim \text{truncN}(4,1) \). This walking speed is then used to calculate the walking times on all connection legs: access, between stops, transfer and egress.
4.7.2.4 Transfers
Previous studies found that the need to transfer between transit lines is associated with a higher disutility than the respective waiting and walking times due to subjective discomfort such as stress, uncertainty and the disruption to on-board activity. BusMezzo accounts explicitly for the transfer-related waiting and walking times. An additional transfer penalty is included in the utility function. The total transfer penalty for alternative \( a_i \) is the sum of transfer penalties on each of alternative’s legs:

\[
TRANS_{i,n} = \sum_{j=1}^{m-1} TRANS_{j,n}
\]

Where \( TRANS_{j,n} \) is the transfer penalty of individual \( n \) caused by transferring from \( L_j \) to \( L_{j+1} \) as part of path alternative \( a_i \). The current implementation in BusMezzo assigns a uniform transfer penalty. The model can be refined to assign different transfer penalties to various mode-combinations, travelers’ groups and transit sites (Iseki and Taylor, 2009; Guo and Wilson, 2010).

4.7.2.5 Monetary cost
According to TCRP (2003c), fare strategies are classified into flat fare structures and differential fare structures. In the case of the former, the fare is fixed regardless of the travel distance or mode. In contrast, with a differential fare structure fares can vary by distance, number of passed zones and mode (e.g. intercity vs. regional service). The report found that the vast majority of transit agencies use a flat fare structure, in particular for urban transit systems that consist of bus, light rail train and metro services. Moreover, most agencies offer a free transfer, usually in the form of a time-interval free pass (e.g. hourly, daily). Many of those agencies that have a differential fare use zonal fares, where trip fare depends on the number of zones that the passenger passes during the trip. This typically implies a fix number of zones for a given OD.

Therefore it can be concluded that in most urban transit systems trip fare is constant for a given OD pair, regardless of the chosen path. In these cases, trip fare is not a factor in the path choice process. In the case of a flat fare structure without free transfers, the number of transfers determines the total trip fare and the monetary cost disutility can be incorporated into the transfer penalty coefficient. Therefore, although this model does not account for the monetary cost, this is not expected to bias the results under most circumstances.
5. Transit Path Choice Model Estimation

A web-based survey was conducted as part of this thesis in order to collect data to estimate the choice-set generation model (CSGM) and the dynamic path choice model (DPCM) that were presented in Chapter 4. An important component of this survey is the explicit investigation of the choice-set composition process. The questionnaire includes both Revealed Preferences (RP) and Stated Preferences (SP) sections (see Hicks and Turner (1999) for details on experiment design). This chapter presents the design of the questionnaire as well as a summary of the results. First, the general questionnaire structure is presented followed by descriptions of each of the five questionnaire parts. Second, the results of the survey are discussed (5.2). Third, the estimation results of the CSGM and the path utility function are presented in Sections 5.3 and 5.4, respectively.

5.1 Questionnaire Design

The questionnaire was designed to enable the collection of data regarding both the choice-set generation and the dynamic path choice processes. Consideration was given to make the choice situations custom and familiar to potential respondents in order to contribute to the realism of their responses. In order to simplify questionnaire design and analysis, questionnaire sections that are associated with a specific transit path choice were limited to trips destined at the Technion in Haifa, Israel. The questionnaire language was Hebrew and the main target population was frequent transit travelers to the Technion – students and staff. However, some questionnaire sections were generic and applicable also to other potential respondents. Table 5.1 presents the five sections that the questionnaire was composed of - general information, chosen transit route, transit consideration-set, dynamic transit path choice and transfer location choice – as well as their respective content, question type and generality.
<table>
<thead>
<tr>
<th>Section</th>
<th>Content</th>
<th>Questions type</th>
<th>Generality</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>General information</td>
<td>Multi-choice</td>
<td>Universal</td>
</tr>
<tr>
<td>B</td>
<td>Chosen transit path</td>
<td>Multi-choice</td>
<td>Path-specific</td>
</tr>
<tr>
<td>C</td>
<td>Transit consideration-set</td>
<td>Multi-choice</td>
<td>Path-specific</td>
</tr>
<tr>
<td>D</td>
<td>Dynamic transit path choice</td>
<td>Binary-choice</td>
<td>Universal</td>
</tr>
<tr>
<td>E</td>
<td>Transfer location choice</td>
<td>Rating</td>
<td>Universal</td>
</tr>
</tbody>
</table>

General information regarding socio-economic properties and travel habits that may affect transit path choice was first collected (Section A). Respondents were asked to report a recent actual transit path choice in order to incorporate real-world path choice in the analysis (Section B). The following SP section refers to the composition of the consideration-set (Section C). In addition, SP questions enable to construct an experimental design that supports the analysis of controlled factors by determining their levels. Moreover, the experimental design allows minimizing the variability of factors that lie outside the scope of this study. Respondents were given a series of dynamic path choice decisions - each with two transit alternatives to choose from (Section D). Finally, the respondents were asked to rate the importance of various attributes when choosing a transfer location (Section E).

The questionnaire was constructed and distributed using a web-based platform called ‘Qualtrics’ (www.qualtrics.com). The main advantages of using a web-based platform are:

- Adaptation – designing an adaptive questionnaire where onward questions depend on previous responses. This attribute allows to present only relevant travel alternatives implied by previous travel choices (e.g. bus lines which are available at a given transfer stop).
Randomization – randomizing questions and alternative answers ordering as well as sampling a subset of questions from a larger set. The later feature is especially useful in order to test various choice scenarios, which is an important advantage in terms of statistical analysis properties.

Data collection – questionnaire distribution, respondent participation and data coding are done with reduces manual effort and are less subject to human errors. In addition, a web-based platform is useful for targeting a larger sample in a shorter time period.

Figure 5.1 presents the flowchart of the questionnaire. The questionnaire started with an opening message. The message described the purpose of the questionnaire, its structure, expected duration and the confidentiality of the information provided, followed by general information questions. According to the information provided on travel habits to the Technion, it was determined whether the respondent is relevant for the trip-specific sections. In case that the respondent is a frequent transit traveler to the Technion, the respondent had to provide details on the most recent transit journey. The third section included the specification, selection and rating of alternative transit paths to the Technion. The fourth section, transit path choice, is relevant for all survey respondents and is the second part for those who are not frequent transit travelers to the Technion. It was followed by a short section concerning the transfer stop choice. The questionnaire was concluded with a closing message paying gratitude to the respondent and providing contact details. The questionnaire was designed to adapt to previous data provided by the respondent. The following sections describe in detail the questionnaire sections one by one. See Appendix C for selected screenshots from the questionnaire.
Figure 5.1: Questionnaire structure flowchart
5.1.1 *GENERAL INFORMATION*

The first section contained several questions regarding the respondent socio-economic attributes in order to analyze their importance in path choice decisions. Respondents were asked to provide the following information:

- Gender
- Age group
- Average income per household member (relative to the national average)
- Car ownership
- Common travel modes (possibly selecting multiple modes)
- Travel frequency to the Technion

Only respondents who are both transit users on a regular basis and travel to the Technion on a regular basis participated in the two following sections (Sections B and C). Otherwise the respondent was transferred directly to the dynamic path choice part (Section D) of the questionnaire.

5.1.2 *CHOOSEN TRANSIT PATH*

Respondents that use transit modes and visit the Technion regularly are asked to report their most recent transit trip to the Technion. This part of the questionnaire collects information regarding respondents’ RP. In order to understand how the chosen transit route was specified, it is necessary to describe the transit system in Haifa metropolitan area.

Haifa is the third largest city in Israel with 265,000 inhabitants in 2009. Haifa metropolitan area is the second largest in Israel with a population of one million. The natural geographical borders of Haifa (to the west and north- the Mediterranean Sea; to the south- the Carmel Mountain) result in two distinctive main entrees to the city (Figure 5.2). Since 2002 the city of Haifa has two central bus stations (CBS) organized in architecture of gate terminals: a southern gate (’Hof-Hacarmel’/Coastline) and an eastern-northern gate (’Hamifratz’/Bay). This architecture implies that all suburban and intercity bus trips have to include a transfer at one of these CBS. Moreover, a hierarchical service design was introduced with major lines connecting the two CBS and minor lines providing service at the district level. Buses account for more than 90% of transit trips in Haifa (Matat, 2010). There are two train stations that are directly linked to each of the CBS. Two additional train stations are located along the coastal railway in
between the CBSs. The coastal railway is used both for intercity lines connecting Haifa to northern and southern destinations including Tel-Aviv metropolitan area as well as for a regional service for some of Haifa’s suburbs.

Figure 5.2: A satellite image of the city of Haifa with the locations of CBS and train stations. The colored areas correspond to the 10 origin zones defined in the survey (drawn in Google earth)

Haifa’s transit architecture with very few restricted and distinguished entrances to the city of Haifa provides an advantageous layout for the design and analysis of the questionnaire. For the purpose of this survey, it was decided to consider only the part of the trip which is within the boundaries of the city of Haifa. Therefore, if the respondent’s origin lies beyond these boundaries, the questionnaire refers only to the part of the trip that started at one of the four possible transfer locations from the regional and national service to the local service.

Transit path trips are classified by the OD pair and the number of trip segments. Since the destination is identical for all respondents (the Technion), the respondents only have to specify their origin stop by using free text and selecting the origin zone from a list of 10 possible origin zones in the city of Haifa (Figure 5.2). The 10 geographical zones were defined to include distinctive neighborhoods and activity
generators and are consistent with previous definitions used by the central bureau of statistics and local transport planners. In addition, the respondents provided the number of trip sections (bus lines, train, park and ride, etc.) used before arriving in Haifa and the number of local buses used for performing the reported trip. Respondents selected the type of traveling ticket that was used for this trip. It should be noted that along with introducing the current transit architecture, the local bus operator (‘Egged’) introduced an hourly pass that allows passengers to board an unlimited number of buses during an hour starting from the first boarding without an extra monetary cost.

The details of the actual trip are essential in order to provide respondents only relevant and valid onward bus lines and transfer stop alternatives. This design results with a conditional sequence of path-adaptive multi-choice questions. For example, if the respondent specified the bay CBS as the origin stop in Haifa and indicated that the local part of the trip included one transfer stop then the respondent was asked three questions: select the bus line that you boarded at the bay CBS from a list of relevant bus lines; select the stop where you had alighted from the bus and transferred to another bus line from a list of possible locations, conditional on the selected bus line and; select the bus line that you boarded at the transfer stop from a list of buses that connect the selected transfer stop and the Technion.

5.1.3 Transit consideration-set
One of the main objectives of this survey is to gain knowledge on how transit users construct their consideration-set. Unlike the chosen alternative, the set of alternatives that people consider before making a decision is not an observable behavior. Therefore this section is a SP part where respondents are asked to provide details on their consideration-set in the following forms:

1) Specify – indicating the number of reasonable alternatives to travel by transit from their origin stop to the Technion. They were instructed to consider only paths that are plausible alternatives that they may actually undertake. The respondents specified each of the alternatives in the same fashion as the chosen transit route details had been reported. The details were given with path-adaptive multi-choice questions. This ‘free association’ report may provide an unbiased insight into respondents’ consideration-sets.
2) Select - selecting all the alternatives that they perceive as plausible from a given set. A master-set, a list of reasonable transit paths was pre-defined for each of the 10 origin zones based on the researcher familiarity with the network and a pilot study. Respondents could also indicate alternatives that they did not include in the ‘free association’ part.

3) Rate - rating each of the transit paths included in the given master-set on a five grades scale. Respondents could indicate that they are unfamiliar with a certain alternative or unable to evaluate it. Alternatives’ ratings are useful when analyzing the relation between the inclusion of alternatives in the consideration-set and their perceived attractiveness.

The design of this section was aimed to facilitate the analysis of consideration-set properties as well as the estimation of a CSGM. As was pointed out in a synthesis of route choice-set research by Bovy (2009), only a couple of previous studies have collected and analyzed data on choice-sets: Prato and Bekhor (2006) in the context of car trips and Hoogendoorn-Lanser and Van Nes (2004) with respect to multi-modal trips that include rail as the main trip leg. Thus, the choice-set data collected in this survey is the first of its kind.

5.1.4 Transit path choice

The formulation and specification of the path utility function requires estimating the importance of various trip attributes in passenger decision making. While there are clear advantages in collecting data on travelers’ actual trip decisions (RP), there also several important drawbacks and limitations. The decision factors that will be included in the analysis are limited to the decision circumstances and may exercise low variability that will not allow drawing conclusions. Moreover, it is difficult to obtain the values of all relevant decision factors (e.g. waiting time, walking time, level of comfort). It also limits considerably the number of observations per respondent due to memory constraints. Automatic data collection methods based on tracking the location of personal mobile devices in addition to AVL and APC data may allow in the future the analysis of RP trip decisions.

An experimental design of SP questions is composed of generic choice conditions that avoid circumstance-specific constraints. In addition, a relative large number of choice scenarios are feasible. Partial factorial design allows avoiding the construction of
all possible factor values combinations, while still enabling the identification and estimation of various factors. The design of a SP choice experiment introduces a compromise between information richness and completeness on one hand, and the bias involved with exhausted and impatience respondents on the other hand. Randomizing scenarios so that each respondent receives a subset of choice scenarios is an effective way to balance between these two requirements (Box, 1978).

The experimental design of the choice scenarios takes into account the number of factors involved with scenario description, number of levels for each factor, number of combined choice scenarios and the desired number of choices per respondent. The choice scenarios were designed to imitate real-world dynamic boarding decisions, where individuals decide whether to board an arriving bus or stay at the stop and wait for another bus. Moreover, the decision factors and their corresponding values are supposed to replicate the incomplete information that transit users possess in reality.

Table 5.2 summarizes the choice experiment factors and their respective levels. The following factors were included in the choice experiment:

- Number of transfers (TRANS) – the number of buses that the individual has to board and alight in order to get from the origin to the destination. Every pair of alternatives includes one alternative with a direct connection and one that includes a single transfer.

- Waiting time for the first bus (WT1) – estimated time till the first bus will arrive at the origin stop. Every pair of alternatives includes one alternative that its first bus arrives immediately (no waiting time), while the other alternative involves some waiting time. Note that there is no real-time passenger information system currently in place in Haifa. Therefore, waiting time figures are not regarded as definitive and may be interpreted as passenger expectation based on frequency, elapsed waiting time and experience.

- Schedule adherence (SA1) – in case that the bus does not arrive immediately then the level of accuracy of the waiting time is given. Certainty level is given as the probability that the bus will come within a 5 minutes window from the specified time, between 5 minutes before the specified time to 5 min after the specified time. This definition was chosen in order to avoid complicated reliability measures that are difficult to grasp (e.g. standard deviation).
- In-vehicle time on the first bus \((IVT1)\) – estimated riding time from the origin stop to the alighting stop from the first bus line. Riding time figures may represent passenger expectations based on experience or timetables.

- Riding conditions on the bus vehicle \((C1)\) – whether there is an available seat on the bus or not. This information is given only for the bus that arrives now. It reflects the possibility to directly observe the riding comfort on an approaching bus.

- Waiting time for the second bus \((WT2)\) – estimated time until the second bus will arrive at the transfer stop in case that the trip includes a transfer. The uncertainty involved with this estimation is not included in this experiment.

- Possibility of denied boarding due to an overcrowded bus \((DB2)\) – the probability of not been able to board the next bus vehicle of the second bus line and therefore having to wait for the consecutive bus on this line. Denied boarding may occur on high-demand lines during peak periods and is a source of dissatisfaction and long waiting times.

- In-vehicle time on the second bus \((IVT2)\) – estimated riding time from the transfer stop to the destination stop from the second bus line in case that the trip includes a transfer.

- Egress walking time \((ET)\) – estimated walking time from the last stop, where the passenger alights, to the final destination. It was decided to give walking time rather than walking distance as walking speed varies among individuals and is perhaps less comprehensible.
Table 5.2: Route choice experiment factors

<table>
<thead>
<tr>
<th>F</th>
<th>Variable</th>
<th>Factor</th>
<th>Relevancy</th>
<th>No. of level</th>
<th>Units</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1</td>
<td>TRANS</td>
<td>Number of transfers</td>
<td>Always</td>
<td>2</td>
<td></td>
<td>0*/1*</td>
</tr>
<tr>
<td>F2</td>
<td>WT1</td>
<td>First bus Waiting time</td>
<td>Always</td>
<td>4</td>
<td>Minutes</td>
<td>0*/5/10/15</td>
</tr>
<tr>
<td>F6</td>
<td>SA1</td>
<td>Schedule adherence</td>
<td>Conditional</td>
<td>2</td>
<td>Percent-ages</td>
<td>50/80</td>
</tr>
<tr>
<td>F3</td>
<td>IVT1</td>
<td>In-vehicle time</td>
<td>Always</td>
<td>4</td>
<td>Minutes</td>
<td>10*/20/30/40</td>
</tr>
<tr>
<td>F8</td>
<td>C1</td>
<td>Available seat</td>
<td>Conditional</td>
<td>2</td>
<td>Binary</td>
<td>Yes/No</td>
</tr>
<tr>
<td>F4</td>
<td>WT2</td>
<td>Second bus Waiting time</td>
<td>Conditional</td>
<td>3</td>
<td>Minutes</td>
<td>5/10/15</td>
</tr>
<tr>
<td>F9</td>
<td>DB2</td>
<td>Probability for denied boarding</td>
<td>Conditional</td>
<td>2</td>
<td>Percent-ages</td>
<td>10/30</td>
</tr>
<tr>
<td>F5</td>
<td>IVT2</td>
<td>In-vehicle time</td>
<td>Conditional</td>
<td>3</td>
<td>Minutes</td>
<td>5/10/15</td>
</tr>
<tr>
<td>F7</td>
<td>ET</td>
<td>Egress walking time</td>
<td>Always</td>
<td>4</td>
<td>Minutes</td>
<td>0*/5/10/15</td>
</tr>
</tbody>
</table>

(* indicates a value that is always assigned to one of the alternative choices)

Applying the experimental design recommended by Box (1978) and Rose and Bliemer (2009), the choice experiment includes 16 choice scenarios. Table 5.3 presents the scenarios, where (+1) and (-1) are common notations for binary construction of choice scenarios. Note that the number of choice combinations is not a simple multiplication as some factors are conditional on other factor values. Each experiment involves a binary choice between the scenario displayed in Table 5.3 and the scenario defined by the complementary matrix. The interpretation of this experimental design in terms of factor values is given in Table 5.4. The total number of paired alternatives is 972. All of those choice experiments were included in the survey and were coded into the questionnaire platform. Each respondent received 16 questions, with each question sampled from a set of scenario combinations assigned to it (between 27 and 108 alternative combinations). Since ‘Qualtrics’ does not have an experimental design
software, the alternatives had to be formulated and coded manually into the questionnaire. This experimental design structure results in a total number of potential questionnaires of $5.23 \cdot 10^{27}$.

Special attention was given to the format and instructions that accompanied this section. The different factors were explained in a straightforward manner followed by a choice experiment example (see Appendix C). Respondents were instructed to choose the alternative that they prefer from each pair of alternatives, emphasizing that there is no correct answer.

Table 5.3: Experimental design of the choice scenarios

<table>
<thead>
<tr>
<th>No.</th>
<th>F1</th>
<th>F2</th>
<th>F3</th>
<th>F4</th>
<th>F5</th>
<th>F6</th>
<th>F7</th>
<th>F8</th>
<th>F9</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>+1</td>
</tr>
<tr>
<td>2</td>
<td>+1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>+1</td>
<td>-1</td>
<td>+1</td>
<td>+1</td>
<td>-1</td>
</tr>
<tr>
<td>3</td>
<td>-1</td>
<td>+1</td>
<td>-1</td>
<td>-1</td>
<td>+1</td>
<td>+1</td>
<td>-1</td>
<td>+1</td>
<td>-1</td>
</tr>
<tr>
<td>4</td>
<td>+1</td>
<td>+1</td>
<td>-1</td>
<td>-1</td>
<td>+1</td>
<td>+1</td>
<td>-1</td>
<td>+1</td>
<td>-1</td>
</tr>
<tr>
<td>5</td>
<td>-1</td>
<td>-1</td>
<td>+1</td>
<td>-1</td>
<td>+1</td>
<td>+1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>6</td>
<td>+1</td>
<td>-1</td>
<td>+1</td>
<td>-1</td>
<td>-1</td>
<td>+1</td>
<td>-1</td>
<td>+1</td>
<td>-1</td>
</tr>
<tr>
<td>7</td>
<td>-1</td>
<td>+1</td>
<td>+1</td>
<td>-1</td>
<td>-1</td>
<td>+1</td>
<td>+1</td>
<td>-1</td>
<td>+1</td>
</tr>
<tr>
<td>8</td>
<td>+1</td>
<td>+1</td>
<td>+1</td>
<td>-1</td>
<td>+1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>9</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>+1</td>
<td>-1</td>
<td>+1</td>
<td>+1</td>
<td>+1</td>
<td>-1</td>
</tr>
<tr>
<td>10</td>
<td>+1</td>
<td>-1</td>
<td>-1</td>
<td>+1</td>
<td>+1</td>
<td>+1</td>
<td>-1</td>
<td>-1</td>
<td>+1</td>
</tr>
<tr>
<td>11</td>
<td>-1</td>
<td>+1</td>
<td>-1</td>
<td>+1</td>
<td>+1</td>
<td>-1</td>
<td>+1</td>
<td>-1</td>
<td>+1</td>
</tr>
<tr>
<td>12</td>
<td>+1</td>
<td>+1</td>
<td>-1</td>
<td>+1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>+1</td>
<td>-1</td>
</tr>
<tr>
<td>13</td>
<td>-1</td>
<td>-1</td>
<td>+1</td>
<td>+1</td>
<td>+1</td>
<td>-1</td>
<td>-1</td>
<td>+1</td>
<td>+1</td>
</tr>
<tr>
<td>14</td>
<td>+1</td>
<td>-1</td>
<td>+1</td>
<td>+1</td>
<td>-1</td>
<td>-1</td>
<td>+1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>15</td>
<td>-1</td>
<td>+1</td>
<td>+1</td>
<td>+1</td>
<td>-1</td>
<td>+1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>16</td>
<td>+1</td>
<td>+1</td>
<td>+1</td>
<td>+1</td>
<td>+1</td>
<td>+1</td>
<td>+1</td>
<td>+1</td>
<td>+1</td>
</tr>
</tbody>
</table>
Table 5.4: Factor values on choice scenarios

<table>
<thead>
<tr>
<th>#</th>
<th>Alternative 1</th>
<th>Alternative2</th>
<th>#combinations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F1</td>
<td>F2</td>
<td>F3</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>5/10/15</td>
<td>10</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>5/10/15</td>
<td>10</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>20/30/40</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>0</td>
<td>20/30/40</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>5/10/15</td>
<td>20/30/40</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>5/10/15</td>
<td>20/30/40</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>5/10/15</td>
<td>10</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>5/10/15</td>
<td>10</td>
</tr>
<tr>
<td>13</td>
<td>0</td>
<td>0</td>
<td>20/30/40</td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>0</td>
<td>20/30/40</td>
</tr>
<tr>
<td>15</td>
<td>0</td>
<td>5/10/15</td>
<td>20/30/40</td>
</tr>
<tr>
<td>16</td>
<td>1</td>
<td>5/10/15</td>
<td>20/30/40</td>
</tr>
</tbody>
</table>

(NR = not relevant)
5.1.5 Transfer stop choice

Very often passengers have several possible locations to transfer between transit lines. A DPCM has to replicate the transfer location. The final section of the questionnaire asked respondents to rate the importance of five factors in choosing where to transfer. The factors were phrased as statements that the respondents had to indicate their position towards them using a five grades scale. The statements referred to the importance of the location of the transfer stop along the route, walking distance between stops, service frequency at the transfer stop and waiting conditions at transfer facilities.

5.2 Summary statistics

The questionnaire was distributed during August 2009. A total number of 158 respondents participated in the survey. A summary and analysis of the results from each of the 5 questionnaire parts is presented in the following sections. Estimation results are presented in Sections 5.3 and 5.4.

5.2.1 General information

From the 158 respondents, 84 are men (53%) and 74 are women (47%). The age distribution was as follows: 20% (N=32) of the respondents are between 18 and 25 years old, 67% (N=106) between 25 and 39 years old and 13% (N=20) above 40. This reflects the high share of students among the respondents. Figures 5.3-5.5 present the distributions of household income, car ownership and common travel mode, respectively. It can be summarized that the common survey respondent is a 25-39 years old man or woman with a mid-high income who lives in a household that owns one car and uses car and bus on a regular basis. Half of the respondents use buses on a regular basis. It is interesting to note that 52% of the respondents that use car frequently use it as an exclusive mode.
Figure 5.3: Relative household income distribution of survey respondents (N=158)

Figure 5.4: Car ownership distribution among survey participants (N=158)
The relative income distribution can be transformed to the 64\textsuperscript{th} percentile of the national income distribution. This income percentile was equivalent in 2009 to approximately 11,500 NIS or 2,200 Euros, before taxes (CBS, 2009). This income percentile corresponds to a car ownership share of 63\% (JIIS, 2008). This is close to the 66\% of the respondents that live in a household that owns at least one car. Relatively to the national and metropolitan mode share of transit, a large share of respondents uses public transport modes on a regular basis (Matat, 2010).

The majority of the respondents, 80\% (N=126) travel to the Technion on a regular basis. The majority of those travelling regularly, travel to the Technion at least three times a week (N=69) (Figure 5.6).
Figure 5.6: Frequency of trips to the Technion (N=126)

5.2.2 RP TRANSIT PATH

The RP section regarding the most recent transit trip is relevant only to those respondents who travel to the Technion and use public transport on a regular basis - 59% (N=93) of the respondents.

The majority of the sample started their trip in Haifa at one of its CBS: 40% (N=38) at the southern gate of Hof Hacarmel CBS and 22% (N=20) at the eastern gate of Bay CBS, as shown in Figure 5.7. Besides the CBSs, the most important origin was Neve Sheannan (a residential neighborhood in proximity to the Technion). Five other origin zones have a small fraction of respondents of 4-5% each. Two origin zones were not selected by any respondent (Hadar and Grand Mall).
Figure 5.7: Number of respondents per origin travel zone (N=93)

Figure 5.8 presents the market share of travel ticket types that were used for the reported trip. A single ticket which is typical to non-commuters was used by 35% of the respondents.

![Bar chart showing travel ticket type share](chart.png)
In line with the distribution of origin zones (Figure 5.7), 40% (N=37) of the respondents started their trip in Haifa and therefore had no trip leg outside Haifa (Figure 5.9). The average number of trip legs was 2.37. Table 5.5 shows how the number of trip legs varies between trips that their origin is in Haifa compared with other trips and where the transfers took place. The relatively high number of trip legs is expected due to the transit system structure (see Section 5.1.2). One third (N=31) of the trips involved some transfer in Haifa. Interestingly, the distribution of trip legs highly resembles the figures reported for the very different transit system of Seoul based on an automated fare collection system (Jang, 2010). More than 80% of the transfers took place at one of two locations – either Ziv center (58%) or Horev center (23%). The attributes of RP paths are further discussed in Section 5.2.3.2 in comparison with paths that are included in the consideration-set.

Table 5.5: Number of trip legs by origin and location

<table>
<thead>
<tr>
<th>Trip Origin \ Location</th>
<th>Before Haifa</th>
<th>In Haifa</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Haifa</td>
<td>---</td>
<td>1.46</td>
<td>1.46</td>
</tr>
<tr>
<td>Elsewhere</td>
<td>1.64</td>
<td>1.34</td>
<td>2.98</td>
</tr>
<tr>
<td>All respondents</td>
<td>0.99</td>
<td>1.38</td>
<td>2.37</td>
</tr>
</tbody>
</table>

Figure 5.9: Number of trip legs distribution (N=93)
5.2.3 SP TRANSIT CONSIDERATION-SET

This section had asked respondents to specify, select and rate path alternatives for the same trip that they reported on the preceding RP section. The size of the choice-set is first investigated followed by an analysis of the attributes of paths that are included in the choice-set.

5.2.3.1 Consideration-set size

Respondents were first asked to state the number of reasonable alternatives that there are considerable chances that they will actually undertake. Then they were asked to specify them one by one. The number of alternatives remained almost the same at the specification phase, as evident in Figure 5.10. The distribution of the consideration-set size is presented in Figure 5.11, with 3 path alternatives being the mode. The average number of specified paths was 3.15, with 75% having no more than 3 alternatives and 88% having up to 4 alternatives. The maximum number of specified alternatives was 9. This distribution resembles closely the consideration-set size distribution presented in Bovy and Stern (1990) based on results of Benshoof (1970) who analyzed the number of routes considered by motorists. Furthermore, these results are consistent with the cognitive psychological theory (see discussion in Hoogendoorn-Lanser, 2005).

Interestingly, the number of specified alternatives was positively correlated with how frequently the respondent travels to the Technion ($r = 0.22$). It may be argued that higher travel frequency results in higher familiarity and awareness of alternative paths.

![Figure 5.10: Cumulative distribution function of number of declared and specified alternatives (N=93)](image-url)
Following the specification of plausible alternative paths, respondents were asked to select all the attractive paths from a master-set that was pre-defined for their OD (see Section 5.1.3). Note that the definition of an alternative path is consistent with the definition used in the proposed model (Section 4.4). Path alternatives were given in terms of bus lines and transfer stops, without any reference to origin stops, access and egress modes. Therefore, the size of the master is rather small with a maximum number of 9 alternatives. The number of trip combinations (specific lines and transfer stop combinations) that can be derived from the given master-sets is in the range of 10 to 44. The average number of paths selected to the consideration-set is 2.32 alternatives. A large share of single-alternative consideration-sets was induced by trips originated at Neve Sheannan (NS) that lies in close proximity to the Technion. If these observations are excluded, the average number of selected alternatives is 2.47. The distribution of the number of selected alternatives resembles that of specified alternatives with a slight shift to the left (compare Figure 5.12 with 5.11). The smaller consideration-sets compared with specified consideration-sets is due to variations in the definition of alternative path. Some respondents specified two parallel routes (‘common lines’) as a single alternative while others treated them as two distinguished alternatives. Therefore, the number of alternatives specified by the respondents can be misleading as it may be the result of different interpretation of the path alternative concept. In contrast, the number of alternatives selected at the selection phase is unambiguous and
consistent with the definition used in our model. These results are in line with previous results from The Netherlands. Fiorenzo-Catalano et al. (2004) and Hoogendoorn-Lanser and Van Nes (2005) found that Dutch travelers reported consideration-sets ranging between one and six path alternatives with an average size of about two alternatives per traveler, based on a large travel habit survey.

![Diagram: Number of selected path alternatives (N=93)](image)

The cumulative distribution function of the ratio between the size of the selected consideration-set and the size of the given master-set is presented in Figure 5.13. Most consideration-sets were within 40-70% the size of the given master-set. It is presumed that the master-set grows with more intensive and connected networks more than the consideration-set size does.
Figure 5.13: The CDF of the ratio between the selected consideration-set and the given master-set

5.2.3.2 Path alternative attributes
The number of transfers is an important determinant of both the inclusion probability and the rating of a path alternative. The inclusion probability of an alternative \( i \) is defined as \( y_{it}^{\text{od}} = \frac{\Sigma_{n} y_{in}^{\text{od}}}{\Sigma_{n} y_{in}^{\text{od}}} \) (see Section 4.52), which is the share of respondents that selected this alternative for a given OD. The inclusion probability of a direct path is 98% compared with only 48% for single transfer alternatives and falling to 25% for alternatives that involve two transfers (Figure 5.14). None of the respondents specified or selected an alternative that requires more than two transfers in the choice-set. It should be noted that each master-set included a direct alternative.

The respondents were also asked to rate the alternatives that were included in the master-set. The overall average score was 3.11 on a 1 to 5 scale. However, the average rating for an alternative that was selected at the previous phase was 3.81, compared with 1.80 for alternatives that were not selected. With respect to the number of transfers, a similar pattern emerged as in the selection phase: the average grades were 4.54, 2.75 and 1.74 for direct, single transfer and two transfer paths, respectively.
In addition to the number of transfers, the ratio between total IVT and the total IVT of the shortest alternative is also a potential explanatory variable of path inclusion and rating. The IVT ratio is defined as:

\[
IVT_{Ri} = \frac{IVT_i}{\min_{a \in A_{od}} IVT_i}
\]

(5.1)

Where:

\[
IVT_i = \sum_{j=1}^{n} (ST_{s_{2j+1}}^{lj} - ST_{s_{2j}}^{lj})
\]

(5.2)

Where \(ST_{s}^{lj}\) is the scheduled departure time of line \(l_j\) at stop \(s\). This factor can be used in the CSGM to filter paths that are very long time-wise compared with other available alternatives. Figure 5.15 shows the cumulative distribution function for IVTR for both the chosen-path (RP) and for paths included in the consideration-set. Only 5% of the chosen-paths have an IVT which is more than 20% longer than the path with the shortest IVT for the respective OD pair. As expected, the criterion for being included in the consideration-set is lower, with 21% exceeding by more than 20% and 14% having an IVT which is more than 1.5 times the shortest path.

Figure 5.14: Inclusion probability and the average rating of path alternatives by the number of transfers
Table 5.6 presents the shares of chosen-paths and consideration-set paths that satisfy various criteria. The number of transfers never exceeded by more than one transfer the minimum number of transfers required for a given OD for chosen-paths. A small share of consideration-set paths includes one extra transfer compared with the minimum required. In addition to the number of transfers and the IVTR, several dominancy rules were examined. As discussed in Section 4.5.1, dominancy rules can be used to exclude alternatives that are dominated – inferior relative to another alternative in at least one aspect and not better-off than this alternative in all other respects.
Table 5.6: The share of chosen and consideration-set paths that satisfy various criteria

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Chosen path</th>
<th>Consideration-set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Up to 1 extra transfer</td>
<td>100.0%</td>
<td>91.4%</td>
</tr>
<tr>
<td>Up to 2 extra transfer</td>
<td>100.0%</td>
<td>100.0%</td>
</tr>
<tr>
<td>IVTR of less than 1.2</td>
<td>94.9%</td>
<td>79.2%</td>
</tr>
<tr>
<td>IVTR of less than 1.5</td>
<td>97.5%</td>
<td>87.3%</td>
</tr>
<tr>
<td>Not dominated by transfers and IVTR</td>
<td>69.6%</td>
<td>48.0%</td>
</tr>
<tr>
<td>(1) Not dominated by transfers and IVTR – disabling low-frequency paths (expression 6.5)</td>
<td>72.2%</td>
<td>62.8%</td>
</tr>
<tr>
<td>(1) + perception threshold of 0.2</td>
<td>93.7%</td>
<td>89.3%</td>
</tr>
</tbody>
</table>

Alternative $a_1$ is dominated by alternative $a_2$ if both conditions- 5.3 and 5.4 – are true and at least one of them holds true also with a strict inequality:

\[
\text{TRANS}_{a_1} \leq \text{TRANS}_{a_2} \quad a_1, a_2 \in A^{ad} \quad (5.3)
\]

\[
\text{IVT}_{a_1} \leq \text{IVT}_{a_2} \quad a_1, a_2 \in A^{ad} \quad (5.4)
\]

Where $\text{TRANS}_i$ is the number of transfers involved with path alternative $i$ and $\text{IVT}_a$ is defined by equation 5.2. Note that each alternative is compared only with alternatives that are in the master-set of the corresponding OD pair. When this dominancy rule is enforced, 30% of the chosen-paths and 52% of the consideration-set paths are excluded. There are two reasons for chosen-paths to be dominated by other alternatives. First, low-frequency direct lines dominate paths that have similar (or even identical) routes but require a transfer. The composition of a background choice-set for the DPCM has to include all the alternatives that are attractive under all possible range of waiting times values (including the case of no waiting time, when the vehicle arrives at the stop). In order to bypass the problem of dominating low-frequency services, it is proposed to disable alternatives that include low-frequency legs from dominating other alternatives. As a result, low-frequency alternatives that are attractive when they are available would still be included in the choice-set as well as high-frequency services that are not dominated by other high-frequency alternatives. The proposed condition can be expressed as:

\[
\frac{\sum_{i \in L_j} 60}{\sum_{i \in L_j} 60/h_i} \leq h_{cr} \quad \forall L_j \in L_a \quad (5.5)
\]
Where \( h_{cr} \) is the threshold headway that distinguishes between frequency-based service and schedule-based service (\( h_{cr} = 15 \) minutes was used in this context). The enforcement of this rule increased the coverage to 72% and 63% of chosen-paths and consideration-set alternatives, respectively.

The second reason for the low share of chosen-paths that fulfill the dominancy rule are marginally longer IVT. Due to information and cognitive limitations individuals cannot distinguish between highly similar alternatives. Hence, the CSGM should not exclude alternatives due to negligible differences. A loose definition of condition 5.4 incorporates a perceptional threshold and can be formulated as:

\[
IVT_{a_1} \cdot (1 + \gamma_{IVT}) \leq IVT_{a_2} \quad a_1, a_2 \in A^{od} 
\]

(5.6)

Where \( \gamma_{IVT} \) is the perceptional threshold parameter that can be interpreted as the minimum ratio of difference required for the individual to identify dominancy. This parameter can represent the mean value of its distribution in the population. When a perception threshold of 0.2 was introduced, more than 94% of chosen-paths were not dominated as well as 89% of the alternatives included in the consideration-set.

**Regression analysis**

Regression analyses were performed in order to investigate the importance of various attributes on the probability to be included in the consideration-set (\( \bar{y}_{1}^{od} \)) and the average rating (\( \bar{r}_{1}^{od} \)). This investigation was performed as a preliminary analysis that may point out which factors should be included in the estimation of the CSGM. Two factors were found to have a consistent and significant explanatory power (\( p < 0.01 \)) for both selection and rating performance: the number of transfers and the ratio between total travel time and the travel time on the shortest alternative included in the master-set of the relevant OD pair, RTTT\(_i\). This ratio is formally expressed as:

\[
RTTT_i = \frac{T_{TT_{i}}}{\min_{a \in A^{od}} T_{TT_{i}}} 
\]

(5.7)

\[
T_{TT_{i}} = \sum_{j=1}^{m_1} (ST_{s_{2:j+1}}^{l_{j}} - ST_{s_{2:j}}^{l_{j}}) + \sum_{j=1}^{m_2} \frac{60}{\sum_{f \in L_j} 60/h_f} 
\]

(5.8)

The first term in Equation 5.8 sums up the scheduled IVT on trip legs, while the second term calculates the total expected waiting time based on the schedule and assuming deterministic headways. In particular, walking times are not incorporated as there is no
detailed data on the first leg from origin to origin stop (the last leg is the same for all alternatives as all the lines have the same stops at the Technion).

The coefficients of the following regression equation are presented in Table 5.7 (along with their t-statistics):

\[ \bar{y}_{i}^{od} = \beta_0^y + \beta_{TRANS}^y \cdot TRANS_i + \beta_{RTTT}^y \cdot RTTT_i \] (5.9)

\[ \bar{r}_{i}^{od} = \beta_0^r + \beta_{TRANS}^r \cdot TRANS_i + \beta_{RTTT}^r \cdot RTTT_i \] (5.10)

The intercepts of the regression equations can be interpreted as the maximum possible value of the dependent variable, which is slightly higher than the absolute maximum values – perfect inclusion probability \( \bar{y}_{i}^{od} = 1 \) and perfect rating \( \bar{r}_{i}^{od} = 5 \). As expected, \( \beta_{TRANS} \) and \( \beta_{RTTT} \) have a negative sign. This implies that an alternative that involves more transfers or has a higher RTTT has a lower probability to be included in the consideration-set and is less attractive. The regression analysis suggests that the number of transfer is the most significant explanatory variable with a large impact on the dependent variables.

Table 5.7: Regression analysis results

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>( \beta_0 )</th>
<th>( \beta_{TRANS} )</th>
<th>( \beta_{RTTT} )</th>
<th>( R^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \bar{y}_{i}^{od} )</td>
<td>1.11 (19.20)</td>
<td>-0.28 (-12.27)</td>
<td>-0.23 (-4.99)</td>
<td>0.40</td>
</tr>
<tr>
<td>( \bar{r}_{i}^{od} )</td>
<td>5.15 (20.65)</td>
<td>-1.42 (-14.83)</td>
<td>-0.68 (-3.38)</td>
<td>0.46</td>
</tr>
</tbody>
</table>

5.2.4 *SP TRANSIT PATH CHOICE*

All the respondents completed the 16 scenarios on the SP transit path choice part of the survey. Figure 5.16 presents the frequency that paths that had a certain attribute were chosen over the alternative path. The direct alternative was preferred over the alternative that required a transfer in 70% of the cases (recall that all scenarios were composed of a direct path versus a single-transfer alternative). The alternative with a shorter IVT was chosen twice more often than the alternative with a longer IVT. The difference is smaller for total waiting times and walking times as also the range of values that were used is smaller (see Table 5.2).
A maximum likelihood estimation of the path utility function is presented in Section 5.4.

5.2.5 SP TRANSFER STOP CHOICE

The final section of the questionnaire asked to rate the importance of five transfer location attributes when choosing between a set of possible transfer locations on a 1 to 5 scale. The average importance grade for each attribute is presented in Figure 5.17, after reversing scales. As expected, the most important criterion for transfer stop choice is that it would have the highest frequency among the possible transfer stops. The second most important criterion is to minimize the walking time involved with the transfer decision, followed by the comfort of waiting conditions at the stop (seats, shelter, toilets, etc.). The importance grades of these three attributes were found to be significantly different from the irrelevancy level at the 0.1% level \((H_0: \bar{X} = 3\) in all cases; \(z(157) = 29.91, p < 0.001; z(157) = 15.41, p < 0.001; z(157) = 6.52, p < 0.001\); for maximum frequency, minimum walking and waiting conditions, respectively). Furthermore, the differences between the three significantly important factors were also found significant (maximum frequency vs. minimum walking: \(t(157) = 5.03, p < 0.001\); minimum walking vs. waiting conditions: \(t(157) = 4.92, p < 0.001\).
In contrast, the respondents were on average indifferent with respect to the location of the transfer stop relatively to their set of possible transfer locations (recalling that 3 indicated indifference). However, an average importance grade can be misleading as it may result from a large variation across the sample population. An analysis of the rating distribution revealed that 17% of the respondents indicated that the location of the stop – either first or last - is highly important in their transfer location decision. These results suggest that the rule of thumb that transfers are concentrated at the merging and demerging points which is used by VIPS (2000) is an unrealistic simplification.

The proposed DPCM accounts explicitly for the two most important transfer stop decision factors: waiting and walking times. The results suggest that the location of the stop relatively to other potential transfer stops is irrelevant for most people. The remaining explanatory factors are site-specific and individual-specific which are currently not accounted for in the model. The importance of site-specific characteristics as the facility layout and perceived comfort and safety was estimated by Wardman et al. (2001), Hensher et al. (2003), Iseki and Taylor (2009) and Guo and Wilson (2010, 2011). These estimations could be incorporated into the utility function by assigning stop-specific transfer penalties.
5.3 TRANSIT CHOICE SET GENERATION MODEL

The choice-set generation model (CSGM) was estimated for the case of deterministic activation and threshold parameters. A total of $\sum_{od} N^o d A^d = 373$ observations were available from the survey. Based on the analysis of selected paths attributes (Section 5.2.3.2), three filtering rules were examined:

(R1) Number of extra transfers
(R2) In-vehicle time (IVT) ratio
(R3) Dominancy rule without dominating low-frequency services

Following the estimation method proposed in Section 4.5.2, the estimation problem has the following non-linear least-squared formulation:

$$\min \sum_{i=1}^{n} \left( y_{in} - z_{ij} \right)^2$$

s.t.

$$z_{ij} = \left[ \sum_{r} \beta_{r} \alpha_{ir} \right] \quad \forall i \in A$$

$$\delta_{\text{trans}_i} = \begin{cases} 1 & \min_{a \in A} |L_i| + \alpha_{\text{ext}}^{\text{trans}} < |L_i| \\ 0 & \text{otherwise} \end{cases}$$

$$\delta_{\text{IVT}_r \text{ratio}_i} = \begin{cases} 1 & \min_{a \in A} \left\{ \sum_{l \in L_i} EIVT_{l_{ij}} \right\} \cdot \alpha_{\text{max}_r \text{ratio}}^{\text{IVT}} < \sum_{l \in L_i} EIVT_{l_{ij}} \\ 0 & \text{otherwise} \end{cases}$$

$$\delta_{\text{not}\_\text{domin}_i} = \begin{cases} 1 & \text{OR} \\ 0 & \exists a_k \in A: \left\{ \left| L_i \right| < |L_k| \land \sum_{l \in L_i} EIVT_{l_{ij}} \leq \sum_{l \in L_k} EIVT_{l_{ij}} \land \frac{60}{\sum_{l \in L_i} 60/n_{ij}} \leq 15 \forall l_i \in L_k \right\} \text{OR} \\ \left\{ \left| L_i \right| \leq |L_k| \land \sum_{l \in L_i} EIVT_{l_{ij}} < \sum_{l \in L_k} EIVT_{l_{ij}} \land \frac{60}{\sum_{l \in L_i} 60/n_{ij}} \leq 15 \forall l_i \in L_k \right\} \text{OR} \\ \left\{ \left| L_i \right| \leq |L_k| \land \sum_{l \in L_i} EIVT_{l_{ij}} \leq \sum_{l \in L_k} EIVT_{l_{ij}} \land \frac{60}{\sum_{l \in L_i} 60/n_{ij}} \leq 15 \forall l_i \in L_k \right\} \text{OR} \end{cases}$$

$$\beta_r \in \{0,1\} \quad \forall r$$

$\alpha_{\text{trans}}^{\text{ext}}$ integer

Where $y_{in}$ is an indicator that takes ‘1’ if individual $i$ includes alternative $l$ in the choice-set and ‘0’ otherwise. $z_{ij}$ indicates whether the alternative fulfills all the filtering rules and hence included in the choice-set generated by the CSGM. $\beta_r$’s and $\delta_r$’s refer to the three filtering rules. $\beta_r$’s are indicators which take ‘1’ if the respective filtering rule $r$ is active and ‘0’ otherwise. $\delta_{ir}$’s indicate whether alternative $i$ fulfills the respective filtering rule $r$. $\alpha_{\text{trans}}^{\text{ext}}$ and $\alpha_{\text{max}_r \text{ratio}}^{\text{IVT}}$ are threshold parameters of the number of extra transfers and the maximum total in-vehicle time ratio. $EIVT_{l_{ij}}$ is the expected in-vehicle time involved with transit line $l_{ij}$ calculated as the difference between the scheduled
times at the relevant stops. \( h^P \) is the planned headway on line \( l \). The critical value for differentiating between low-frequency and high-frequency services was set to 15 minutes \( (h_{cr} = 15) \).

The estimation procedure required the specification of 3 binary activation parameters \((\beta' \text{'s})\) and 2 threshold parameters \((\alpha \text{'s})\) - the number of extra transfers and IVT ratio. The dominancy rule is based on these two dimensions where only alternatives that have a joint headway of less than 15 minutes on each transit leg can dominate.

The estimation was carried out in Excel. Figure 5.18 presents the performance of various solution combinations as a function of the IVT ratio threshold parameter. \( T_0, T_1 \) and \( T_2 \) correspond to the possible values of \( \alpha_{\text{ext}}^{\text{trans}} \). The activation of the dominancy rule is noted by \(-D\) versus \(-N\) in case the dominancy rule is inactive. When \( T_0 (\alpha_{\text{ext}}^{\text{trans}} = 0) \), this distinction is redundant as the two cases yielded the same results.

![Figure 5.18: Objective function value as function of the IVT threshold](image)

The optimal solution is obtained for \( T_2-D \) \((\alpha_{\text{ext}}^{\text{trans}} = 2\) and the dominancy rule is activated\) when \( \alpha_{\text{max,ratio}}^{\text{IVT}} \) is within the range: \( 1.23 \leq \alpha_{\text{max,ratio}}^{\text{IVT}} \leq 1.32 \). The performance of \( T_1-D \) and \( T_2-D \) is very similar. It is evident that the activation of the dominancy rule improved the performance across the range of IVT ratio threshold levels. Interestingly, when the dominancy rule is activated the performance function has a single local minimum. When \( \alpha_{\text{ext}}^{\text{trans}} = 0 \), the dominancy rule plays no role because no
OD pair had more than one direct path alternative. The minimum value remains unchanged for values larger than 1.23 since no direct alternative is excluded due to higher IVT ratios in our dataset.

The formulation of the objective function implies that its value corresponds to the number of false predictions by the estimated solution. The share of correct identifications is hence calculated as (Table 5.8):

$$CI = 1 - \left[ \frac{\sum_{l_{in}}^{\text{assumed}} N_{od}^{A_{od}} \left( y_{jin} - z_{jin} \right)^2}{\sum_{od} N_{od}^{A_{od}}} \right] = TN + TP$$  (5.11)

Where the nominator is the objective function value. False identifications can be sorted into two classes: false positive - the model does not generate an alternative that was included by the respondent - and false negative - the model generates an alternative that was not included by the respondent.

Table 5.8: Matching between reported and generated consideration-set

<table>
<thead>
<tr>
<th>Alternative is included in respondent’s choice-set \ Alternative is included in the generated choice-set</th>
<th>No</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>True negative (TN)</td>
<td>False positive (FP)</td>
</tr>
<tr>
<td>Yes</td>
<td>False negative (FN)</td>
<td>True positive (TP)</td>
</tr>
</tbody>
</table>

In the context of path-set generation, false positive is a higher concern than false negative since the non-inclusion of an alternative implies that it has a zero probability to be chosen later on (Bovy, 2009). Table 5.9 presents the optimal solution for each combination of activation parameter values – the objective function value as well as the optimal threshold parameters’ values.
Table 5.9: Optimal solutions under all possible activation combinations

<table>
<thead>
<tr>
<th>#</th>
<th>$\beta_1$</th>
<th>$\beta_2$</th>
<th>$\beta_3$</th>
<th>Optimal solution</th>
<th>CI (%)</th>
<th>FN (%)</th>
<th>FP (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>176 (,)</td>
<td>53</td>
<td>47</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>146 (,)</td>
<td>61</td>
<td>28</td>
<td>11</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>137 (1,1.3)</td>
<td>63</td>
<td>27</td>
<td>10</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>113 (1,1.3)</td>
<td>70</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>123 (0,)</td>
<td>67</td>
<td>2</td>
<td>31</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>123 (0,)</td>
<td>67</td>
<td>2</td>
<td>31</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>120 (1,1.3)</td>
<td>68</td>
<td>18</td>
<td>14</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>113 (2,1.3)</td>
<td>70</td>
<td>15</td>
<td>15</td>
</tr>
</tbody>
</table>

The first case does not involve any filtering rules. This corresponds to including all the potential paths in the master-set. In this case, there cannot be any false positives and the share of correct identifications corresponds to the share of selected alternatives. When filtering rules are applied, the share of false negative is reduced at the cost of introducing some false positive identification. This is expected as some alternatives are excluded from the choice-set and hence a lower probability to include the alternatives that were not reported by respondents while having a higher probability to exclude alternatives that were considered by respondents. In particular, the activation of R1 without R2 (cases 5 and 6) yields solutions that eliminate all alternatives that imply an extra transfer ($\alpha_{\text{trans}}^{\text{ext}} = 0$). This solution has a very low share of false negative but eliminates reported alternatives that have extra transfers.

The highest share of correct identification (70%) was obtained under combinations 4 and 8 with equal shares of false positive and false negative. Since the highest number of extra transfers in the master-set was 2, the solution obtained in the last case implies that R1 has no impact. Hence, the two optimal solutions are equivalent.

The optimal parameter values indicate that the best fit to the reported choice-sets is achieved by excluding all path alternatives that have a total IVT that is more than 30% higher than the shortest path alternative for their OD. The dominancy rule that its inclusion contributed to the objective function is the strict dominancy definition with a
low-frequency correction, as described above. The two solutions that yielded the best result imply that there is no limitation on the maximum acceptable number of extra transfers. This may be due to the main drawback of our dataset - the majority of the respondents are distributed between only two OD pairs. Moreover, the alternatives given in the proposed master-set did not include more than 2 extra transfers. In case a threshold with a value of 2 or more exists, the data does not allow its estimation.

Verification

In order to assess the CSGM, a simplified version of Haifa’s transit network was given as an input to the initialization phase of BusMezzo. The network contains all the potentially relevant lines to travel from each of the 10 origin zones included in the survey to the Technion (see Figure 5.2). Line frequencies and travel times were given based on the actual timetables.

The two central bus stations are the origins for 62% of the respondents, as shown in Figure 5.7. These are therefore the focus of this analysis. The master-sets generated by the CSGM were compared with the choice-sets reported by respondents in the survey. There are two main performance measures for the reproduction success of the CSGM (e.g. Guo and Wilson, 2010; 2011). First, the coverage rate indicates the inclusion rate of reported paths in the generated set which is calculated as follows:

\[
CR_{od} = \frac{\sum_{a \in A^d} \delta_a^g}{|A^d_p|} = \frac{TP}{TP + FP}
\]  

(5.12)

Where \( A^d_r \) is the collective set of reported alternatives for a given OD pair (all the paths that were given in the master-set and were selected by some respondents) and \( \delta_a^g \) is an indicator that takes ‘1’ if alternative \( a \) is included in the generated choice-set, \( A^d_g \), and ‘0’ otherwise.

The coverage rate is a measure of how effective the CSGM is. The complementary figure is the inclusion rate of generated choice-sets in the reported set of paths, a measure of efficiency.

\[
IR_{od} = \frac{\sum_{a \in A^d_g} \delta_a^r}{|A^d_g|} = \frac{TP}{FN + TP}
\]  

(5.13)

Where \( \delta_a^r \) is an indicator that takes ‘1’ if alternative \( a \) is included in \( A^d_r \) and ‘0’ otherwise. Both measures are reported in Table 5.10. An additional coverage rate was calculated for the chosen alternative in the RP section.
Table 5.10: Performance measures of the CSGM based on survey data

<table>
<thead>
<tr>
<th>Origin travel zone</th>
<th>Coverage rate of the selected CS</th>
<th>Inclusion rate</th>
<th>Coverage rate of the chosen path</th>
<th>Number of generated paths*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bay CBS</td>
<td>100.0%</td>
<td>100.0%</td>
<td>100.0%**</td>
<td>18 (5)</td>
</tr>
<tr>
<td>Coastline CBS</td>
<td>90.2%</td>
<td>100.0%</td>
<td>97.2%</td>
<td>15 (3)</td>
</tr>
<tr>
<td>Other origins</td>
<td>89.6%</td>
<td>100.0%</td>
<td>97.2%</td>
<td>64 (15)</td>
</tr>
</tbody>
</table>

*The number of generated paths corresponds to the number of alternative line combinations with the number of obtained merged path alternatives given in brackets.
**Few chosen-paths that involved long walking distances were not included in the consideration-set section. When these paths are removed, the coverage rate is 100% for the chosen paths.

There is a perfect match between the reported and the generated choice-sets for the Bay CBS origin zone – all reported paths are reproduced by the model. Moreover, the model does not generate any path that is not reported by some respondent (FN=FP=0). Other origin zones also have high coverage rates. The CSGM did not generate any path that was not selected to the choice-set by any of the respondents for any of the OD pairs. It should be noted that in general a high inclusion rate can indicate a too extensive generation model that generates paths that have negligible shares. However, this result should be interpreted in the context of a relatively small and simplified network, where the number of paths in the universal-set is small. Nevertheless, generated choice-sets are still larger than reported choice-sets since the generated choice-set is a collective choice-set for a group of individuals travelling between an OD pair.

The coverage rate for the RP alternative is very high for all origin zones. An investigation into the uncovered paths revealed that they either included long walking distances that exceeds the maximal threshold (0.5 km). It can be concluded that the CSGM reproduced the vast majority of path alternatives that were reported by travelers.

The inclusion rate cannot be directly assessed in the model since it does not incorporate probabilistic choice sets. The deterministic master-set is used throughout the dynamic path choice model (DPCM). BusMezzo was used to simulate traveler
decisions with a demand that corresponds to the share of respondents for each OD pair. The probability to board an arriving transit vehicle can be interpreted as the share of individuals that will take a path alternative when it becomes available. This in turn could be considered the dynamic equivalent of whether an individual considers a path alternative to be a plausible and attractive and hence included in the choice-set. Hence, the probability to board – the share of simulated travelers that board an alternative when it became available – could be compared to the inclusion rate. Even though these notions are not identical, they are expected to be strongly correlated.

Figure 5.19 presents the simulated probability to board versus the inclusion rate. It includes only path alternatives for the Bay CBS or the Coastline CBS, due to sample size concerns. The dots are scattered close to the 45 degrees line, indicating a very good match between the two measures. The weighted correlation, after accounting for the number of individuals associated with each OD pairs, is \( r = 0.973 \). Figure 5.20 represents the comparison for the average rating of path alternatives, which also has a very high fitting with \( r = 0.962 \).

Figure 5.19: Comparison of inclusion rate vs. simulated probability to board
Travelers assignment results cannot be compared with the reported chosen-paths because the relevant network includes low-frequency lines and hence random arrival assumptions are invalid.

5.4 PATH UTILITY FUNCTION ESTIMATION
The data collected on the transit path choice section of the survey (Part D in Figure 5.1) refers to attributes of path alternatives in the context of a boarding decision. Respondents were asked to choose between a path alternative that is associated with an arriving bus and a path alternative that implies to stay and wait at the stop. This representation intends to imitate the dynamic conditions of a dynamic path choice process. Alighting and connection path decisions were not considered. It is presumed that the underlying decision factors and their relative importance remain the same along the trip and therefore the path utility function takes the same form over different trip stages. Hence, the estimated values can be used for specifying the path utility function coefficients in BusMezzo. The current implementation assumes highly adaptive decision makers. A more comprehensive estimation of the DPSM would enable to determine the potential interdependence between successive decisions and the level of adaptation exercised by travelers.

The database has a total of 2524 records (16 questions * 158 respondents with 4 missing records) that are available for estimating the path utility function. Each record
contains the attributes of the sampled choice scenario from the dataset of 972 alternative scenarios. The estimation results of the best MNL specification are presented in the following section. All maximum likelihood estimations were performed in Biogeme software (Bierlaire, 2003). Then, the estimated values are discussed and compared with previous studies (Section 5.4.2).

5.4.1 MODEL SPECIFICATION AND ESTIMATION RESULTS
Numerous model specifications for the path utility function were tested. These alternative specifications included both path and individual attributes. None of the individual attributes such as age, income and travel habits was found to be a significant explanatory variable. The specification that yielded the best estimation results is presented in this section. It incorporates all the travel time components and the number of transfers. These are the primary variables that determine the path choice as they are the most important quality of service measures. In addition, secondary quality of service measures as comfort and reliability are also included. This model provides a useful specification that contains the important trade-offs between time components, number of transfers, seat availability and service reliability. The travel-time and transfer attributes are included in the current implementation of the DPCM in BusMezzo, as was presented in Section 4.7.2. The remaining path decision factors can be incorporated into the DPCM in the future by assigning their respective anticipated values. This can be done for example by analyzing a large number of simulation runs and deriving the expected values of these variables through day-to-day learning.

The estimated coefficients and their respective t-statistics are presented in Table 5.11. The specified model is linear in the parameters with a MNL model structure. Since each respondent had 16 observations there are potentially individual-effects that should be accounted for in the estimation process. Hence, individual effects were specified for each of the variables to account for the panel dataset. Individual effects were found significant for five of the seven path attributes. Furthermore, respondents might be biased by inertia. In order to test the possibility that inertia biased respondents' choices, a model with an additional explanatory variable which takes the previous choice (in terms of alternative's label) was estimated. The inertia coefficient was found to be irrelevant as its estimated coefficient was zero with t-stat of zero. Hence, respondents' choice cannot be explained by the label attached to their previous choices.
Table 5.11: Estimated coefficient values for the simplified MNL model

<table>
<thead>
<tr>
<th>Variable</th>
<th>Units</th>
<th>Estimated values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transfers</td>
<td>Number of transfers</td>
<td>-0.371 (-2.76)</td>
</tr>
<tr>
<td>In-vehicle time</td>
<td>Minutes</td>
<td>-0.0497 (-11.89)</td>
</tr>
<tr>
<td>Waiting time</td>
<td>Minutes</td>
<td>-0.0930 (-10.07)</td>
</tr>
<tr>
<td>Egress time</td>
<td>Minutes</td>
<td>-0.0907 (-10.08)</td>
</tr>
<tr>
<td>Available seat</td>
<td>Binary indicator</td>
<td>0.539 (4.70)</td>
</tr>
<tr>
<td>Schedule adherence</td>
<td>Percentage X 100</td>
<td>0.0528 (10.31)</td>
</tr>
<tr>
<td>Denied boarding</td>
<td>Percentage X 100</td>
<td>-0.0345 (-5.50)</td>
</tr>
<tr>
<td>Standard deviation of IVT</td>
<td></td>
<td>0.0277 (6.45)</td>
</tr>
<tr>
<td>Standard deviation of waiting time</td>
<td></td>
<td>0.0280 (3.29)</td>
</tr>
<tr>
<td>Standard deviation of walking time</td>
<td></td>
<td>0.0660 (7.51)</td>
</tr>
<tr>
<td>Standard deviation of available seat</td>
<td></td>
<td>0.618 (3.95)</td>
</tr>
<tr>
<td>Standard deviation of denied boarding</td>
<td></td>
<td>0.0313 (7.13)</td>
</tr>
<tr>
<td>Observations</td>
<td></td>
<td>2524</td>
</tr>
<tr>
<td>Parameters</td>
<td></td>
<td>12</td>
</tr>
<tr>
<td>Null LL</td>
<td></td>
<td>-1749.50</td>
</tr>
<tr>
<td>LL at convergence</td>
<td></td>
<td>-1249.42</td>
</tr>
<tr>
<td>Rho-square</td>
<td></td>
<td>0.286</td>
</tr>
<tr>
<td>Adjusted rho-square</td>
<td></td>
<td>0.279</td>
</tr>
</tbody>
</table>
All the variables that were included in the choice experiment were found to have a significant explanatory power. All of the coefficients have the expected sign. All coefficient values are significant at the \( p < 0.005 \) level. The inclusion of individual effects resulted in a higher explanatory power based on adjusted rho-square and the Log-Likelihood ratio test with \( p < 0.001^3 \).

As can be expected, the disutility varies between different time components. Waiting time is associated with the highest inconvenience followed closely by walking time. A minute of waiting and walking is equivalent to 1.87 and 1.82 minutes of IVT, respectively. In addition, transfer penalty is equivalent to 7.5 minutes of IVT or 4 minutes of waiting or walking.

The comfort coefficient suggests that having an available seat is equivalent to approximately 12.5 minutes of IVT. It may be argued that the comfort effect would be better captured when accounting for the respective on-board time. However, an alternative model specification which defined the comfort variable as the result of seat indicator times the relevant IVT was found inferior. It may be the result of implicit assumptions on the expected standing time before a seat would become available. In addition, a decrease of 10% in the probability that the next bus will arrive within a 5 minutes window from the expected arrival time is associated with a disutility equivalent to 5.67 minutes of waiting time. This may reflect that the disutility of lower schedule adherence is made up of not only the additional waiting time but also the discomfort involved with uncertainty. This may also be the case with denied boarding. An increase of 10% in the probability of denied boarding is associated with the same discomfort as 3.7 waiting minutes. The additional expected waiting time due to denied boarding varies among networks depending on service frequencies.

\[ ^3 (-2\left[L(\hat{\beta}_R) - L(\hat{\beta}_U)\right]) = -2[-1309.20 + 1249.42] = 119.56 > \chi^2_{0.9999,8} = 26.12 \]
5.4.2 DISCUSSION
All the factors included in the experimental design were found to have a significant explanatory power. All the coefficients have the expected signs. The following compares the estimated values with previous findings reported in the literature.

Travel time components
A large number of studies have estimated the relative importance of travel time components based on SP and RP data. Wardman (2004) conducted a meta-analysis of public transport values of time studies. The meta-analysis concludes that the values of both walking and waiting time are less than twice the value of IVT and found no clear relationship between the two based on a large set of studies. The overall mean ratio was found to be 1.76 for waiting time and 1.68 for walking time. The corresponding estimated values are 1.87 and 1.82 which are within the range of values reported in the literature. Nuzzolo et al. (2011) reported a Logit model that was estimated for a schedule-based model that had been calibrated based on data from the city of Naples, Italy. The values of the calibrated parameters resemble closely the estimated values with a ratio of 1.88 between waiting time and IVT. The calibration of the transit-related coefficients by Wahba and Shalaby (2009b) for a dynamic transit model yielded substantially different results from the common values reported in the literature.

Wardman (2004) also found that the relative importance of walking and waiting time varies by transit mode, journey distance and to a lesser extent also by passenger type (e.g. commuter, business, leisure). However, the segmentation of travel time components by traveler groups did not yield an improvement in our estimation results.

Transfers
According to the estimation results, transfer penalty is equivalent to 7.5 IVT or 4 minutes of waiting or walking. Wardman (2001) found that the average transfer penalty in studies of the British railways is equivalent to 17.6 IVT. However, he mentioned the low frequency and the design of the experiment as potential reasons for the high estimation. Moreover, urban transit users are expected to be less reluctant to transfer. Within the context of a high-frequency urban system, Iseki and Taylor (2009) found in their review a large range of transfer penalty values. The estimated values varied
between modes and transfer locations, with transfers between buses having a higher transfer penalty than transfers in the rail system. Guo and Wilson (2004) conducted a RP survey in Boston’s metro system and estimated transfer penalty to be equivalent to be in the range of 4.3-15.2 minutes of walking time for various station types. The value estimated by the survey is at the lower end of this range. However, their estimation is based on an on-board survey and is therefore limited to transit users. In a later path choice model analysis, Guo and Wilson (2010) found transfer penalty to be equivalent to 4.2 IVT on an average London underground station. VIPS (2000) uses a transfer penalty in the range of 5 to 10 IVT.

Clearly, there is a large range of values in the literature with respect to transfer penalty. Note that the survey referred to transferring between buses and that respondents were provided with expected waiting time for the next vehicle which plausibly reduced the uncertainty involved with the transfer alternative. In addition, the survey considered only two levels of transfer – none or single. The results of Horowitz (1981) suggest that this may still be generalized as an increasing number of transfers have a linear effect on the utility function.

**Comfort**

Comfort conditions were recently acknowledged as an important decision factor. Horowitz and Zlosel (1981) already found that passengers are not so sensitive to the crowding level on-board as long as there is an available seat. Our results that having an available seat is a significantly important factor in transit path choice in the context of urban trips is in agreement with the results of Hensher et al. (2003) and Nuzzolo et al. (2011). The estimated comfort coefficient is equivalent to 12.5 IVT, the same magnitude as the 14 IVT reported by Nuzzolo et al. (2011). The disutility caused by overcrowded conditions was not estimated in this model. It can be assumed that seat availability is known only for approaching vehicles when making a boarding decision or in case of RTI. However, the effect of comfort needs further investigation to evaluate its impact for the entire range of crowding conditions. Schmöcker et al. (2011) reported the value of 2.7 times the actual IVT in the case of overcrowding for the London underground.
**Schedule adherence and denied boarding**

These two aspects of service reliability were seldom studied in the context of path choice models. Note that unlike the impediment of late or early arrivals that is analyzed in the context of trip departure choice, the schedule adherence coefficient estimated in this model refers to the risk associated with a certain path alternative. Hensher and Prioni (2002) estimated that each minute of bus delay is equivalent to 1.82 IVT, while our estimations suggest that 10% increase in the probability of being late more than 5 minutes induced a disutility of 5.67 IVT. As is also the case for denied boarding, it is plausible that the values of these coefficients are dependent on the local transit network characteristics. For example, passengers travelling in a highly dense and frequent network may be less sensitive to the probability of a late vehicle or not being able to board the next arriving vehicle as they know that it will not take long before another vehicle will arrive. Hence, it is not clear whether these values can be generalized and their incorporation in the DPCM is not straightforward.

In conclusion, a review of previous value of time studies and SP surveys of transit path choice factors suggest that estimated values from survey results are consistent with previous findings. The estimated path utility function values were given as input to BusMezzo for all the path decisions in the transit assignment case studies described in Chapter 7.
6. EVALUATION OF HOLDING CONTROL STRATEGIES

Service reliability is one of the main objectives of transit operators. In the context of high-frequency urban services, unreliable service results in long waiting times, bunched vehicles, long delays, uneven passenger loads and poor capacity efficiency. In addition, having a more reliable transit performance can also imply lower operations costs and more efficient crew management. This chapter analyzes and evaluates various control strategies designed to improve transit performance. The purpose and design issues involved with holding control strategies are first outlined along with a review of previous studies. Section 6.2 presents how the transit operations model was applied on a high-frequency bus line in Tel-Aviv, Israel. This case study included an analysis of the transit performance and model validation. In addition, the affects of holding control strategies on the performance of this line were assessed. A more comprehensive multi-perspective evaluation was performed for a trunk line in Stockholm, Sweden. Detailed empirical analysis of AVL data was followed by a simulation study of various holding strategies. The promising results of this assessment led the regional transit authority, SL, to pursue the implementation of the proposed holding strategy in a field trial. The practical considerations of such an experiment are discussed (Section 6.3). A method for optimizing the number and location of time point stops is presented in Section 6.4. The solution methods use BusMezzo as an evaluation tool and were applied on the Stockholm line. The simulation model was also used for the analysis of the interaction between holding strategies and different boarding regimes on a common transit corridor (Section 6.5). This line of research demonstrates the capabilities of BusMezzo as an evaluation tool of transit operations, in particular in the context of APTS.

6.1 BACKGROUND AND RELATED STUDIES
Transit operations involve several inherent sources of uncertainty including dispatching time from the origin terminal, travel time between stops and dwell time at stops. These stochastic factors are interrelated through the relation between the number of waiting passengers, headway between consecutive buses and dwell time as well as the propagation of delays through trip chaining. These interrelations may cause
random variations to propagate so that a late vehicle will have to pick up more passengers causing it to be further late while the succeeding vehicle increasingly catches up. This process leads to the well-known bunching phenomenon.

The bunching phenomenon is expected to escalate along the route, a notion supported by previous studies of AVL bus data (El-Geneidy et al., 2011; Ruan and Lin, 2008; Byon et al., 2011). Chen et al. (2009) analyzed the performance of a large sample of bus routes in Beijing and related service characteristics. They found that both total route length and the distance from the origin terminal are negatively correlated with measures of schedule adherence and service regularity as unreliability propagates along the line. Bellei and Gkoumas (2010) analyzed the headway pattern along a bus route, based on a stochastic simulation model. They found an increasing negative correlation between consecutive headways. The headway distribution became sparser towards the end of the route having a bimodal form concentrated around very short and very long headways. As can be expected, the magnitude of this pattern increased with shorter planned headways, higher demand levels and higher travel time variability. Therefore, operators have to take a proactive approach and control the service to prevent the otherwise inevitable irregularity reinforcing feedback.

Transit control strategies consist of a wide variety of operational methods aimed to improve transit performance and level of service. Holding strategies are among the most widely used transit control methods aimed to improve service regularity by regulating departure time from stops according to pre-defined criteria (Abkowitz and Lepofsky, 1990). The strategy contains a set of rules that determine at which stops along the route departure times will be subject to regulation and which criteria are used for determining the departure time. These stops are known as time point stops (TPS).

There are three main decisions involved in implementing holding strategies: the number of TPS, their location along the route and the holding criteria. Although hypothetically all stops might be defined as TPS, a typical bus line has only a few TPS (such as main transfer and CBD locations). Several studies (Turnquist and Blume, 1980; Abkowitz and Engelstein, 1984; Wirasinghe and Liu, 1995; Liu and Wirasinghe, 2001) suggested that TPS should be located at the beginning of a sequence of high-demand stops. In addition, in order to minimize delays caused by holding passengers on-board, stops characterized by high levels of through passengers (passengers staying on board) should be avoided when considering TPS layout (Hickman, 2001). Previous studies that
have formulated a deterministic analytical model reached contradictory results. Some of these studies concluded that only the original terminal should be defined as a TPS (Eberlein et al., 2001), while others suggested that multiple holding locations are beneficial for minimizing total passenger-time (Sun and Hickman, 2008). Senevirante (1990) found that the relation between the standard deviation of the headway and the number of TPS can be formulated as a second degree polynomial. Hence, it was concluded that beyond a certain number of TPS, which depends on the specific line characteristics, the marginal contribution of an additional time point turns negative.

While holding strategies are aimed at reducing passenger waiting times they may also introduce longer travel times for passengers on-board. Several previous studies took this into account by formulating a single compensatory objective function or by assessing multi-criteria analysis (Abkowitz and Engelstein, 1984; Wirasinghe and Liu, 1995; Rossetti and Turitto, 1998; Cortes et al., 2010). A comparison between holding control strategies that rely only on local information to APTS-based strategies was conducted by Dessouky et al. (2003). They studied holding strategies aimed at improving schedule coordination at transfer hubs and concluded that the best holding control strategy was the global optimized strategy, which is also the most demanding in terms of data and technology requirements. This strategy requires forecasting the arrival times of connecting buses, number of transferring passengers and expected number of boarding passengers at downstream stops.

Holding strategies differ in their criteria for departure time from a TPS and the information that their implementation requires. They are classified according to whether they are based on a schedule or a headway criterion. Schedule-based holding enforces buses that arrive early at a TPS to wait at the stop until a pre-defined slack from their scheduled departure time. The equivalent mathematical definition is:

\[
ET_{s,l}^k = \max (SET_{s,l}^k - s_{s,l}, AT_{s,l}^k + DT_{s,l}^k)
\] (6.1)

Where:
- \(ET_{s,l}^k\) - departure (exit) time for line \(l\) on trip \(k\) from stop \(s\)
- \(SET_{s,l}^k\) - scheduled departure (exit) time
- \(s_{s,l}\) - non-negative slack size
- \(AT_{s,l}^k\) - actual arrival time
- \(DT_{s,l}^k\) - dwell time
A simulation-based study tested the effect of different slack sizes on the generalized passenger travel time and found the optimal slack size to be zero (Vandebona and Richardson, 1986). The same result was obtained by Liu and Wirasinghe (2001) who incorporated an optimization process into a simulation model.

Headway-based holding strategies imply that if the current headways between the transit vehicle and the preceding and/or subsequent vehicles do not fulfill a minimal headway requirement, then the transit vehicle has to wait at the stop until the requirement is met. In case the headway requirement takes into account only the headway from the preceding bus, $ET_{s,l}^k$ can be expressed as follows:

$$ET_{s,l}^k = \max (AT_{s,l}^{k-1} + \alpha H_{s,l}^k, AT_{s,l}^k + DT_{s,l}^k)$$  \hspace{1cm} (6.2)

Where:

$H_{s,l}^k$ - planned headway between trips $k - 1$ and $k$ on line $i$

$\alpha$ - threshold ratio parameter

The threshold ratio specifies the minimal allowed headway relative to the planned headway. Headway-control strategies are intended for short headways, when maintaining even headways reduces passengers' waiting times. Schedule-control strategies are more likely to be useful if the frequency is such that passengers tend to follow the timetable (Strathman et al., 1999). Both Turnquist and Blume (1980), who searched analytically for the optimal threshold criteria, and Fu and Yang (2001), who studied it using a simulation model, concluded that the optimal threshold level is in the range of 0.6 to 0.8 times the planned headway.

Headway-based strategies are not limited to the consideration of the headway from the preceding bus. It is also possible to incorporate the headway to the next bus into the holding strategy. In order to have even headways on both sides, the holding criterion is based on the mean headway:

$$ET_{s,l}^k = \max \left( AT_{s,l}^{k-1} + \frac{(AT_{s,l}^k - AT_{s,l}^{k-1}) + (AT_{m,l}^{k+1} + SRT_{m,s} - AT_{s,l}^k)}{2}, AT_{s,l}^k + DT_{s,l}^k \right) = \max \left( AT_{s,l}^{k-1} + \frac{AT_{m,l}^{k+1} + SRT_{m,s} - AT_{s,l}^k}{2}, AT_{s,l}^k + DT_{s,l}^k \right)$$  \hspace{1cm} (6.3)

Where:

$m$ – index of the last stop that was visited by bus trip $k - 1$

$SRT_{m,s}$ - scheduled riding time between stops $m$ and $s$
The fraction in the first expression is the average between the headway from the preceding bus and the headway from the succeeding one. Applying this control strategy at origin terminals in a simulation model of urban rail operations was found promising (Koutsopoulos and Wang, 2007). Instantaneous AVL data allow implementing this strategy also at intermediate stops along the route. Note that this holding strategy is independent of the planned headway. Nevertheless, Daganzo (2009) analyzed the performance of a similar adaptive control strategy based on bus-to-bus communication and concluded that the deviation from the schedule and the deviations of the headway are small and bounded under realistic assumptions.

In addition, it is possible to combine the two headway-based strategies (6.2) and (6.3) in order to restrict the maximum allowable holding time:

\[ ET_{s,l}^k = \max \left( \min \left( \frac{AT_{s,l}^k - 1 + AT_{m,l}^{k+1} + SRT_{m,s} - AT_{s,l}^k}{2}, AT_{s,l}^{k-1}, \alpha h_{s,l}^k \right), AT_{s,l}^k + DT_{s,l}^k \right) \]  

All the above holding control strategies were implemented in the simulation model of BusMezzo.

6.2 CASE STUDY I: BUS LINE 51, TEL-AVIV

6.2.1 EXPERIMENT DESCRIPTION

The transit simulator was applied for the evaluation of the operations of line 51 in the Tel Aviv metropolitan area. The simulation model was first used for analyzing the transit conditions under the current no control conditions. Second, the simulation assessed transit performance under various candidate holding strategies. Line 51 is a high demand urban line which connects a dense satellite residential city to the central business district (Figure 6.1). Its 14 kilometer long route runs on a heavily congested urban arterial. In fact, the route of line 51 constitutes a substantial part of the first light rail train (LRT) line in Tel-Aviv metropolitan area which is currently under construction. The bus line includes 30 stops on the inbound direction and 33 on the outbound direction. The line route and demand profiles for the inbound and outbound directions are shown in Figure 6.2. The scheduled headway during the peak period is 6-

\footnote{The content of this section was included in the following publications: Cats O., Burghout W., Toledo T. and Koutsopoulos H.N. (2010). Mesoscopic modeling of bus public transportation. Transportation Research Record, 2188, 9-18; Cats O., Burghout W., Toledo T. and Koutsopoulos H.N. (2010). Evaluation of real-time holding strategies for improved bus service reliability. Proceedings of 13th International IEEE Annual Conference on Intelligent Transportation Systems, Madeira, Portugal.}
8 minutes and the average running time is 49 minutes inbound and 41 minutes outbound. Currently, there is no service control in place along this line. In the case study, an implicit representation of traffic conditions was adopted (see Section 3.2.2) with running times between stops assumed to follow lognormal distributions, with means equal to the scheduled times.

Figure 6.1: The route of bus line 51 in Tel-Aviv metropolitan area (arrows point the location of proposed time point stops)
Figure 6.2: Schematic route and load profile for inbound (up) and outbound (down) directions of line 51
At both trip ends, recovery times were calculated based on the 85th percentile of the trip travel times, calculated according to the lognormal distribution (TCRP, 2000). These recovery times were then used as minimum requirements when determining the fleet assignment. Layover times are already integrated into the scheduled times.

6.2.2 Analysis of the Current Conditions and a Validation
The outputs of the simulation were tested against two sets of real-world data. First, video traffic records were available from two bus stops – Stop 28 in the inbound direction and stop 4 in the outbound direction for the period 06:30-08:30. Figure 6.3 shows the observed and simulated headway distributions in these two stops. Two-sample Kolmogorov-Smirnov tests were conducted in order to compare the distributions of the observed and simulated headways at these two stops. The hypothesis that the observed and simulated headways are derived from the same distribution cannot be rejected (D=0.204 and D=0.253 compared with $D_{7,0.05} = 0.486$ and $D_{6,0.05} = 0.521$, respectively).

For the second part of the validation, a dataset of observed running times between intermediate stops along the bus line during the AM peak period was compared with simulated running times. The observed dataset contains bus arrival times for stops 13 through 27 on the inbound route. Figure 6.4 (upper image) presents the trajectories according to observed and simulated data in the section covered by the data. It is evident that the simulated trajectory replicates the observed trajectory closely. It is important to note that both simulated and observed running times incorporate dwell times at stops.

The bottom image on Figure 6.4 shows the upper and lower bounds of the 95% confidence interval of the means of simulated and observed running times. The simulated and the observed intervals overlap continuously along the presented trajectory. The hypothesis that the simulated and observed running times are drawn from the same distribution cannot be rejected at the 95% level for any of the stops. In addition, the simulated and observed overall running times between stop 13 and stop 27 were compared. The hypothesis that the observed and simulated running times are derived from the same distribution cannot be rejected based on the Kolmogorov-Smirnov test (D=0.384 compared with $D_{9,0.05}=0.432$).
Figure 6.3: Headway distribution at stop 28 on the inbound direction (up) and at stop 4 on the outbound route (down)
Figure 6.4: A partial trajectory of inbound route – its mean (up) and the upper and lower bounds of 95% confidence interval of the mean (down)

Note that vehicle trajectories are the result of the interaction between running times between stops, headways, passenger arrival and dwell times – the positive feedback loop that drives the bunching phenomenon. In addition, the explicit modeling of trip chaining may replicate cases of late departures from the origin stop. This validation is a first attempt to test the simulation against real-data. It is clearly insufficient and has to be followed up by more comprehensive calibration and validation efforts.
In addition, the assumption that passenger arrival processes follow the Poisson distribution was tested using boarding counts from stops 13 through 27 on the inbound route and stops 4 through 19 on the outbound route. Based on the Kolmogorov-Smirnov test, the hypothesis that passenger arrivals at stops follow the Poisson distribution cannot be rejected for all stops with the exception of stop 21 on the inbound direction. This stop is characterized by low-frequency events of large numbers of boarding passengers. This is probably due to passengers transferring from the nearby train station.

6.2.3 Scenario design
The four holding strategies defined by expressions (6.1) to (6.4) in addition to the base case scenario with no control strategy were studied. The number and location of TPS, as well as the slack size or threshold headway ratio, are determined according to common values and methods in the literature (Abkowitz and Engelstein, 1984; Liu and Wirasinghe, 2001). In order to maintain comparable settings, the same number and location of TPS were used across all strategies (shown in Figure 6.1). The rule-of-thumb that TPS should be located at the beginning of a sequence of high demand stops was applied on line 51. Based on the boarding profiles, three TPS (7, 13 and 21) were defined on the inbound route and two on the outbound route (8 and 20). While dispatching from the origin terminal is determined by the schedule and vehicle availability, the second stop on each direction was defined as a TPS in order to regulate the service from the beginning of the trip.

After the number and location of TPS was set, the holding criteria parameter values had to be defined. Schedule-based holding was simulated with a slack size of zero \( (s_{ij} = 0 \, \forall i, j) \), which implies that buses do not depart from TPS before their scheduled time. In order to examine the sensitivity of the results to the threshold ratio parameter value, the holding strategy defined by (6.2) was simulated with three levels of threshold ratio: 0.6, 0.7 and 0.8, following recommended values from previous studies (Turnquist and Blume, 1980; Fu and Yang, 2002). In summary, seven scenarios were evaluated, as follows:

- No control (strategy 0)
- Schedule-based control (strategy 1, equation 6.1)
- Headway-based control based on the preceding bus (strategy 2, equation 6.2 with three values of $\alpha$)
- Headway-based control based on the mean of the headway from the preceding and the expected headway to the next bus (strategy 3, equation 6.3)
- Combination of strategies 2 and 3 (strategy 4, equation 6.4 with $\alpha = 1.0$)

The simulations were conducted for the peak period between 6:30am and 9:30am with flat peak hour passenger demand and headway of 6 minutes for the inbound direction and 7 minutes for the outbound direction. Since BusMezzo includes several interrelated stochastic components – passenger arrival and alighting processes; dwell time; departure time from origin terminal; travel time and; recovery time – it is essential to conduct multiple replications for output analysis. The standard deviation of the headway is an important service measure that is the outcome of a complex interaction between all random processes in the system. Given this output measure, the number of required repetitions can be calculated using the following formula (Dowling et al., 2004):

$$N(m) = \left( \frac{S(m) \cdot t_{m-1, (1-\alpha)/2}}{\bar{X}(m) \cdot \varepsilon} \right)^2$$

(6.5)

Where:

- $N(m)$ - number of replications required given $m$ initial simulation runs
- $\bar{X}(m)$ – estimated mean based on a sample of $m$ simulation runs
- $S(m)$ - estimated standard deviation based on a sample of $m$ simulation runs
- $\varepsilon$ - allowable percentage of error of the estimate $\bar{X}(m)$ of $\mu$
- $\alpha$ - level of significance

Given $\varepsilon = 0.05$ and $\alpha = 0.05$, then $N(60) = 47.22$ at the worst case, indicating that the initial 60 replications are sufficient for the validation. It should be noted that different applications or output measures may require different number of repetitions depending on the desired level of accuracy. Hence, 60 three-hour simulation runs were conducted for each scenario. The total execution time for the 60 runs was about 10 seconds on a standard PC.

6.2.4 ASSESSMENT OF HOLDING STRATEGIES
The detailed representation of transit operations in the simulation model allows analyzing the system performance ranging from the level of a single trip or a specific
stop to overall system performance measures. Figure 6.5 presents the coefficient of variation of the headway along the inbound direction for the peak hour (7:30am-8:30am). The TPS are marked with vertical dashed lines. As the distance from the origin terminal grows, the variability accumulates and causes the propagation of bus service unreliability. It is evident that under headway-based holding strategies there is an immediate decrease in the coefficient of variation of the headway after each TPS, preventing the continuous propagation of headway variability. Moreover, mean-headway strategies (s=3 and s=4) are the most efficient in reducing headway coefficient of variation. The second TPS (stop 7) is especially useful as it corrects irregular dispatching from the origin terminal, when the information from the next bus becomes available. For the preceding-headway strategy (s=2), $\alpha = 0.7$ yields the highest reduction. As the schedule-based strategy is designed to improve schedule adherence objectives, it is not surprising that it is inefficient when it comes to headway regulation.

Figure 6.5: Coefficient of variation of the headway under various holding strategies (inbound direction)
Figure 6.6 presents a time-space diagram showing example trajectories of a sequence of three bus vehicles, displaying a pair of chained trips (one outbound followed by one inbound trip). Bus trajectories under no control ($s=0$) and mean headway-based strategy ($s=3$) are presented. In each trajectory, sloped segments represent running time between stops and vertical segments correspond to dwell times and holding times at TPS. Recovery times between trips at both terminals are also apparent in the figure. This diagram provides a convenient tool for analyzing operation strategy dynamics. As can be clearly observed, buses are getting bunched systematically when there is no control strategy, while the applied holding strategy is very effective in regulating the service by creating even headways at TPS. For example, on the first trip, bus 2 is getting increasingly bunched with bus 1, with TPS holding bus 2 and preventing continuous bunching. When there is no control strategy, the two buses depart closely, get bunched and even overtake each other along the route. If a schedule-based dispatching is not an operational constraint (e.g. transfer hub, labor agreements), then fully headway-based operations could be implemented in order to reduce the initial headway variability (Koutsopoulos and Wang, 2007).
At the system level, several measures of performance were calculated for each scenario. Table 6.1 summarizes these measures for the inbound direction. While there is no obvious relationship between schedule adherence and service headway, the probability to adhere to the planned headway decreases for shorter headways (Chen et al., 2009). Service regularity is evaluated by measuring headway variability, as for high-frequency services the main performance objective is to regulate headways and avoid bunching of consecutive buses. Headway coefficient of variation was calculated for each stop along the route. The reported statistics are the mean values across all stops in each direction. Note that the average headway during the entire simulated peak period is constant as the number of dispatched buses does not change. Headway-based holding strategies reduced headway variability, resulting in shorter passenger waiting times. However, they also imply an increase in the average total travel time which may impose higher operational costs. The coefficient of variation decreases up to 20% under the preceding-headway strategies and more than 45% when the mean-headway strategy was restricted by the planned headway (s=4) was implemented. The mean-headway
strategy (s=3) resulted in service regularity gains at the cost of an increase in the average travel time. However, imposing a limit on holding time for this information demanding mean-headway strategy (s=4), achieved the highest improvement in service reliability, with a moderate increase in travel time that is not higher than the preceding-headway strategy (s=2, $\alpha = 0.8$).

Table 6.1: Service measures of performance for inbound direction under various scenarios

<table>
<thead>
<tr>
<th>Scenario</th>
<th>$\alpha$ (CV(h))</th>
<th>Average waiting time per pass. (sec)</th>
<th>Average increase in total travel time (sec)</th>
<th>Early arrivals (%)</th>
<th>Late arrivals (%)</th>
<th>On-time arrivals (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>---</td>
<td>0.34</td>
<td>200.43</td>
<td>0.00</td>
<td>0.44</td>
<td>29.44</td>
</tr>
<tr>
<td>1</td>
<td>---</td>
<td>0.34</td>
<td>200.29</td>
<td>26.63</td>
<td>0.38</td>
<td>27.58</td>
</tr>
<tr>
<td>2</td>
<td>0.6</td>
<td>0.30</td>
<td>195.37</td>
<td>12.09</td>
<td>0.31</td>
<td>31.07</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>0.28</td>
<td>194.12</td>
<td>25.85</td>
<td>0.23</td>
<td>32.90</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>0.27</td>
<td>192.72</td>
<td>46.82</td>
<td>0.23</td>
<td>40.15</td>
</tr>
<tr>
<td>3</td>
<td>---</td>
<td>0.23</td>
<td>188.97</td>
<td>79.70</td>
<td>0.22</td>
<td>41.36</td>
</tr>
<tr>
<td>4</td>
<td>1.0</td>
<td>0.18</td>
<td>185.14</td>
<td>44.78</td>
<td>0.24</td>
<td>37.95</td>
</tr>
</tbody>
</table>

Another important measure of service reliability is on-time performance. For a scheduled headway of 6 minutes, a bus is considered in this study to adhere to schedule in the peak hour at a specific stop, if it arrives between 1 minute early and 3 minutes late compared to its scheduled arrival (e.g. SL, 2009). Due to the small share of early arrivals, the schedule-based strategy (s=1) did not improve substantially the share of on-time arrivals compared with no control (s=0). Headway-based strategies may cause systematic deviation from the schedule especially with higher values of the headway ratio parameter ($\alpha$).

The evaluation of holding control strategies requires taking into consideration their effect on several measures. On the one hand, an effective holding strategy improves service reliability by reducing headway variability or increasing schedule adherence which result in shorter passenger waiting times, more even crowding levels and lower fleet costs due to layover considerations. On the other hand, holding
strategies may cause delays to passengers on-board and longer travel times that may require higher fleet costs. As expected, higher values of headway thresholds result in both higher probability to trigger holding and longer holding times (Figure 6.7). The majority of the buses are not held at stops in all holding scenarios. The small portion of long holding times (more than 2 minutes) under mean-headway strategy (s=3) is eliminated by introducing an upper limit on the holding time (s=4).

Figure 6.7: Distribution of holding time at time point stops under various holding strategies (inbound direction)

Figure 6.8 summarizes the trade-off between the decreased waiting time per passenger (y-axis) and the average increase in in-vehicle time (IVT) (x-axis) caused by the implementation of holding control strategies. The dark square denotes the no control strategy without any delay caused by holding and the benchmark for waiting time, assuming that passengers arrive randomly at stops. Such an analysis can identify dominated alternatives – alternatives that can be eliminated as they are worse than another alternative in at least one performance measure and are not better than it in any other performance measure. In our case, none of the holding control strategies are dominated by the no-control scenario. The mean-headway strategy (s=3) is dominated by its restricted version (s=4) as it results in both longer average passenger waiting times and riding times. The schedule-based strategy (s=1) is also dominated by the preceding-headway strategy with $\alpha = 0.6$ and $\alpha = 0.7$. It is worth noting that in the case of longer headways, when passengers arrive at stops according to the schedule and high
importance for transfer coordination, a schedule-based strategy that improves schedule adherence may result in shorter waiting times. The preceding-headway strategy together with the no-control scenario and the restricted mean-headway strategy compose the set of non-dominated strategies in our analysis. Higher thresholds obtained larger gains in terms of reduced waiting time but also an increase in passenger delay due to holding times. This relationship seems to have a linear pattern within the recommended range of values. In order to determine which alternative is the optimal strategy for a specific case, it is common to compose a compensatory function by defining the relative importance of the different objectives. For example, if passenger waiting and IVT are the only two relevant objectives, and assuming the common ratio of 2 between their relative disutility (passenger waiting time is twice as important as passenger IVT), all the non-dominated strategies are beneficial.

Figure 6.8: Trade-off between passenger in-vehicle delay and waiting time under various holding strategies (inbound direction)

In conclusion, all candidate holding strategies improved service regularity substantially. In particular, an analysis of the results suggests that a holding strategy
based on the mean-headway from the preceding bus and the next bus, restricted by a maximum allowable holding time \((s=4)\), is especially efficient. This strategy yields significant gains in terms of improved service regularity and shorter passenger waiting times. In addition, its narrower distribution of headways has the potential to compensate for the increase in total travel time, by requiring a smaller fleet size for carrying out the same service frequency under a given layover policy.

### 6.3 Case Study II: Bus Line 1, Stockholm\(^5\)

The Tel-Aviv case study showed that the implementation of a headway-based holding control strategy that is based on real-time AVL data can improve service regularity. The results of the even-headway strategy were especially promising. However, the applicability of this strategy depends on its impacts on fleet considerations and driver scheduling. Improved regularity has potential benefits for both passengers and operators, while longer travel times caused by holding buses at stops is the drawback of introducing holding control. Therefore, an analysis of holding strategies has to investigate the trade-off for both passengers and fleet management.

More detailed and comprehensive AVL and APC data was available for trunk lines in Stockholm. The trunk lines, widely known as ‘blue lines’ in Stockholm, were introduced gradually starting from 2005 as part of a public transport reinforcement package that complemented a congestion charging trial (Kottenhoff, 2006). The four trunk lines that operate in the inner city account for approximately 60% of the total number of bus trips in this area (SL, 2006).

These lines are characterized by high frequency, articulated bus vehicles, designated lanes at main streets, high level of signal priority and real-time arrival information at stops. Line 1 was chosen as a case study for a multi-perspective assessment of control strategies because of its typical high-demand inner-city line characteristics.

---

6.3.1 Line Description
Several holding strategies were applied in BusMezzo on bus line number 1 in Stockholm. The line route connects Frihamnen, the main harbor in the eastern part of the city, the city center, a business district and western residential areas in Stora Essingen (Figure 6.9). The line includes 33 stops on the eastbound direction route (ER) and 31 stops on the westbound route (WR). Transit performance is analyzed for the afternoon peak period between 15:30 and 18:00.

Figure 6.9: The route of bus line 1 in Stockholm inner-city

Figure 6.10 presents passenger load profiles for both directions for the peak afternoon time interval. Note that boarding and alighting bars refer to the axis on the left, while through passengers (number of on-board passengers who do not alight at this stop) and passenger load curve refer to the axis on the right side.
6.3.2 Analysis of the current performance

The performance of line 1 was analyzed based on detailed and comprehensive AVL data for a week of regular operations in May 2008. Each bus that operates in Stockholm is equipped with a computer that is called BusPC and is located at the driver's cabin. This system enables radio communication with the control centre, receiving text messages and the automatic display and recording of real-time information based on the AVL system. The software is provided and supported by INIT. The current BusPC display shows the three next TPS or terminals and the scheduled time at these stops (Figure 6.11). In addition, a measure of schedule adherence is shown continuously at the top right corner of the BusPC screen. It is calculated based on the actual location of the bus.
A plus sign indicates that the bus runs ahead of the schedule and a minus sign indicates that the bus runs behind schedule. The scheduled adherence measure is given at the half minute level.

![Bus Driver Display](image)

Figure 6.11: A snapshot of the current bus driver display

Figure 6.12 presents actual and scheduled vehicle trajectories on the afternoon peak hour for a representative day. The three TPS are clearly evident by vertical lines that reflect the holding time in case of early arrivals. The graph illustrates the propagation of the bunching phenomenon. For example, in the case of the first and the second displayed trajectories, the first bus becomes increasingly late and the second bus catches up until the buses are completely paired. The current holding control does not prevent this process. In the case of the last couple of trajectories, although they departed with headway of 5 minutes, their headway increases to 12 minutes towards the end of the route.
The contract between the regional transit authority, SL, and bus operators determines that latter may be subject to penalties based on their performance. Service punctuality is an important clause in this contract. Figure 6.13 shows the share of early, on-time and late departures based on SL criteria from each stop along both directions. TPS are highlighted. The overall pattern is that the share of on-time departures deteriorates along the route from a level of above 0.7 down to 0.35. The share of early departures decreases at TPS due to holding of early arrivals. However, this affect does not last with an immediate increase in subsequent stops. Moreover, even at TPS a considerable share of the buses (10-20%) depart early. The share of late departures is not affected by TPS as the current control strategy does not handle buses that are behind schedule.
The AVL data facilitates the analysis of the speed profile along the line by considering the time interval between departure from one stop and arrival at the next stop. The average running speed is in the range of 7 to 40 km/h. As expected, the average speed is higher at both edges of route directions and lower at the middle part that runs through the inner-city.

The continuous schedule deviation display enables drivers to adjust their speed accordingly. In order to investigate the extent of these potential speed adjustments, the correlation between the average speed and the corresponding measure of schedule adherence that was displayed on the BusPC system was calculated between each pair of stops.
consecutive stops. The displayed measure was approximated by using the following formula:

\[ \text{OTD}_s^k = \lfloor ET_s^k - SET_s^k \rfloor_{30} \]  \hspace{1cm} (6.6)

Where \( \text{OTD}_s^k \) is the on-time display for the driver on trip \( k \) when driving between stops \( s \) and \( s+1 \). The operator \( \lfloor \cdot \rfloor \) indicates rounding down \( Y \) to the closest divider by \( x \). \( ET_s^k \) and \( SET_s^k \) are the exit (departure) time and actual arrival time of trip \( k \) from stop \( s \), respectively. The average speed is positively correlated with the schedule adherence measure on most route segments, as presented in Figure 6.14. The overall correlation is 0.2, suggesting that drivers adjust their speed between stops to reduce their deviation from the scheduled time. However, these adjustments are restricted by traffic dynamics, signals and speed limits. Moreover, stronger correlations tend to occur at segments preceding TPS (marked by dashed vertical lines). This observation suggests that drivers adjust their speeds just before approaching those stops in order to make it within the desired time window. Therefore, there are indications that drivers can and do adjust their speeds based on time-dependent service performance. Moreover, these adjustments depend on the TPS layout - locations where the performance is measured. Note that the speed adjustment on each segment contributes to the on-time performance downstream.
Ingemarson (2010) conducted an analysis of line 1 running times before and after the introduction of priority measures and congestion charge in Stockholm. She found that running speeds remained unchanged. These findings should be interpreted in the context of the current schedule-based control. Timetables are not merely a
reflection of running times but rather an important determinant of running time. Under the current holding strategy, drivers follow the timetable both through holding at stops and speed adjustments between stops. Hence, as long as the timetables do not reflect the updated traffic conditions, buses would not exploit them. Figure 6.15 presents the distribution of the total trip time on the ER that was derived from the AVL data. The current timetable trip time is 51 minutes which corresponds to the 83rd percentile of the distribution.

As a high-frequency line, the main determinant of level of service is service regularity. Figure 6.16 presents headway distributions at the origin terminal and the three TPS on both line directions based on the AVL departure times during the afternoon peak period. The distribution is narrowest at origin terminals with a central value that corresponds to the planned headway. However, even at the origin terminal there is a high variation in headways- 15-18% of headways less than 2 minutes and more than 10% longer than 7 minutes. The shares of very short and very long headways increase along the route, with no correction at TPS. The coefficient of variation of the headway doubles along the route from an already high level of 0.6 to 1.2 – indicating that service regularity deteriorates considerably along the route under the current control strategy. An analysis of the data suggests that there is high variability in
terminal operations between days. An analysis of a longer period is needed in order to investigate it further.

Figure 6.16: Headway distribution at terminals and TPS along the ER (up) and WR (down).

6.3.3 EXPERIMENT DESCRIPTION
The case study represents in detail the bus line characteristics based on empirical data. The operational characteristics of line 1 were analyzed in detail based on AVL and aggregate passenger demand data obtained from APC in order to represent them adequately in the simulation model. Since the case study focuses on a specific bus line, travel times are regarded as an exogenous process that results from time-dependent traffic conditions in the transport network. Travel times between each pair of consecutive stops were analyzed and travel times on all links were found to follow the Log-Normal distribution based on Chi-Square and Kolmogorov-Smirnov goodness-of-fit
tests with a confidence level of 95%. The parameters that yielded the best goodness-of-fit for each link were given as input to the stochastic travel time generator in BusMezzo. As travel times between consecutive links are potentially dependent due to queue propagation and network configuration, the correlation between travel times on each pair of consecutive links was calculated. All correlations were found to be less than 0.3, thus link travel times are regarded as independent stochastic processes. A complete description and detailed data analysis are available at Larijani (2010).

The Headway of bus line 1 is between 4 and 5 minutes during the entire afternoon peak period. The real-world timetable was given as input to BusMezzo. In addition, vehicle scheduling is simulated according to the actual trip chaining used for operating the bus line. The coefficients of the dwell time function are based on values calibrated for local data by the metropolitan transit agency.

Passenger demand is represented in this case study in terms of arrival rates and alighting fractions (see Section 3.4.1). This level of representation enables to capture the interaction between passenger activity at stops and transit performance and allows analysis of the impacts of various holding strategies on the level of service. Passenger arrivals follow a Poisson process as it is a high-frequency line. In fact, the operator publishes the planned headway rather than the scheduled arrival time at stops during the peak periods. Along the bus route there are three major transfer stops from the metro and bus systems that can potentially cause non-random arrival patterns. However, service frequency for all cases is very high (more than 40 vehicle arrivals per hour) and therefore the passenger arrival process is assumed to be random at all stops. Time-dependent passenger demand rates were obtained from aggregated passenger demand and dwell time data. There are two underlying assumptions used in this procedure:

1. The dwell time function is assumed to be dominated by the boarding procedure, hence:

   \[ DT_{s,l}^k = lost\_time + \beta_b \cdot B_{s,l}^k \]  \hspace{1cm} (6.7)

   Where:

   - \( DT_{s,l}^k \) - dwell time for trip \( k \) on line \( l \) at stop \( s \)
   - \( lost\_time \) – constant delay
\( \beta_b \) – service time per boarding passenger

\( B^k_{s,l} \) - number of boarding passengers

2. All arriving passengers are able to board the first arriving vehicle (capacity constraints are not bounding). Therefore, the generation rate can be estimated based on the dwell time relation and the service frequency by applying the following calculation:

\[
\lambda_{s,l}^\tau = \frac{\bar{d}_{s,l} - \text{lost time}}{\beta_a} \cdot \frac{60}{H_l^\tau} \tag{6.8}
\]

Where:

- \( \lambda_{s,l}^\tau \) - passenger generation rate at stop \( s \) for line \( l \) during time period \( \tau \)
- \( H_l^\tau \) - planned headway for line \( l \) during time period \( \tau \)

Note that these assumptions are not used in the simulation model but are solely used for the demand segmentation procedure. For each 30 minutes interval the corresponding passenger demand was estimated using the locally calibrated relationship between dwell time and the number of boarding passengers.

**Scenario design**

As mentioned in Section 6.1, the design of holding strategies involves the determination of TPS layout as well as the enforced holding criteria. This case study explores the effects of TPS layouts under various holding strategies. In order to analyze and evaluate different holding strategies, the case study consists of three schemes for determining TPS layout and three rules for defining the holding criteria. In line with the common practice among bus operators (van Oort et al., 2010), bus lines in Stockholm are regulated using a schedule-based holding control. There are three TPS along the route of bus line 1 (stops 10, 17 and 23 on ER and stops 10, 17 and 24 on WR) where buses are being held if they arrive earlier than the scheduled time. In addition to the base case scenario of the current schedule-based holding strategy, two headway-based holding schemes aimed towards improving service regularity were implemented: a strategy based on a minimum headway requirement from the preceding bus (denoted by \( MH \) and defined by equation 6.2 with \( \alpha = 0.8 \)) and; a strategy based on even headways between the preceding bus and the following bus (denoted by \( EH \) and defined by equation 6.4 with \( \alpha = 1.0 \)).
The current TPS were selected based on network configuration by identifying the main transfer stops from the metro system (marked on Figure 6.9). Alternatively, TPS can be selected based on passenger demand and operational characteristics. As previous studies concluded, TPS should be located at the beginning of a sequence of high-demand stops while avoiding stops characterized by high levels of through passengers (Figure 6.10). Moreover, since holding strategies aim to improve service regularity, it is useful to analyze the trend along the route for relevant measures (e.g. punctuality, variability of the headway) and identify critical points. These points may be associated with segments that experience high travel time variability, that contribute to service irregularity or have irregular passenger activity patterns. Applying those techniques and rules-of-thumb on bus line 1 yielded four proposed TPS nicely distributed in each direction: stops (10,15,20 and 27) on the ER and stops (6,14,20 and 25) on WR.

Although hypothetically all stops can be defined as TPS, departure times are usually regulated only at a small subset of stops along a bus line. The availability of Automated Data Collection (ADC) equipment and communication systems enable to instruct drivers continuously. The third layout uses each stop along the route as a TPS, where the vehicle can be potentially held. This layout can be useful to spread the control mechanism over the entire line and prevent the propagation of discrepancies. Van Oort and Van Nes (2009) reported the implementation of a schedule-based strategy at all-stops on a LRT line in The Netherlands, where it showed substantial benefits.

The experimental design results in nine holding scenarios based on the combination of three holding criteria and three sets of TPS as summarized in Table 6.2. For each scenario 10 simulation runs of the afternoon peak period were conducted. Using the standard deviation of the headway, an outcome of complex interactions between interrelated stochastic processes in the system, 10 repetitions yielded an allowable error of less than 8%. The total execution time for the 10 runs was less than 2 seconds on a standard PC. All of the reported results are the average of the 10 replications for each scenario.
Table 6.2: Experimental design for holding scenarios

<table>
<thead>
<tr>
<th>Holding criteria \ Time point locations</th>
<th>Schedule-based</th>
<th>Minimum headway</th>
<th>Even headway</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current</td>
<td>S1</td>
<td>MH1</td>
<td>EH1</td>
</tr>
<tr>
<td>Proposed</td>
<td>S2</td>
<td>MH2</td>
<td>EH2</td>
</tr>
<tr>
<td>All stops</td>
<td>S3</td>
<td>MH3</td>
<td>EH3</td>
</tr>
</tbody>
</table>

6.3.4 RESULTS AND DISCUSSION

6.3.4.1 SERVICE REGULARITY

Holding strategies were analyzed at different levels, as the detailed representation of bus operations in BusMezzo enables to evaluate system performance and level of service at various levels from a specific trip or stop to overall system measures. The effect of TPS on service irregularity as measured by the coefficient of variation of the headway is clearly evident in Figure 6.17. As in the previous case study, service unreliability propagates along the route. Headway variability decreases significantly immediately after a TPS, restraining the continuous increase in service irregularity. The same pattern is obtained from implementing holding strategies at the alternative TPS locations. With the current TPS layout, the even-headway strategy (EH1) is the most efficient strategy yielding lower coefficient of variation of the headway at almost any given point along the route. In the case of control at all stops, service regularity improves remarkably under headway-based strategies with almost no propagation of variability when even-headway strategy is in place (EH3). The schedule-based strategy is ineffective in regulating the service except of marginal benefits when scheduled adherence coincides with regularity. Compared with the AVL data (Figure 6.16), the simulated data has a similar pattern but with a much lower variability when dispatching form the terminal. This may be due to an underestimation of the error term in expression 3.12 which accounts implicitly for trip chaining from other lines and depots. A comprehensive validation will allow the fine-tuning of the model.
The complete headway distribution under all the nine scenarios is presented in Figure 6.18. Presenting these distributions together enables to detect the main patterns and identify the probabilities of extreme headway values which indicate bunched buses. The first conclusion is that the holding strategy defines the headway distribution more than the TPS layout does. Headway-based strategies result in much narrower distribution with considerably lower probabilities for very short and very long
headways than in the case of schedule-based strategy. Secondly, it is evident that in the case of headway-based strategies the distribution is similar for the current and proposed TPS layout but becomes much narrower for the all-stops layout.

![Graph showing headway distribution under various time point layouts and holding strategies combinations](image)

Figure 6.18: Headway distribution under various time point layouts and holding strategies combinations

Table 6.3 presents a number of measures of performance at the system level for each scenario. The coefficient of variation of the headway presented in the table is the mean value over all stops. Headway-based strategies reduced headway variability substantially compared with schedule-based holding. The EH strategy performed better than the MH strategy. In addition, regulating at all stops improves the regularity considerably while the proposed set of TPS results in slightly better service reliability than the current TPS locations.

The improvement in service regularity results in shorter passenger waiting times which were calculated based on the disaggregate output data and take into account the extra waiting time caused by denied boarding in case of capacity constraints. Furth and Muller (2006) suggested that the total waiting time for high-frequency services corresponds to the 95th percentile of the headway distribution due
to budgeted waiting time. The underlying assumption is that passengers plan their trip so that their total travel time would not exceed a certain value on 95% of their trips. As is evident in Figure 6.18, the headway-based strategies narrow down the respective waiting time value considerably with a reduction of 40% when EH3 is compared with the base case scenario. In addition, two measures of bunching were calculated:

- ‘Bunching1’ - the share of headways which are shorter than 60 seconds or longer than twice the planned headway, as was used by Milkovits and Wilson (2010).
- ‘Bunching2’ - The share of headways that deviate from the planned headway by more than 50% - either up or down. This is the definition proposed by TCRP (2003a, p. 3-48).

The two measures follow a similar trend, except of the case of S2. Very low bunching rates of less than 5% are obtained for headway-based control at all stops. The corresponding Level of Service (LOS) scores in terms of headway regularity are also included in the table. These scores are derived from an ordinal scale that was established by TCRP (2003a) and is shown in Table 6.4. The results suggest that moving from the holding control that is the current practice (S1) to other strategies will improve the level of service from frequent bunching up to the level of clockwork performance.
Table 6.3: Service measures of performance under various holding scenarios

<table>
<thead>
<tr>
<th>Scenario</th>
<th>CV (h)</th>
<th>Avg. waiting time per pass. (sec)</th>
<th>Total waiting time (sec)</th>
<th>Bunching1 (%)</th>
<th>Bunching2 (%)</th>
<th>LOS</th>
<th>Avg. standing time per pass. (sec)</th>
<th>On-time arrivals (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>0.54</td>
<td>173</td>
<td>558</td>
<td>7.9</td>
<td>30.3</td>
<td>D-E</td>
<td>80</td>
<td>79.2</td>
</tr>
<tr>
<td>S2</td>
<td>0.54</td>
<td>174</td>
<td>552</td>
<td>4.5</td>
<td>32.5</td>
<td>D-E</td>
<td>81</td>
<td>76.9</td>
</tr>
<tr>
<td>S3</td>
<td>0.50</td>
<td>165</td>
<td>520</td>
<td>6.0</td>
<td>26.5</td>
<td>D</td>
<td>87</td>
<td>80.0</td>
</tr>
<tr>
<td>MH1</td>
<td>0.39</td>
<td>160</td>
<td>489</td>
<td>2.2</td>
<td>14.6</td>
<td>B-C</td>
<td>63</td>
<td>69.8</td>
</tr>
<tr>
<td>MH2</td>
<td>0.37</td>
<td>158</td>
<td>486</td>
<td>2.0</td>
<td>12.1</td>
<td>B-C</td>
<td>61</td>
<td>67.4</td>
</tr>
<tr>
<td>MH3</td>
<td>0.26</td>
<td>146</td>
<td>425</td>
<td>0.5</td>
<td>4.0</td>
<td>B</td>
<td>65</td>
<td>71.4</td>
</tr>
<tr>
<td>EH1</td>
<td>0.35</td>
<td>151</td>
<td>418</td>
<td>1.6</td>
<td>11.0</td>
<td>B-C</td>
<td>58</td>
<td>78.7</td>
</tr>
<tr>
<td>EH2</td>
<td>0.31</td>
<td>147</td>
<td>400</td>
<td>0.9</td>
<td>8.1</td>
<td>B-C</td>
<td>56</td>
<td>76.7</td>
</tr>
<tr>
<td>EH3</td>
<td>0.18</td>
<td>141</td>
<td>335</td>
<td>0.3</td>
<td>2.6</td>
<td>A</td>
<td>58</td>
<td>67.3</td>
</tr>
</tbody>
</table>

Table 6.4: Transit Level of Service based on headway regularity
(from: TCRP, 2003a, page 3-48)

<table>
<thead>
<tr>
<th>Level</th>
<th>CV(h)</th>
<th>Bunching</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.00-0.21</td>
<td>&lt;1%</td>
<td>Service provided like clockwork</td>
</tr>
<tr>
<td>B</td>
<td>0.22-0.30</td>
<td>&lt;10%</td>
<td>Vehicles slightly off headway</td>
</tr>
<tr>
<td>C</td>
<td>0.31-0.39</td>
<td>&lt;20%</td>
<td>Vehicles often off headway</td>
</tr>
<tr>
<td>D</td>
<td>0.40-0.52</td>
<td>&lt;33%</td>
<td>Irregular headways, with some bunching</td>
</tr>
<tr>
<td>E</td>
<td>0.53-0.74</td>
<td>&lt;50%</td>
<td>Frequent bunching</td>
</tr>
<tr>
<td>F</td>
<td>&gt;0.74</td>
<td>&gt;50%</td>
<td>Most vehicles bunched</td>
</tr>
</tbody>
</table>

Holding strategies could also impact crowding measures. A more regular service increases the probability of having an available seat since passenger load is distributed more evenly between buses. A lower share of bunching avoids the pairing of empty and
overcrowded vehicles. The average standing time per passenger is used as the crowding measure and is a proxy for the level of comfort. It was calculated for a given line \( l \) in the following manner:

\[
AST_l = \frac{\sum_k \sum_s \left[ RT_{s,s}^k \max \left( \frac{0.1L_{s+1}^k}{seats_f} \right) + DT_{s,s}^k \max \left( \frac{0.1L_{s-1}^k - A_{s-1}^k}{seats_f} \right) \right]}{\sum_k \sum_s B_{s+1}^k} \tag{6.9}
\]

Where \( RT_{s,s}^k \) is the riding time from stop \( s \) to stop \( s+1 \) on trip \( k \) of line \( l \). \( DT_{s,s}^k \) is the dwell time. \( L_{s+1}^k \) is the passenger load when approaching stop \( s \). \( A_{s-1}^k \) and \( B_{s+1}^k \) are the corresponding numbers of alighting and boarding passengers, respectively. \( seats_f \) is the number of seats on vehicle type \( f \) that is used to operate trip \( k \).

This calculation sums up all the time components when passenger load exceeds the number of seats. As expected, the average standing time per passenger was reduced by up to 30% when headway-based strategies were implemented.

The last column on Table 6.3 refers to schedule adherence, following again the criteria of arrival within the time window of 1 minute early and 3 minutes late compared to the timetable (SL, 2009). The highest on-time performance is obtained under S3. Interestingly, although EH strategy does not incorporate the schedule into the holding criteria, its implementation resulted in the same level of on-time performance as the schedule-based scenarios, except of the case of all-stops control.

As in the Tel-Aviv case study, the evaluation considers the trade-off between average passenger waiting times and the average increase in passenger IVT caused by its implementation. Figure 6.19 displays how each of the holding strategy scenarios performs on both passenger-time dimensions. In this case, the reference point for waiting times is the hypothetical case of perfectly even headways which imply average waiting time of half the planned headway. It is evident that EH scenarios dominate MH scenarios regardless of TPS locations. In the case of schedule-based strategy, the current TPS dominate the proposed locations.
Headway-based strategies resulted in shorter passenger waiting times in the cost of longer IVT compared with schedule-based holding. According to value of time studies, the ratio between waiting time and IVT is in the range of 1.5-2.0 (Wardman, 2004). Based on these values, the EH strategy results in substantial overall time savings compared with schedule-based strategy. The ratio between the weighted reduction in waiting time and the weighted increase in IVT is 4 or 3 for EH1 and EH3, respectively.

6.3.4.2 Operational Considerations
From the operator perspective, holding strategies have the potential to improve fleet management certainty at the cost of longer total travel times. The result of these two factors in terms of fleet costs depends on the trip travel time distribution as holding strategies are expected to simultaneously increase the average value and reduce its variability. Figure 6.20 presents the total trip time distribution for the ER where according to the timetable the total running time is 51 minutes. Two-sample Kolmogorov-Smirnov tests were conducted in order to compare the distributions of the observed and simulated total trip times (for scenario S1). The test results are that the hypothesis that the observed and simulated distributions are derived from the same distribution cannot be rejected ($D=0.125$ while the critical value is $D_{14,0.05} = 0.349$).
This indicates that BusMezzo reproduced the total trip distribution. It is evident that headway-based strategies yielded a narrower travel time distribution. According to the simulation results, 80% of the trips complete the trip within less than the scheduled time under the strategy that is currently used (S1). This percentage rises to 92% under EH1 and climbs to 99% in case of EH3.

Figure 6.20: Total travel time distribution under various time point layouts and holding strategies combinations

Table 6.5 summarizes some operational measures of performance. The average total cycle time (a bi-directional chain) increases by 1-2 minutes to 100 minutes when MH strategy was implemented compared with schedule-based and EH strategies. This result is consistent with previous findings of Van Oort et al. (2010) which compared a schedule-based and a MH strategy. Furthermore, holding at all stops prolonged the average cycle time by 0.5-1.5 minutes compared with the case of the same strategy and the current TPS layout. This is expected as the holding mechanism is activated more often and reacts before allowing bus movement to compensate for earlier deviations. However, the average running time does not determine the timetable and fleet assignment design. Instead, the common practice among bus operators is to use the 85th or the 90th percentile of trip travel time distribution as the design criterion (TCRP,
2000). Hence, in order to study the effect of holding strategies on fleet assignment, we compare the 90\textsuperscript{th} percentile of total cycle time (a bi-directional chain). The EH strategy narrows the cycle time distribution and yields a reduction of 1.5-3 minutes in the 90\textsuperscript{th} percentile value. This reduction of 3\% in the total cycle time has positive consequences for both operators and passengers. These findings reinforce the conclusions of Daganzo (2009) from an analytical study on a similar holding strategy. In contrast, the MH strategy leads to a small increase of around half a minute. A reduction of 0.5 minute is gained in case of regulating by the schedule at all stops (S3).

### Table 6.5: Operational measures of performance under various holding scenarios

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Average cycle time (sec)</th>
<th>90\textsuperscript{th} percentile of cycle time (sec)</th>
<th>Average holding time per vehicle run (sec)</th>
<th>Average delay at relief point (sec)</th>
<th>SD of delay at relief point (sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>5895</td>
<td>6269</td>
<td>51</td>
<td>222</td>
<td>145</td>
</tr>
<tr>
<td>S2</td>
<td>5916</td>
<td>6268</td>
<td>56</td>
<td>225</td>
<td>131</td>
</tr>
<tr>
<td>S3</td>
<td>5934</td>
<td>6226</td>
<td>81</td>
<td>181</td>
<td>110</td>
</tr>
<tr>
<td>MH1</td>
<td>5944</td>
<td>6274</td>
<td>173</td>
<td>239</td>
<td>154</td>
</tr>
<tr>
<td>MH2</td>
<td>6010</td>
<td>6298</td>
<td>217</td>
<td>254</td>
<td>127</td>
</tr>
<tr>
<td>MH3</td>
<td>6021</td>
<td>6306</td>
<td>207</td>
<td>215</td>
<td>82</td>
</tr>
<tr>
<td>EH1</td>
<td>5874</td>
<td>6168</td>
<td>130</td>
<td>226</td>
<td>134</td>
</tr>
<tr>
<td>EH2</td>
<td>5880</td>
<td>6074</td>
<td>141</td>
<td>236</td>
<td>97</td>
</tr>
<tr>
<td>EH3</td>
<td>5957</td>
<td>6078</td>
<td>175</td>
<td>220</td>
<td>66</td>
</tr>
</tbody>
</table>

A shorter cycle time implies a higher certainty that vehicles would be available on-time for their next scheduled trip. The current level of fleet certainty can be sustained by updating the timetable based on the 90\textsuperscript{th} percentile which would result in a higher frequency for the same fleet size. Alternatively, the current frequency might be operated by a smaller number of vehicles. However, fleet size depends on drivers and vehicle schedules, labor agreements and of course the integer nature of this variable. Hence, the reduction in cycle time may not necessarily be translated into fleet size savings. In addition to ordinary scheduled buses, bus operators typically have reserved
buses that can be allocated as a response to various conditions (e.g. high demand, vehicle breaks down, and large deviations from the schedule). The number of reserved buses may be reduced as buses operate more regularly and with a higher certainty to make it on time to the next trip.

The level of service of high-frequency services depends mainly on headway regularity and therefore the main operational objective is to maintain even headways between consecutive vehicles. However, schedule-based strategy does not require real-time data communication and is also suitable for crew management. Some bus operators use driver schedules that include driver replacement at intermediate stops, also known as relief points. In case there are relief points along the line, this is an additional concern as it is especially important to have high schedule adherence at these stops. Driver relief points may be a potential hindrance to applying headway-based strategies, as schedule adherence is the main concern for driver shift scheduling.

The average and standard deviation of the delay at the relief point (‘Fridhemsplan’) on the westbound route are included in Table 6.5. The complete distribution is presented in Figure 6.21. On this route the relief point is towards the end of the route and therefore subject to more uncertainty. Note that the relief point is also a TPS in the current layout. The average delay is the lowest under S3, while the standard deviation of the delay is the lowest for EH3. Interestingly, the headway-based strategies do not imply longer delays compared with the base case of S1. Moreover, the standard deviation of the delay is significantly lower in the cases of MH3 and EH3. The probability of a very late arrival (more than 5 minutes late) is more than double on S1 compared to EH1. These results suggest that headway-based strategies can even improve the punctuality in the relief point, an important objective of crew management and fleet assignment and an important concern for labor unions.
The analysis of the results highlights substantial potential benefits from implementing an even-headway strategy that regulates the headways with the purpose of equalizing the headway from the preceding bus with the headway from the succeeding bus. This strategy improves the service reliability substantially, leading to passenger time savings, reduced operating costs as well as better schedule adherence at the relief point. Therefore, the even-headway strategy is a promising operation and management strategy in particular when implemented at all the stops along the route.

6.3.4.3 Robustness
The overall robustness of the EH strategy has to be considered with respect to human factors and BusPC design. The potential impacts of practical considerations involved with the implementation of the proposed strategy were analyzed. The evaluation was based on simulating the peak hour only. The following factors were incorporated into the control strategy in BusMezzo: driver display preciseness – exact time difference in seconds vs. half minute level as provided by BusPC; compliance rate – the share of bus drivers that follow the control strategy with the remaining drivers assumed to consistently disregard the strategy and depart without holding; maximum holding time -
an upper bound for the holding time at each TPS. These three factors were embedded in the EH control strategy as follows:

\[
ET_{s,l}^k = \max \left( \min \left( AT_{s,l}^{k-1} + \left[ \frac{AT_{m,l}^{k+1} + SRT_{m,s} - AT_{s,l}^{k-1}}{2} \right], AT_{s,l}^{k-1} + \alpha h_{s,l}^k \right), AT_{s,l}^{k} + DT_{s,l}^{k} \right) \quad (6.10)
\]

\[
ET_{s,l}^k = \max \left( \min \left( AT_{s,l}^{k-1} + \left[ \frac{AT_{m,l}^{k+1} + SRT_{m,s} - AT_{s,l}^{k-1}}{2} \right], AT_{s,l}^{k} + DT_{s,l}^{k} + \gamma_{\text{max}} \right), AT_{s,l}^{k} + DT_{s,l}^{k} \right) \quad (6.11)
\]

\[
ET_{s,l}^k = \begin{cases} 
\max \left( \min \left( AT_{s,l}^{k-1} + \left[ \frac{AT_{m,l}^{k+1} + SRT_{m,s} - AT_{s,l}^{k-1}}{2} \right], AT_{s,l}^{k-1} + \alpha h_{s,l}^k \right), AT_{s,l}^{k} + DT_{s,l}^{k} \right) & \delta_k = 1 \\
\frac{AT_{s,l}^{k} + DT_{s,l}^{k}}{\alpha} & \text{otherwise} 
\end{cases} \quad (6.12)
\]

Where \( \gamma_{\text{max}} \) is an upper bound for holding time and \( \delta_k \) is an indicator that equals 1 if the driver on trip \( k \) complies with the regulations at TPS.

The results of this experiment are presented in Table 6.6. Both EH1 and EH3 were simulated with two imperfect compliance rates of 50% and 75% of the drivers complying with the applied headway strategy while the others simply ignore the BusPC display. The table shows the average coefficient of variation of headways along the line and the relative change compared with the base case of the same strategy with the ideal holding operations: perfect display preciseness, perfect compliance and without the enforcement of maximum holding time enforced. Both strategies are surprisingly robust with respect to driver compliance, presumably due to their cooperative nature where adjacent vehicles can correct for a non complying vehicle and mutual corrections. In an additional test of compliance error at the stop visit level (low awareness or stop-specific constraints that are not associated with a particular driver), the performance of EH3 was only negligibly affected as even the same vehicle can correct itself at the next stop if needed.

The introduction of maximum holding time comes at a higher price in terms of reduced service regularity. It is a practical constraint as long holding times are unacceptable by passengers on-board. Moreover, the preciseness of the BusPC display adds another disturbance into the system that reduces the effectiveness of the holding strategy compared with the hypothetical case of ideal holding conditions. It is important to keep in mind that all these design factors hinder also the current operations and should be regarded as a sensitivity analysis. Furthermore, possible speed adjustments between stops were not modeled in the simulation. Hence, the results may underestimate the benefits from the continuous cooperative nature of the proposed strategy.
Table 6.6: Effects of holding implementation on headway coefficient of variation

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Driver display preciseness</th>
<th>Compliance rate</th>
<th>Max holding time</th>
<th>CV(h)</th>
<th>Increase in the CV(h) compared with respective case (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EH1</td>
<td>Perfect</td>
<td>1.0</td>
<td>None</td>
<td>0.35</td>
<td>+0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.75</td>
<td>None</td>
<td>0.35</td>
<td>+1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.5</td>
<td>None</td>
<td>0.37</td>
<td>+7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.0</td>
<td>1 min</td>
<td>0.42</td>
<td>+21</td>
</tr>
<tr>
<td></td>
<td>30 sec</td>
<td>1.0</td>
<td>1 min</td>
<td>0.46</td>
<td>+31</td>
</tr>
<tr>
<td>EH3</td>
<td>Perfect</td>
<td>1.0</td>
<td>None</td>
<td>0.18</td>
<td>+0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.75</td>
<td>None</td>
<td>0.19</td>
<td>+4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.5</td>
<td>None</td>
<td>0.19</td>
<td>+8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.0</td>
<td>1 min</td>
<td>0.21</td>
<td>+19</td>
</tr>
<tr>
<td></td>
<td>30 sec</td>
<td>1.0</td>
<td>5 min</td>
<td>0.20</td>
<td>+13</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.0</td>
<td>1 min</td>
<td>0.23</td>
<td>+26</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.5</td>
<td>None</td>
<td>0.24</td>
<td>+34</td>
</tr>
</tbody>
</table>

6.3.5 From simulation to a field study
The analysis of the simulation results highlights potential benefits from implementing the EH strategy at all stops. The analysis suggests that implementing this strategy on trunk line 1 will have positive impacts on reliability and hence on passenger waiting times and crowding levels. Moreover, it will benefit fleet operations, while maintaining the schedule adherence in general, and at the relief points in particular. In addition, the underlying inter-dependent mechanism of the EH strategy can be useful in preventing the current speed pattern of slowing down just before approaching TPS that is reinforced by the schedule-based control strategy.

The findings from the analysis of AVL data and the simulation study were discussed with SL and the bus operator of line 1, Keolis, in order to design guidelines for its implementation. It was decided to design a field trial to test the applicability of this strategy and evaluate its performance in real-world conditions. The experiment is scheduled to take place on October 2011. The actual implementation of the proposed control scheme involves complications that require careful consideration as discussed below.
Tenure contract

The current contract conditions between SL and the local bus tenure are an impediment for a potential trial. The contract defines a threshold criterion for the punctuality at TPS of 82%. The operator has to pay penalties for a poorer performance and receives bonuses for a better performance (Ingemarson, 2010 – p.9). Of course the operator has to be assured that the current incentive scheme which is based on on-time performance would not be enforced during the trial period so that it will not result in high penalties on its expense. SL agreed to wave off all penalties associated with the operations of line 1 and all the potentially affected lines during the trial period in order to guarantee the operator’s commitment to the experiment.

Previous attempts

Although a number of studies (Liu and Wirashinge, 2001; Dessouky et al., 2003; Daganzo, 2009) have indicated the benefits of holding strategies, a number of implementation studies have also shown that these benefits are not always realized in practice. Previous reports on field trials of control strategies designed to improve service regularity showed limited results. Important lessons could be drawn from a previous attempt to improve the regularity of line 1 in Stockholm (SL, 2003). During the autumn of 2002, a dedicated dispatcher at the control centre instructed two mobile traffic controllers how to regulate the service with the objective of improving the regularity. The dispatcher was the only person that had access to real-time AVL data. Moreover, the control strategy was defined vaguely with no clear holding criterion. The report of the pilot study concludes that the headway control in addition to other measures that were introduced simultaneously led to small regularity improvement.

A field study in Chicago reported by Pangilinan et al. (2008) investigated a control strategy equivalent to the EH strategy. The trial was also based on a dedicated dispatcher at the control room and supervisors located at key stops along the route. Again, the supervisors passed on the dispatcher instructions to the drivers with only the dispatcher having direct access to real-time AVL data. The authors found that the dispatcher workload did not allow him to detect, not to mention respond to, every service regularity problem even after they simplified the conditions that require intervention. They concluded that this was the main hindrance in their field study.
Nevertheless, service regularity during the trial period improved compared with the previous unsystematic control scheme.

**Driver display**

The implementation of the proposed strategy is facilitated by the capability of the BusPC to display a headway-based measure in addition to the current schedule adherence measure. The additional indicator refers to how far the bus is from being exactly in the middle between the proceeding and the succeeding buses. An indicator with a plus sign indicates that the bus is too close to the bus in front while a negative one means that the bus is too close to the bus behind. If the bus is exactly in the middle then the indicator displayed is zero. This measure can be embedded into the BusPC screen and is behaviorally consistent with the current practice as “plus” requires waiting or slowing down and “minus” to speed up (as illustrated in Figure 6.22, the new measure marked with red circle). On the one hand, having both figures can be a source of confusion especially since drivers are probably very much used to follow the current figure. On the other hand, it can alert drivers that something unordinary is going on and remind them of the trial.

![Figure 6.22: An illustration of the trial period driver display](image)

**Control room**

Dispatchers at the control room have a dynamic display of the real-time location of each bus. The dispatcher monitors the performance, communicate with drivers in case of need (via radio communication or text messages) and initiate interventions (e.g. insert reserve buses, short turning). Carrel et al. (2010) studied the control room dynamics and the main factors that influence controllers decisions based on direct observations at
the control room of the tube Central Line in London operated by TfL. They stress that control room decisions still rely to a large extent on personnel judgment and informal practices. The controllers tend to choose local interventions that fulfill all the constraints and are manageable and flexible but may not be the desired one from a system-wide perspective. An important observation with respect to the proposed field-study is the dominancy of scheduled adherence as a decision factor even when it introduces irregularities. This is mainly due to fleet and crew considerations as the control room needs to keep the vehicles’ and drivers’ operations running on-time. In addition, the cognitive effort involved with keeping track of potentially conflicting objectives as service regularity and vehicle and crew considerations may lead controllers to concentrate on the more constrained dimensions.

In light of the reports of Carrel et al. (2010) and Pangilinan et al. (2008), it is highly recommended that line 1 will be monitored by a dedicated dispatcher during the trial period as it will require special care. In addition, it is important that the dispatcher will have a display that will ease the detection of irregular runs that require dispatcher’s attention (e.g. visualization of distances between buses).

**Design**

The headway-based control should be applied only during high-frequency periods. Other periods (e.g. nights, weekends) have to be exempted. In addition, in the case of an incident (e.g. vehicle breaks down, accident) it may be necessary to either dismiss a certain bus (‘take it out of the system’) or switch the operation to schedule-based mode to prevent unwanted holding times.

According to the simulation results treating all stops as TPS (EH3) has more significant gains in terms of service regularity than restricting holding to the current set of TPS (EH1). Enabling to hold at each stop prevents the accumulation of irregularity instead of resolving it only at the next TPS further downstream. Another advantage is that holding times are spread between more stops, hence requiring shorter holding time per stop as was the case when a schedule-based holding at all stops was applied on a LRT line in The Netherlands (Van Oort and Van Nes, 2009). However, this approach may be difficult to implement due to local traffic dynamics, stop capacity constraints and the human factor. A possible compromise might be that drivers should keep an even headway at all stops by holding if necessary based on their judgment of local conditions.
while making sure that they fulfill the criteria at the three current TPS. A continuous control may also contribute to changing the current pattern of speed adjustments just before approaching a TPS.

**Driver and vehicle scheduling**

Driver and vehicle scheduling are important constraints in an actual implementation of the proposed strategy. Regulating the dispatching from the origin terminal can potentially prevent some of the initial variability introduced already at the beginning of the route. The results of Pangilinan et al. (2008) suggest that headway-based dispatching from the origin terminal plays an important role, an issue highlighted also by Van Oort and Van Nes (2009). Hence, it is preferable to assign a set of dedicated drivers to line 1 during the trial period. However, transit lines do not operate as a closed system. Regulating the headway at terminals may be complicated in real-world operation conditions since drivers and vehicles are circulated between lines. Hence, it may be worthwhile to treat the terminal as a TPS both in terms of schedule and headway. In other words, buses may depart later than the timetable in case the headway from the preceding bus is too short. This is aimed to prevent the departure of bunched buses from the origin terminal which is evident under the current control scheme (Figure 6.16).

**Driver behavior**

The operator monitors the on-time performance but there are no incentive scheme to adhere to the schedule at the individual driver level. The on-time performance analysis (Figure 6.13) indicated that a considerable share of the drivers depart early from TPS in contrast with the current holding strategy. This may be due to break regulations; local traffic conditions or poor compliance.

Drivers experience stress related to the uncertainty involved with schedule adherence. The proposed operational scheme could potentially reduce bus drivers’ stress as it relieves the need to refer all the time to the timetable. Furthermore, while the current strategy does not help a late driver, the proposed strategy allows the succeeding driver to ‘correct’ his/her relative location by holding back.
Speed adjustments based on the even-headway measure should be encouraged. This possibility was explored analytically by Daganzo and Pilachowski (2011). It can act as a balancing force to ‘push forward’ buses instead of just holding them back. The analysis of speed patterns and its correlation with schedule adherence (Figure 6.14) suggests that drivers can and do adjust their speeds based on the service performance. These adjustments depend albeit on the TPS layout - the locations where the performance is measured. Clearly, drivers cannot perfectly adjust their speeds as they are subject to traffic dynamics, signals and speed limits.

The trial depends on the collaboration of drivers and control centre personnel. The following channels may be used for disseminating information: group meetings; brochures; newsletter; a slip on the driver schedule; post reminders at depots and on driver cabins; a pop-up message on BusPC.

**Evaluation**

The assessment of an actual field implementation should be based primarily on before-after analysis of AVL and APC data. The comparison has to be made for equivalent months that have similar traffic conditions and passenger demand levels. The integration of performance analysis and post-trial interviews will allow drawing conclusions on the applicability of the proposed scheme and possible refinements. In case that the control strategy is to be adopted on a larger scale, then the analysis may propose regularity measures that could be incorporated into the incentives scheme.

The evaluation of a field trial should also investigate its impact on riding times. Although running times are not an explicit objective of the proposed strategy, with its implementation the whole system is expected to get closer to the speeds enabled by the priority measures. Thus, it can help to resolve some conflicts between transit authorities and operators on the construction of timetables. The impacts of the proposed strategy could be summarized by assigning monetary values to the different evaluation aspects – passenger-time, fleet considerations and labor-time – and calculating the net effect. Such an approximation would support the assessment of the strategy by the transit authority.
6.4 Optimizing the Number and Location of Time Point Stops

6.4.1 Related Work
How should the operator determine which stops will be used as TPS? This practical question is the subject of on-going research efforts using either numerical analysis or simulation studies. In general, too few TPS may be insufficient for preventing/correcting the bunching phenomenon causing long passenger waiting times and uneven crowding. Too many TPS may in contrast lead to long travel times and increased operational costs. Section 6.3.3 investigated an alternative TPS layout that did not perform better than the current layout although it was determined based on the results of previous studies. In spite of the fact that the case study found holding at all stops to be advantageous over the alternative layouts, it is not necessarily the optimal layout.

Recently, several studies formulated the holding control problem as an optimization problem. The objective function is a measure of total cost and the holding times are the decision variables. Delgado et al. (2009) formulated the problem as a deterministic mathematical program and solved it for a circular route. Yu and Yang (2009) and Grube and Cipriano (2010) applied a genetic algorithm for the holding control problem where the decision variables were the holding time at each stop and the objective was to minimize total waiting times. However, both studies did not analyze the spatial allocation of the holding times. Cortes et al. (2010) applied a genetic algorithm for a pre-defined set of TPS and possible holding times. A similar problem formulation was solved by Mazloumi et al. (2011) using an ant colony optimization technique with a discrete set of possible holding times. However, in all these studies the number and/or the subset of potential location of the stops were predetermined.

This study formulates the time point layout (TPL) problem as an optimization problem. The number and location of time points are considered simultaneously. The following section formulates the problem as a multi-objective optimization problem which reflects passengers and operators perspectives. Section 6.4.3 and 6.4.4 presents two alternative solution methods and their respective results. Finally, Section 6.4.5 compares the results with the no control scenario and the TPS that are currently used by the operator and draws conclusions. For the complete description and results see Mach-Rufi (2011).
6.4.2 PROBLEM FORMULATION

The passenger-related objective function includes all passenger time components in order to capture the total travel time from the origin stop to the destination, as follows:

$$ f_1(\bar{x}) = \sum_{i \in I} [\gamma_w \cdot w_n(\bar{x}) + \gamma_r \cdot r_n(\bar{x}) + \gamma_d \cdot d_n(\bar{x}) + \gamma_h \cdot h_n(\bar{x})] $$

(6.13)

Where $w_n$, $r_n$, $d_n$ and $h_n$ are the waiting, riding, dwelling and holding times of passenger $i$, respectively. The sum of the three latter components is the IVT. $I$ is the set of all the passengers using the analyzed system. $\bar{x}$ is the binary vector of decision variables which corresponds to the list of stops along the line. Each value in $\bar{x}$ takes '1' if the respective stop is a TPS and '0' otherwise. The weights assigned in this study to the objective function components were $\gamma_w = 2$, $\gamma_r = 1$, $\gamma_d = 1$ and $\gamma_h = 1.5$ in order to account for the larger burdensome associated with waiting and holding times.

The operator-related objective function is based on the total travel time distribution criterion used for determining the number of vehicles required to operate a certain line with a given frequency. The common practice among transit operators is to use the 90th percentile of the distribution (TCRP, 2000):

$$ f_2(\bar{x}) = p_{90} \left( TTT(\bar{x}) \right) $$

(6.14)

Where $p_{90} \left( TTT(\bar{x}) \right)$ is the 90th percentile of the vehicle total travel time distribution. Hence, the objective function is a vector defined as: $F(\bar{x}) = \{ f_1(\bar{x}), f_2(\bar{x}) \}$.

BusMezzo was used for evaluating TPS configurations as it represents the major sources of uncertainty and enables the application and evaluation of adaptive control strategies within a short running time. The use of a simulation model allows the incorporation of many stochastic and dynamic transit operations processes which compose a complex system. This analysis expands on previous studies by considering both the optimal number and optimal location of the time points, while holding times are based on a real-time headway-based strategy. All the measures of performance were derived from simulating the operations of line 1 in Stockholm. Assumptions regarding how many vehicles are affected by each control decision become redundant. Based on the results reported above for both Tel-Aviv and Stockholm case studies, the even-headway strategy was applied throughout the optimization process.

The optimization problem was solved using a greedy heuristic approach and a genetic algorithm. The algorithms were programmed in MATLAB. The program revised the input, called repetitively BusMezzo for simulating each candidate solution, stored the relevant output data and calculated the objective functions. In order to guarantee a
maximum allowable error of 5% with respect to the standard deviation of the headway, all reported results are based on 50 replications.

6.4.3 Greedy Algorithm

6.4.3.1 Description
A greedy (or myopic) algorithm is a heuristic which selects the locally optimal choice at each stage without taking into account potential implications on the global solution. This heuristic involves the sequential solution of simpler sub-problems. The greedy algorithm was applied to solve the TPL problem as presented in Figure 6.23. Note that this algorithm was applied only with respect to the passenger-time objective function (expression 6.13). The initial solution is the no control scenario, where all decision variables are set to zero. At every step of the algorithm, an additional TPS is added at the stop that minimizes \( f_1(\bar{x}) \). This process continues as long as the objective function is improved by the additional TPS. Note that this heuristic involves the evaluation of \((|S| - k + 1)\) scenarios at each stage, where \(S\) is the set of stops and \(k\) is the stage counter.

The pseudo code of the greedy algorithm is as follows:

<table>
<thead>
<tr>
<th>Step</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1</td>
<td>Initialization - ( k = 1, \bar{x}^0 = 0 )</td>
</tr>
<tr>
<td>Step 2</td>
<td>No control scenario - Run BusMezzo with ( \bar{x}^0 ), calculate ( f_1(\bar{x}^0) )</td>
</tr>
<tr>
<td>Step 3</td>
<td>Try each stop - For ( j = 1 ) to ( N ): If ( x_j^k = 0 ), then set ( b_j^k(1, ..., j, ..., N) = (0, ..., 1, ..., 0) ), run BusMezzo with ( \bar{x}^k = \bar{x}^{k-1} + b_j^k ) and calculate ( f_1(\bar{x}^k) ).</td>
</tr>
<tr>
<td>Step 4</td>
<td>Find minimum - ( \bar{x}^{k+1} = \arg{\min_{j} f_1(\bar{x}^k)} )</td>
</tr>
<tr>
<td>Step 5</td>
<td>Check improvement - If ( f_1(\bar{x}^{k+1}) \geq f_1(\bar{x}^{k-1}) ) then stop and give as an output ( \bar{x}^{k-1}, f_1(\bar{x}^{k-1}) )</td>
</tr>
<tr>
<td>Step 6</td>
<td>Check stop criterion - If ( k = 2N ) then stop and give as an output ( \bar{x}^{k+1}, f_1(\bar{x}^{k+1}) )</td>
</tr>
<tr>
<td>Step 7</td>
<td>Advance stage counter - Set ( k = k + 2 ) and return to step 3.</td>
</tr>
</tbody>
</table>
**6.4.3.2 Results and Discussion**

The greedy algorithm was performed without the stopping criterion in order to assess its performance throughout the entire range. The number of iterations equals therefore the number of stops along the route. Each stage determined an additional TPS where the lowest objective function value was obtained. The objective function decreases sharply after the first TPS was introduced and reaches its minimum value with the introduction of the third TPS. Hence the algorithm would have stopped after the third iteration if the stopping criterion was enforced (Figure 6.24). The objective function remains low until the 9th time point was introduced and then rises considerably until

![Greedy algorithm flowchart](image-url)
the 20th time point has the same level as the no control scenario. The all stops scenario has an objective function value that is 15% higher than the no control scenario.

Figure 6.24: Objective function value for the complete run of the greedy algorithm

The main determinants of the variation of the objective function value are the holding and waiting times. Figure 6.25 shows the different trends that these two components exercise and provide insight on their accumulated affect on the objective function value. Total pass-holding time increases as the number of TPS increases, especially after the 9th iteration. In contrast, the total passenger-waiting time initially decreases but barely changes afterwards. Minimum waiting times are obtained at the 10th iteration with a 23% decrease compared with the no control scenario. The trend of the total-waiting time is consistent with the one reported by Mazloumi et al. (2011) for an increasing number of TPS in the optimization process. In their case it dropped until the third TPS and then remained relatively flat. However, the trend of the holding time cannot be compared since they constrained the total slack time over the route.
Figure 6.25: Evolution of the holding and waiting times

Figure 6.26 presents the coefficient of variation of the headway over the iterations of the greedy algorithm. A dramatic improvement in service reliability is achieved already by the first TPS which leads to a 50% reduction in the coefficient of variation along the route. The general trend is that adding more TPS reduces the coefficient of variation. This stands in contrast to the analysis performed by Senevirante (1990). However, it did not consider dynamic control strategy and transit operations.

Figure 6.26: Evolution of the coefficient of variation of the headway
Figure 6.27 presents the sequence in which the time points were selected. The three stops that make up the solution of the greedy algorithm are 13th, 1st and 7th, selected in this order. Most of the stops that were selected on early iterations are located at the first half of the route, where regulating the service affects a large share of downstream stops. Few stops towards the end of the route were chosen at early stages of the algorithm. This is due to their very marginal affect on the objective function value since they do not impose delays on many passengers on-board.

![Figure 6.27: Time point stop location in the complete greedy algorithm run](image)

6.4.4 Genetic Algorithm

6.4.4.1 Description
Genetic algorithms apply the principles of the natural selection theory in the domain of optimization methods. These algorithms adopt the principles of evolution and survival of the fittest as well as concepts as mutation, crossover and migration in order to describe complex stochastic systems. It is considered a powerful and robust tool to
reach close-to-optimum solutions. Greedy algorithms were applied within a large range of optimization problems (Siranandam and Deepa, 2008).

In the terminology of genetic algorithms, each feasible solution is an individual and the set of solutions considered at each stage is referred to as population/generation. An individual has to be represented in the form of a chromosome (a sequence of genes). In our TPL problem, the chromosome is the vector of binary decision variables \((\bar{x})\) and the fitness of an individual is the value of the corresponding objective function \((f_1(\bar{x}))\). The individuals in each generation will converge progressively to a space of solutions that is closer to the optimal solution. However, the algorithm does not guarantee that the optimal solution would be found in the last generation due to the stochastic nature of both the objective function and the solution method. The progress of the genetic algorithm when applied to the TPL problem is presented in Figure 6.28.

The genetic algorithm was performed with a population size of 20 individuals. The initial probability to be a TPS (for a gene to be ‘1’) was set to 15%. The two individuals with the best performance from each generation were reproduced into the next generation (‘elitism’). The rest of the solutions had a reproduction probability that is proportional to their performance. The crossover of genes in the reproduction process was uniform. This implies that if both parent solutions agree on a certain gene (either having or not having a TPS at a certain location) then the child solution will inherit this property. If the parents have a contradictory gene – one having a time point at a certain location and the other not – then the child has a probability of 50% to have a TPS there. In addition, 20% of the genes are the result of mutation that flips that gene of the parent solution. The stopping criterion was based on the number of generations with the algorithm terminating after the 25th iteration.
The genetic algorithm was also applied for the multi-objective problem. Here the objective function is a vector defined as: \( F(\bar{x}) = \{f_1(\bar{x}), f_2(\bar{x})\} \). The Pareto front describes the set of solutions that are not dominated by any other solution - non-inferior solutions. The procedure of the genetic algorithm is essentially the same as for the single objective function with the evaluation step calculating also \( f_2(\bar{x}) \).

6.4.4.2 Results and Discussion
The results of the genetic algorithm were first analyzed by examining the objective function value across generations (Figure 6.29). The overall trend is that the average and the standard deviation values decrease as the algorithm progresses until the 15th generation. During this phase the crossover plays the dominant role and leads to a more homogeneity gene pool with low objective function values. Later on the mutation procedure starts to be the main force behind the variability in the gene pool, producing new genotypes that push the algorithm for further exploration.
The evolution of the waiting time does not show a clear trend, albeit there is a slight decrease until the 11th generation (Figure 6.30). In contrast, the holding time shows a dramatic decrease of both the average and the standard deviation values until the 13th generation, where the algorithm starts to explore new solution spaces via mutations (Figure 6.31). Hence, the improvement in the objective function value is primarily due to obtaining TPL that reduce the holding times imposed on passengers on-board. The reduction in waiting times of passengers further downstream plays only a secondarily role.
Since the genetic algorithm does not reach a unique solution, the analysis of the emerging TPL is probabilistic. Figure 6.32 presents the share of individuals that have a certain stop as a TPS at different generations (initial population, $5^{th}$, $10^{th}$ and $15^{th}$). In the initial population each stop has the same probability to become a TPS. As the algorithm progresses, some stops have increasingly higher probability to be selected over other stops (e.g. 6, 7, 13, 15). Interestingly, the set of individuals that obtained the
minimum objective function values all have one to three TPS. The highest probability for having a TPS is found at stops 13 to 15, followed by a second TPS at either stop 6 or 7 and a third TPS at stop 21. This suggests that it is important that the TPL will be spread evenly over the route.

![Graph showing frequency of time point locations at various generations.](image)

Figure 6.32: Frequency of time point locations at various generations

When the genetic algorithm was run with the multi-objective functions, the algorithm evolves towards generating individuals that are closer to the Pareto front. Figure 6.33 presents the population of the initial, 3\textsuperscript{rd}, 6\textsuperscript{th}, 9\textsuperscript{th}, 12\textsuperscript{th} and 15\textsuperscript{th} generations. There is a clear decreasing trend in both objective function values as the algorithm progresses. The final Pareto front contains only two individuals which correspond to the same solution with the 13\textsuperscript{th} stop being a TPS. It is remarkable that both objective functions are targeted simultaneously and no clear trade-off is apparent in the form of a non-inferior front. Instead, the minimum of one objective function coincides with the minimum obtained for the counterpart function.
Figure 6.33: Individuals obtained by the multi-objective genetic algorithm by various generations

6.4.5 CONCLUSIONS

Solutions comparison

The results of the greedy and genetic algorithms (the best solution of both the single and multi objective functions and an additional prominent solution from the last generation) were compared with the no control scenario and the TPL that is currently in place. Table 6.7 presents the four passenger time components as well as the two objective function values. The comparison of the different solutions reveals that the current TPL results in long holding times which deteriorate the performance from both passenger and operator perspectives. In fact, its performance is equivalent to the no control scenario in terms of passenger-times and even worse in terms of fleet considerations.

The solutions of both the greedy and the genetic algorithms reduce the generalized passenger cost function by 11% compared with the current TPL - mainly due to shorter passenger-holding times (Figure 6.34). In addition to passenger time savings, a more regular service is expected to be advantageous also in terms of a more even passenger load on-board. The crowding measure of average standing time per passenger was used (Equation 6.9). It is evident that even though the optimization process did not account for passenger comfort explicitly, the solutions of the greedy and genetic algorithms reduced the average standing time substantially. This is expected,
because of the underlying relation between the coefficient of variation of the headway and the distribution of passenger loads. Our results are in line with the more uniform passenger loads reported by Delgado et al. (2009).

Table 6.7: Summary of results for selected solutions

<table>
<thead>
<tr>
<th>Solution</th>
<th>Total waiting time (pass-hr)</th>
<th>Total riding time (pass-hr)</th>
<th>Total dwell time (pass-hr)</th>
<th>Total holding time (pass-hr)</th>
<th>$f_1(x)$ (pass-hr)</th>
<th>$f_2(x)$ (min)</th>
<th>Avg. standing time per pass. (sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No control</td>
<td>675</td>
<td>799</td>
<td>668</td>
<td>0</td>
<td>2816</td>
<td>35</td>
<td>103</td>
</tr>
<tr>
<td>Current (10,17,24)</td>
<td>531</td>
<td>754</td>
<td>634</td>
<td>231</td>
<td>2797</td>
<td>39</td>
<td>95</td>
</tr>
<tr>
<td>Greedy (1,7,13)</td>
<td>532</td>
<td>745</td>
<td>617</td>
<td>48</td>
<td>2498</td>
<td>35</td>
<td>89</td>
</tr>
<tr>
<td>Genetic1/M-Genetic (13)</td>
<td>526</td>
<td>740</td>
<td>623</td>
<td>48</td>
<td>2487</td>
<td>35</td>
<td>88</td>
</tr>
<tr>
<td>Genetic2 (6,14,21)</td>
<td>527</td>
<td>747</td>
<td>624</td>
<td>60</td>
<td>2515</td>
<td>36</td>
<td>94</td>
</tr>
</tbody>
</table>

Figure 6.34: Time components of selected solutions
**What is the main determinant - number or location?**

The analysis also considers the fundamental question: which factor is the main determinant of the holding control performance – the number or the location of the TPS? Our analysis suggests that the performance is more sensitive to the latter. All the individuals that were generated by the algorithm during all the generations were grouped by the number of TPS to compare their performance distribution (Figure 6.35). The histograms follow a similar pattern with a large variation over the range of performance under various TPL. Nevertheless, fewer TPS have a noticeable higher frequency at the lower range of the spectrum. A higher number of TPS shifts up the distribution. However, the conclusions that can be drawn from this analysis are limited as the sample is biased by the evolution of the algorithm.

![Histograms of the objective function value for different number of TP, based on all the individuals generated during the genetic algorithm](image)

Figure 6.35: Histograms of the objective function value for different number of TP, based on all the individuals generated during the genetic algorithm.

A more elaborative analysis has to consider the entire population of solutions with a given number of TPS. All possible solutions with a single (30 solutions) or a couple of time points \( \binom{33}{2} = 528 \) solutions) were simulated in BusMezzo and the corresponding objective functions were calculated. Figure 6.36 presents the distribution of the objective function for the case of one or two TPS. In addition, the value of the no control scenario is highlighted. Its performance is equivalent to the 80th percentile of the two TPS distribution. Moreover, the TPL that is the current practice is worse-off...
than 75% of the possible combinations of a couple of TPS. In other words, the operator has a good chance to reduce passenger-times by choosing two TPS randomly rather than keeping the current TPL. The results highlight the importance of selecting the correct TPS as the performance varies considerably among solutions with the same number of TPS. These conclusions are consistent with the analysis of Mazloumi et al. (2011).

![Figure 6.36: Histograms for all the feasible solutions of single (33) and double (528) time point layouts](image)

**Computational time**

The computational time is proportional to the number of solutions that have to be executed in BusMezzo. The running time of 50 replications needed for the evaluation of a single scenario was 25 seconds on a standard PC. The number of solutions that need to be assessed in the greedy algorithm is \( \sum_{k=1}^{K} (|S| - k + 1) \), where \( K \) is the number of iterations needed before the stopping criterion is fulfilled. In the case of the genetic algorithm the number of assessed solutions is the number of iterations times the population size. Our analysis required the execution of 127 and 320 scenarios for the greedy and genetic algorithms, respectively. The greedy algorithm was very efficient in this case as it yielded very similar results to the genetic algorithm. Nevertheless, this conclusion cannot be generalized based on this analysis as it may be subject the properties of the studied line or the specific parameters. The number of solutions
considered by the genetic algorithm is still a very small fraction of the universal set of solutions \( \sum_{i=1}^{\lfloor S \rfloor} \left( \lfloor S \rfloor \right) \). Note that the greedy algorithm is path-dependent with each step being highly dependent on the previous results of the stochastic model. In contrast, the genetic algorithm guarantees a higher level of robustness.

**Multi-objective**

The multi-objective optimization procedure takes into account the potentially contradictory perspectives of passengers and operators. The analysis revealed that not only is there no clear trade-off between the two objectives, but rather that they are virtually consistent. This can be further elaborated by examining the factors that compound them. The passenger-time objective function (Equation 6.13) is written again for convenience:

\[
 f_1(\bar{x}) = \sum_{i \in I} \left[ \gamma_{w} \cdot w_n(\bar{x}) + \gamma_{r} \cdot r_n(\bar{x}) + \gamma_{d} \cdot d_n(\bar{x}) + \gamma_{h} \cdot h_n(\bar{x}) \right]
\]

Total passenger waiting times are calculated over all passengers and given a Poisson arrival process it is a function of the average headway and its coefficient of variation. Only the latter varies between solutions. The variations of the second expression are due to traffic conditions rather than control strategies. The dwell time can be linearly approximated by the number of boarding passengers. The total passenger-dwell time penalizes crowded buses with long dwell times; hence more even headways will contribute to the minimization of this expression. The total-holding times has a more complex interrelation with service regularity.

The second objective, the 90th percentile of the total trip time, can be decomposed into the vehicle riding time, dwell time and holding time. Each one of these trip components contributes to the variation in the total trip time. The riding time between stops is in fact independent of the control strategy assuming that timetables are designed adequately. The variability of the total dwell time at the vehicle-level depends on the variability in the number of boarding passengers and therefore on headway regularity. Moreover, from the single vehicle perspective, the more irregular the service is the control strategy will enforce longer holding times that will have a higher variability. In conclusion, the value of both objective functions will decrease overall with more regular services that exercise more even headways. Therefore, TPL solutions that minimize the coefficient of variation of the headway are expected to
perform well on both objective functions. This was also the result in the case presented in Section 6.3.3.

6.5 Interaction of Holding Control Strategies and Boarding Regimes on a Common Corridor

Dwell times are one of the main sources of service uncertainty. As discussed on Section 3.3.2, the dwell time depends on the boarding procedure and door configuration. In particular, a payment method that allows passengers to board through all doors is expected to reduce the time spent at stops since boarding passengers are typically the dominant factor in the dwell time function. Moreover, the dwell time variability may decrease as the impact of variations in passenger volumes on the dwell times will diminish. The importance of reducing dwell time variability as part of the effort to improve service reliability was highlighted in the trial report of SL (2003). Hence, moving to boarding through all doors may reduce passenger travel times not only due to shorter dwell times but also because of shorter waiting times as the service becomes more regular.

The analysis of service regularity and holding strategies was until now limited to the case of a single line. However, in reality transit lines interact with each other, especially in the common case of a shared corridor. A certain share of passenger demand may be indifferent between overlapping lines and hence consider the service regularity in terms of the joint vehicle arrival pattern. Common lines may also have different characteristics in terms of frequency, vehicle capacity, boarding regime and control strategy. BusMezzo was used for analyzing the effects of boarding procedure on transit performance on a common corridor and its interaction with holding strategies.

The complete descriptions of data collection, scenario design and simulation results are available in West (2011). The analysis revealed the importance of accounting for the interrelated transit operation processes because they exercise a positive feedback loop. This highlights the potential application of a transit simulation tool that models transit operations dynamics and their interaction with passenger path choice. Passenger loads were very different from those suggested based on the static assignment approach towards the common-line problem, because the simulation captures the effect of joint regularity. Interestingly, the analysis suggested that service regularity and travel times on low-frequency lines would benefit from the introduction
of boarding from-all-doors on high-frequency lines that share the same corridor. The combination of headway-based holding strategy and boarding from-all-doors was especially effective in improving service regularity.
7. IMPACTS OF REAL-TIME INFORMATION

A second set of case studies is concerned with evaluating the impacts of real-time information (RTI) on traveler decisions. Such an analysis requires the dynamic modeling of transit operations and traveler decisions. This chapter opens with an overview of the potential gains from RTI and previous attempts to analyze its impacts. A framework for modeling RTI in BusMezzo is then presented in Section 7.2. The prediction of next vehicle arrival at stop emulates the common method for generating RTI. The impacts of RTI on traveler decisions are incorporated into traveler decision within the multi-agent transit simulation model. Sections 7.3 and 7.4 present case studies which demonstrate the capabilities of BusMezzo by studying the effects of RTI under service disruptions and various levels of RTI. This chapter concludes with suggestions for potential applications and future research direction.

7.1 POTENTIAL BENEFITS OF REAL-TIME INFORMATION PROVISION

Transit systems are increasingly equipped with information and communication technologies in order to improve passenger level of service and support fleet management (FTA, 2000). APTS such as AVL were first used for improving operations and management. Later on, they were also utilized to provide RTI to passengers (FHA and FTA, 2000). The impacts of RTI on traveler behavior in the context of car traffic have been studied extensively (for a review see Lappin and Bottom, 2001; Chorus et al., 2006).

In the context of transit systems, RTI can refer to the remaining time until the arrival of the next vehicle, information on service disruptions, crowding conditions or prescriptive journey planners. Previous studies that evaluated the expected benefits from the deployment of RTI systems concentrated on passengers' perception and

---

6The content of this section was included in the following publications: Cats O., Koutsopoulos H.N., Burghout W. and Toledo T. (2011). Evaluating the role of real-time transit information provision on dynamic passenger path choice. Transportation Research Record, In press and; Cats O., Burghout W., Toledo T. and Koutsopoulos H.N. Modeling real-time transit information and its impacts on travelers' decisions. Submitted to Transportation Research Record, 2012.
satisfaction. Several empirical studies (TCRP, 2003b; Mishalani et al., 2006; Dziekan and Kottenhoff, 2007) found that perceived waiting times decreased substantially after the deployment of RTI systems, although they were still overestimated by passengers. Other studies suggested that the deployment of RTI displays at stops leads to an increase in the overall level of satisfaction (Zhang et. al, 2008; Caulfield and O’Mahony, 2009). Moreover, a recent empirical study by Watkins et al. (2011) found that users of RTI from personal mobile devices experienced shorter actual waiting times. In addition to level-of-service impacts, RTI could be used as a demand management tool that enables to better utilize the capacity of the transit system. Moreover, the prediction of transit conditions can be used for estimating the downstream effects of various control strategies. This would enable a more proactive approach to operations and management.

The provision of RTI can support traveler path choice decisions and enables an adaptive behavior that may result in time savings. The impacts of providing RTI at stops and on-board regarding transfer options and the associated waiting times were analyzed by Nökel and Wekeck (2008, 2009). This information enables more informed transfer choices and alighting decisions. Based on numerical simulation analysis the authors concluded that the additional time savings generated by providing the information on-board are within the same order of magnitude as those obtained from providing RTI at stops. Coppola and Rosati (2009) developed a simulation model for evaluating RTI regarding expected arrival times and crowding conditions. They found that RTI on bus arrival led to an increase in passengers’ waiting time and a decrease in their in-vehicle time (IVT).

A review of transit assignment models by Liu et al. (2010) highlighted the importance of RTI modeling in the recent emergence of dynamic travel decision models. The analysis of RTI requires dynamic representation of both the supply and the demand aspects of transit systems. Transit service needs to be represented in terms of individual vehicle runs and a corresponding timetable. Time-dependent passenger demand needs to be represented and loaded on specific vehicles. The loading process has to include a path choice model that takes into account time-dependent properties such as expected waiting time, travel time and level of comfort. Hickman and Wilson (1995) developed an analytical framework for adaptive transit path choice model that enables to evaluate the influence of RTI regarding the next arrival of each bus line. They
considered different levels of RTI accuracy, assuming that passengers make the best use of the available information through a deterministic network loading model. Nuzzolo et al. (2001) proposed a stochastic dynamic path choice model that includes a learning mechanism for both en-route and day-to-day dynamics. This model was extended to include attributes that are either experienced by travelers, such as elapsed waiting time and crowding conditions, or anticipated by them based on their experience or RTI (Nuzzolo et al. 2011).

None of the evaluations carried out in previous studies considered the interaction between supply and demand dynamics at the network level. This study overcomes this drawback by analyzing RTI within a transit operations and assignment simulation model.

7.2 MODELING REAL-TIME INFORMATION IN BUSMEZZO

7.2.1 MODELING APPROACH
The modeling of RTI has to include the dynamic representation of the elements that produce instantaneous data that is processed in order to generate predictions on transit conditions further downstream. In addition, the influence of the disseminated RTI on travelers' decisions has to be represented dynamically. Figure 7.1 presents the framework for RTI modeling in BusMezzo. It shows inputs (parallelograms), data processing (ovals), models (rectangles) and outputs (rounded rectangles). The core of the RTI modeling (dashed frame) consists of three modules: transit operations (supply side - in blue), the generation and processing of RTI (in purple) and travelers' path decisions (demand side – in orange).
Figure 7.1: Framework for real-time information modeling in BusMezzo

The simulation of traffic dynamics and transit operations in BusMezzo emulates transit performance and the production of automated data collection (ADC) methods, such as AVL and APC. This data can be processed in order to generate predictions on future transit conditions that will be disseminated to travelers. These predictions can also be utilized when applying control strategies that will in turn affect future transit performance. The dissemination of RTI may influence travelers' decisions and ultimately passenger flows. Passengers’ progression in the network is dictated by the decisions they make in reaction to transit conditions, such as transit vehicle arrivals at stops, their capacity constraints and the consideration of elapsed and anticipated waiting times. At the same time, travelers’ decisions affect transit performance through the effect of passenger flows on crowding, dwell times and their secondary implications on service reliability.

Note that individual passenger cars, transit vehicles and travelers are generated and progressed in the simulation model simultaneously. Hence, the simulation model is not a sequential process of generation and dissemination of RTI as both travelers and
operators can adapt their behaviors. These adaptations effect the performance of the transit system and hence also on the information that is generated later on.

7.2.2 GENERATION OF REAL-TIME INFORMATION

7.2.2.1 Related work

ADC data facilitates the generation of real-time predictions on transit performance. These predictions could be used by both travelers and transit operators. RTI can be disseminated to passengers in various forms (descriptive, quantitative or prescriptive) and channels (fixed stop screens, on-board displays or on-line applications). ADC and RTI can also support the design of more elaborative real-time control strategies, such as conditional transit signal priority, holding strategies, short-turning and expressing. The purpose of these strategies is to improve transit performance by direct intervention. In an analysis of control strategies to improve service coordination at transfer hubs, Dessouky et al. (2003) concluded that the most effective strategy required data on real-time transit conditions for all relevant vehicles. Furthermore, Shalaby and Farhan (2004) showed that predictions of the downstream effects of various control strategies could enable a more proactive approach to operations and management (e.g. assessing the impacts of a holding strategy on predicted downstream trajectories).

The prediction of the remaining travel time until arriving at downstream stops can be based on historical data or real-time AVL data. The latter can potentially result in more accurate estimation of traffic conditions. Previous studies applied various methods for bus arrival predictions as regression models, artificial neural networks (ANN), Kalman filter, support vector machines and statistical pattern recognition (Cathey and Dailey, 2003; Chen et al. 2004; Shalaby and Farhan, 2004; Jeong and Rilett, 2005; Yu and Yang, 2009; Padmanaban et al., 2010; Vu and Khan, 2010). These methods follow the general prescription proposed by Cathey and Dailey (2003). Typical inputs to these models are historical travel times, scheduled arrival times and AVL data for the last few vehicles. Table 7.1 summarizes the prediction methods that were applied in previous studies and the data sources that were incorporated. Most prediction models do not include dwell times explicitly in their arrival time prediction, even though they are an important source of travel time variability. Dwell time predictions can be enhanced by predicting downstream passenger flows based on recent APC records and anticipated headways. In addition, APC data could be utilized for generating RTI on anticipated crowding conditions at downstream stops.
### Table 7.1: Summary of previous studies on bus prediction models

<table>
<thead>
<tr>
<th>Authors</th>
<th>Prediction methods</th>
<th>Inputs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chen et al. (2004)</td>
<td>Artificial neural networks adjusted by Kalman filter</td>
<td>Historical data clustered by day of week, time of day, weather and stretch; Running time of the current bus on the last stretch.</td>
</tr>
<tr>
<td>Shalaby and Farhan (2004)</td>
<td>Kalman filter</td>
<td>Historical running times and passenger generation rates; Running time of the previous bus on the next stretch.</td>
</tr>
<tr>
<td>Jeong and Rilett (2005)</td>
<td>Historical, Regression, Artificial neural networks</td>
<td>Historical arrival times and dwell times; Schedule adherence at the last observation of the current bus.</td>
</tr>
<tr>
<td>Yu and Yang (2009)</td>
<td>Support vector machine learning method</td>
<td>Historical running times, speeds of the current and previous buses on the current and previous stretches</td>
</tr>
<tr>
<td>Padmanaban et al. (2010)</td>
<td>Kalman filter</td>
<td>Running times of the current bus and the two preceding buses</td>
</tr>
<tr>
<td>Vu and Khan (2010)</td>
<td>Statistical pattern recognition</td>
<td>Historical running times; Running times of current and previous buses on the last stretch.</td>
</tr>
<tr>
<td>Baptista et al. (2011)</td>
<td>K-nearest neighbors</td>
<td>Historical running and dwell times</td>
</tr>
</tbody>
</table>

#### 7.2.2.2 Real-time information generator in BusMezzo

The approach adopted in the current implementation of BusMezzo is to imitate these approaches and generate RTI based on historical data as expressed in timetables and the real-time AVL data of the current vehicle. Figure 7.2 presents the process of generating RTI for the remaining time until the next vehicle arrival on line l at a given stop s. The procedure is based on searching for the next trip (k) that is expected to visit
the stop and use its arrival time at the last recorded location \((m)\). The RTI generator restores the last recorded visit and checks whether it is a boundary case (first or last trip). If the next vehicle has already started on the next trip that will visit the stop then the expected arriving time is:

\[
EAT_{s}^{k,v} = AT_{m}^{k,v} + (ST_{s}^{k,v} - ST_{m}^{k,v}) \tag{7.1}
\]

Where \(AT_{m}^{k,v}\) and \(ST_{s}^{k,v}\) are the actual arrival time and scheduled arrival time of vehicle \(v\) at location \(m\) on trip \(k\), respectively.

This prediction assumes that future arrival times will maintain the current deviation from the schedule. However, if the next vehicle did not start trip \(k\) yet, the vehicle may recover from the deviation before starting the next trip. In order to recover, the deviation from the schedule needs to be shorter than the layover time at terminal \(n\). Otherwise, the expected departure time from the terminal is based on the minimal allowable layover time. This can be formally expressed as:

\[
EAT_{s}^{k,v} = \begin{cases} 
ST_{s}^{k,v} + EAT_{n}^{k-1,v} - \text{MAXAT}_{n}^{k-1,v} & \text{if } EAT_{n}^{k-1,v} > \text{MAXAT}_{n}^{k-1,1} \\
ST_{s}^{k,v} & \text{otherwise}
\end{cases} \tag{7.2}
\]

\[
EAT_{n}^{k-1,v} = AT_{m}^{k-1,v} + (ST_{n}^{k-1,v} - ST_{m}^{k-1,v}) \tag{7.3}
\]

\[
\text{MAXAT}_{n}^{k-1,v} = ST_{n}^{k,v} - \text{MINRT}_{n}^{k-1,v} \tag{7.4}
\]

Where \(\text{MINRT}_{n}^{k-1,v}\) is the minimal recovery time needed at terminal \(n\) for vehicle \(v\) after completing trip \(k - 1\).

This scheme is aimed to replicate the method that is commonly used by transit agencies for generating real-time arrival information (TCRP, 2003b). Moreover, it improves the predictions for vehicles that are still running on earlier trips on their schedule. Nevertheless, arrival predictions based on this scheme may be inaccurate as it is assumed that the remaining travel time is the same as the historical average. Therefore, it could be presumed that predictions become more accurate as the distance to the downstream location decreases.
7.2.3 Real-time Information and Traveler Decisions

The RTI that is disseminated to travelers affects their perceptions of system conditions and hence has the potential to influence their travel decisions. The dynamic path choice model (DPCM) in BusMezzo considers each passenger as an adaptive decision maker and represents its progress in the transit system. This is a multi-agent approach for modeling complex systems by representing the strategies of individual agents. Dia (2002) discussed the feasibility of an agent-based approach for modeling the impacts of advances travel information systems (ATIS) on driver behavior. Each travel decision depends on traveler’s current expectations with respect to future travel attributes such
as waiting times and travel times. It is assumed that in the absence of RTI, travelers’ expectations in the context of an urban transit system rely on their prior-knowledge of the transit network, which is derived from the timetable travel times and planned headways.

The information that is available to a traveler when making a certain decision is determined by the dissemination means and their locations as well as by individual characteristics (i.e. prior-knowledge and experience, availability of personal mobile device). BusMezzo was designed to enable the modeling of different levels of information at different decision stages. The spatial distribution of information display can be modeled by indicating the type and comprehensiveness given at different stops and vehicles (e.g. stops with no information vs. hub stations with complete information). This is incorporated into traveler’s perception and hence determines the assessment of potential alternatives.

The increasing availability of personal mobile devices such as smart phones provides travelers with instantaneous access to RTI during their entire trip. The share of individuals that have access to RTI by using a personal mobile device can be specified in BusMezzo. For an individual with RTI access, the individual level of information overrides the level of information determined by the public information display available at the network location. An additional source of information that does not require RTI is the elapsed waiting time at stop. In case that a traveler experiences waiting time that exceeds the anticipated waiting time by a certain threshold, the connection decision is reconsidered and the traveler may choose to walk to another stop.

The information disseminated to passengers varies among transit systems. It can be classified based on the following aspects:

- **Type** – waiting times (planned headways, static timetables or real-time predictions); in-vehicle travel times (static timetables or real time predictions); comfort (historical or real-time crowding levels) and; disruptions (descriptive, predictive, prescriptive).
- **Location** – pre-trip, at stops and on-board
- **Comprehensiveness** – with respect to the local stop, cluster of connected stops (i.e. transit hub) or the entire system
The exact combination of these aspects determines the level of information that is available to passengers at each trip stage and the respective decision. Following previous studies (Hickman and Wilson, 1995; Coppola and Rosati, 2009; Nökel and Wekeck, 2009), it is assumed that passengers perceive RTI as credible and incorporate it into their adaptive decision process. This assumption is supported by several empirical studies that reported that 70-100% of passengers use RTI as a source of information when available (TCRP, 2003b). Hence, the value of anticipated attributes such as waiting time, IVT and level of crowding takes the value from the highest level of information available. For example, let us consider a traveler that does not have a personal mobile device with access to RTI. Hence, the walking decision from the origin to the first transit stop is based on the traveler’s prior-knowledge. If the traveler arrives at a stop with real-time arrival information on the local stop then the traveler’s boarding decision relies on the RTI waiting times, while the remaining travel attributes are based on prior-knowledge. In case there is no RTI on-board, then the alighting decision relies entirely on prior-knowledge. When the traveler alights at a certain stop, a connection decision takes place. If the transfer stop is equipped with RTI display that covers also nearby stops then the immediate waiting time component in this decision is based on RTI.

The DPCM in BusMezzo also includes the potential reconsideration of traveler decisions due to information that becomes available (Figure 4.2 in Section 4.3). Each time RTI becomes available (e.g. traveler arrived at stop, boarded a vehicle, new piece of information) and is significantly different from traveler’s previous expectations then his/her perceptions needs to be updated and may be followed by a new decision. For example, a traveler that had no access to RTI at the origin and arrived at a stop with RTI may update the anticipated waiting time if it is very different from the prior-expectations. This in turn may lead to choosing another origin stop, based on the information that is currently available to the individual. In the future, the model could be enhanced by considering the process of information acquisition as travelers may initiate an inquiry on their personal device.
7.3 Case study I: Stockholm Metro Network

7.3.1 Description and design

7.3.1.1 Experiment description

As a proof of concept, the dynamic transit operations and assignment model is applied to the Stockholm Metro network. This network consists of seven routes clustered into three main lines identified by their color: blue (T10-T11), red (T13-T14) and green (T17-T19), as shown in Figure 7.3. Cervero (1995) gives an overview of the transit commuting patterns of the 'Stockholm's rail-served satellites' along the metro lines. The complete network was coded into BusMezzo, with the real-world time tables and walking distances between platforms. The network consists of 210 platforms situated at 100 stations. The Metro operates high-frequency service, with scheduled headways of 5 minutes on each branch. Passengers are assumed to plan their trip without considering the timetables, implying a Poisson arrival process at the origin stop. However, Metro dispatching is regulated based on the timetable. A schedule-based holding control is applied at all stops with the scheduled departure time as the earliest exit time. For a discussion of holding strategies in the context of urban rail operations see Koutsopoulos and Wang (2007).

The choice-set generation model (CSGM) was used as a pre-process step to the simulation runs. It resulted in 14,699 alternative paths for the entire network. The execution time was 3 minutes and 10 seconds on a standard PC.
7.3.1.2 Scenarios Design
The case study considers passenger information systems with RTI regarding the next vehicle arrival time at various levels of coverage and comprehensiveness. The following levels of information comprehensiveness were implemented and simulated in BusMezzo:
- **No RTI**: Passengers have no access to RTI, all travel decisions rely on prior knowledge.
- **Platform/Stop RTI**: RTI is available at stops and rail platforms regarding all transit services departing from the specific location.
- **Station/Cluster RTI**: RTI is available at stops and rail stations regarding all transit services departing from all platforms and bus stops within a single station/hub or a walking distance of up to 500 meters.
- **Network RTI**: RTI is available regarding all transit services in the network to all individuals through personal mobile devices.

In Stockholm Metro network, stops consist of separate platforms for each main line. Platform-level RTI is available at all platforms. Therefore RTI at the platform-level is regarded as our base case scenario. Furthermore, all routes of the same main line have a common platform. Therefore, choosing a platform stop is equivalent to choosing a metro line. RTI at the platform-level can influence passenger boarding decisions only when travelling to a stop that is not served by all line branches. Otherwise, there are no path-choice implications to providing RTI at the platform-level, as passengers are indifferent between different routes of the same main line. In addition, it is not realistic that passengers will choose to walk to a nearby stop in case of information on long waiting times as headways are short relative to walking times. In contrast, providing RTI at the station-level can be utilized for choosing another platform in the connection decision. Providing real-time arrival and riding times at the network-level may influence passenger path decisions at all stages, including alighting decisions.

In addition, the case study includes three operational conditions. The base case scenario of normal operations and two possible disruptions were examined (see Figure 7.3):

- **Regular operations (R)** - Normal operating conditions with real-world travel times and timetables.
- **Disruption, riding times (DR)** - A 15 minute delay in riding time on the Blue Line from Fridhemsplan to T-Centralen.
- **Disruption, frequency (DF)** - A reduction in frequency on the Green Line from a total of 18 vehicles per hour to 6.
RTI provision can be particularly advantageous in the case of service disruptions, causing longer than expected riding or waiting times. These kinds of disruptions may be caused by mechanical, operational, or technical problems. For example, the Stockholm Metro was subject to major service disruptions during the winter of 2010 due to frozen tracks caused by extreme weather conditions.

The experimental design consists of three levels of RTI provision (1- platform; 2- stop; 3- network) and three network operational conditions (R- regular; DR- riding time disruption; DF- frequency disruption) resulting in nine scenarios (Table 7.2). For each scenario, 10 simulation runs were conducted for a three hour period with uniform passenger demand. The number of replications was found to be sufficient with an allowable error of 2% for the average passenger travel time, which is the outcome of interactions between all random processes in the system.

<table>
<thead>
<tr>
<th>Operation conditions / Level of RTI provision</th>
<th>Regular conditions (R)</th>
<th>Disruption at green line service frequency (DF)</th>
<th>Disruption at blue line riding times (DR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Without RTI (1)</td>
<td>R1</td>
<td>DF1</td>
<td>DR1</td>
</tr>
<tr>
<td>RTI for all platforms at the same stop (2)</td>
<td>R2</td>
<td>DF2</td>
<td>DR2</td>
</tr>
<tr>
<td>RTI for the whole network (3)</td>
<td>R3</td>
<td>DF3</td>
<td>DR3</td>
</tr>
</tbody>
</table>

### 7.3.1.3 Choice Specification

The utility function for path alternative $i$ when evaluated by traveler $n$ at time $t$ took the following form:

$$U_{i,n}(t) = \beta_i^{\text{wait}} TAWT_{i,n}(t) + \beta_i^{\text{IVT}} TAVT_{i,n}(t) + \beta_i^{\text{walk}} TACT_{i,n} + \beta_i^{\text{trans}} TRANS_i + \varepsilon_{i,n}$$ (7.5)

Where $TAWT_{i,n}(t)$ and $TAVT_{i,n}(t)$ are the time-dependent anticipated waiting time and IVT, respectively. $TACT_{i,n}$ is the expected walking/connection time and $TRANS_i$ is the number of transfers involved with the path alternative. $\beta$’s are the corresponding coefficients and $\varepsilon$ is a generic error term. Coefficient values were derived from the estimated values of the path utility function (see Section 5.4). Each individual is
assigned with coefficients sampled from a normal distribution to account for the heterogeneity of preferences in the population. Trip fare is fixed for the entire network and hence does not affect passenger path decisions. Walking speeds are sampled from a truncated normal distribution with a mean and standard deviation of 4 km/hr and 1 km/hr, respectively.

The different information scenarios may affect passenger expectations regarding waiting times and IVT and therefore the disutility associated with an alternative. Note that the level of uncertainty involved with a path alternative is not included in the utility function. It is assumed that passengers regard the RTI as accurate and fully incorporate the RTI that is available at each decision point.

7.3.2 Results and Discussion
The results show that passenger stop and line choices are affected by the service disruption scenarios and incorporate the available level of RTI. The analysis of the results focused on the OD pair of Stadshagen (S) and Gamla Stan (G) (Figure 7.3). Focusing on a single OD pair enables a clear interpretation of the results.

There are no directs lines connecting stops (S) and (G). There are two possible transfer stops: Fridhemsplan (F) and T-Centralen (T). The CSGM has generated three alternative paths for this OD pair:

- Path A: Blue Line to (F) and transfer to the Green Line
- Path B: Blue Line to (T) and transfer to the Green Line
- Path C: Blue Line to (T) and transfer to the Red Line

The generated path alternatives were merged according to joint transfer stops and lines. The CSGM guarantees that each alternative path contains all the routes that utilize the same stretch. Paths that include more than a single transfer were eliminated during the CSGM since they were dominated by more attractive alternatives. The network configuration and the corresponding travel attributes associated with the relevant components are illustrated in Figure 7.4.
Figure 7.4: Network configuration and travel attributes of the relevant trip components

Table 7.3 summarizes the average total journey time and its components of IVT and out-of-vehicle time. Figure 7.5 presents the distribution of passengers between the three possible paths. Compared with the base case scenario of platform-level information, providing real-time arrival information on all platforms in a transfer stop can be beneficial in the case of (T). Moreover, providing real-time arrival and riding times for the whole network may influence both alighting decisions (T or F) and connection decisions (Red or Green Lines in the case of transferring at stop T).

In the base case scenario, 63% of the passengers choose to transfer at (T) since the riding time between (F) and (T) is three times longer on the Green Line than on the Blue Line. In addition, when transferring at (T), 55% of the passengers transfer to the Green Line (path B) due to the higher frequency and slightly shorter walking distances between platforms. It should be noted that the multinomial Logit model used in this
application may overestimate the probability of transferring at (T) due to its well-known IIA characteristic. More general discrete choice models can account for correlations among alternatives, a problem that received only little attention in the transit path context (Schuessler and Axhausen, 2007).

Under normal operating conditions with platform-level RTI (R1), the average journey time is 18 minutes. Waiting times and walking times account for 49% of the total time. When station-level RTI was provided, the shares of passengers choosing to transfer at (F) and (T) did not change, as this information does not affect alighting decisions. However, passengers transferring at (T) utilized this information when choosing the platform and line that minimize their waiting time. This led to almost equal shares between paths B and C and 3% savings in total journey time. When passengers have access to network-level information, the share of passengers transferring at (F) decreased relatively to the other information scenarios. This is due to the effect of the information at the origin stop on the alighting stop decision. This shift resulted in time savings because of substantial reduction in the average time spent on-board. Nevertheless, it simultaneously implies longer out-of-vehicle time as transferring at (T) involves longer walking distances.

Table 7.3: Average passenger journey time components

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Total journey time [sec]</th>
<th>Change in total journey time</th>
<th>In-vehicle time [sec]</th>
<th>Out-of-vehicle time [sec]</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1</td>
<td>1081</td>
<td></td>
<td>554</td>
<td>527</td>
</tr>
<tr>
<td>R2</td>
<td>1046</td>
<td>-3.2%</td>
<td>557</td>
<td>489</td>
</tr>
<tr>
<td>R3</td>
<td>1035</td>
<td>-4.3%</td>
<td>538</td>
<td>497</td>
</tr>
<tr>
<td>DF1</td>
<td>1418</td>
<td></td>
<td>553</td>
<td>865</td>
</tr>
<tr>
<td>DF2</td>
<td>1293</td>
<td>-8.8%</td>
<td>545</td>
<td>748</td>
</tr>
<tr>
<td>DF3</td>
<td>1260</td>
<td>-11.1%</td>
<td>523</td>
<td>737</td>
</tr>
<tr>
<td>DR1</td>
<td>1771</td>
<td></td>
<td>1116</td>
<td>655</td>
</tr>
<tr>
<td>DR2</td>
<td>1733</td>
<td>-2.2%</td>
<td>1115</td>
<td>617</td>
</tr>
<tr>
<td>DR3</td>
<td>1603</td>
<td>-9.5%</td>
<td>1054</td>
<td>549</td>
</tr>
</tbody>
</table>
In the case that the frequency on the Green Line is sharply reduced due to service disruptions (DF), the total journey time obviously increases compared to the regular conditions scenario. This is due to longer waiting times, while IVT remain unchanged. This increase can be reduced by providing more comprehensive RTI to passengers. With station-level information, more passengers transferring at (T) choose to continue with the Red Line rather than the Green Line resulting in an increase of 44% in the market share of Path C compared with the platform-level RTI scenario. Note that the availability of station-level RTI can only influence passengers’ connection decision (to which stop to walk) and not at which stop to alight. When RTI regarding expected arrival times at downstream stops is available to passengers at the time they make the alighting decision, the share of Path A decreases compared to the platform-level scenario. More than 16% of the passengers that transferred at (F) choose to continue on the Blue Line to (T), where there are more attractive transfer alternatives. Compared with the base case, RTI yields in this case substantial time savings of 9% when provided at the station-level and 11% when it is provided at the network-level.

When a disruption that causes severe delays on the Blue Line occurs (DR), passengers experience longer travel times. This is mainly due to longer IVT. Waiting times also increase as delays are propagated and affect trip chaining and service regularity. Providing RTI at the station-level did not affect passengers’ decisions in this case, as service disruption almost did not influence arriving times at (T). However,
When passengers were informed about the expected delay, they shifted dramatically to (F) – an increase of 30% in the market share of path A to avoid the disrupted service segment. Information provision results in this case in a reduction of over 9% in average passenger journey time compared to the base case.

When lacking RTI, passengers carry out decisions based on their prior knowledge. Therefore, path market shares are almost the same for all base case scenarios, regardless of operational conditions. Under all operational conditions, RTI provision affected passenger path decisions and resulted in substantial time savings. Note that due to the utility function specification, the uncertainty imposed by service disruption is not taken into account. Presumably, route choice shifts could be more dramatic if service disruption involves high uncertainty levels. As expected, increasing level of RTI comprehensiveness leads to increased time savings. However, the marginal benefit from providing additional RTI depends on network configuration and service conditions. These factors determine the importance of more informed platform choice and alighting decisions.

The previous scenarios are based on the real-world timetable. However an investigation of the Blue and Green lines schedules at (F) has indicated that there is a potential for improving transfer coordination. The coordination was based on setting the scheduled time of the Green Line at (F) to follow the scheduled time of the Blue Line. The exact coordination was calculated according to the 80th percentile of the distribution of walking time between the two platforms. The departure times of the Green Line trains from all other stops were shifted accordingly, basically implying slightly shifted dispatching times from the origin terminal. In order to benefit from this coordination, passengers have to be informed with RTI regarding downstream transfer stops. This scenario was simulated under regular operational conditions, when the effect of scheduled coordination can be evaluated.

The results indicate that the transfer coordination improvement can lead to substantial journey time benefits. The shifted dispatching times of the Green Line resulted in a decrease of 18% in average waiting time compared with the base case under the same level of information provision. However, the overall time savings are only 5% because more passengers transferred at (F) - taking a slower line. But by doing so they reduced their waiting and walking times. Perhaps some of the time savings
yielded from coordination could be achieved even in the absence of RTI by communicating it to travelers and day-to-day learning.

The dynamic transit assignment enables us to analyze how passenger path choice evolves during the simulation time and the varying operational conditions in the transit network. For example, similar aggregated market shares can be obtained from various time-dependent load performances. Figure 7.6 presents the coefficient of variation of boarding passengers at stop (F) on the Green Line. The variation was calculated over all trips of Green Line routes. Note that the coefficient of variation varies between scenarios that yielded the same market shares (see Figure 7.5). The variability is virtually the same for platform-level and stop-level scenarios with the same operation conditions, since both levels do not influence passenger alighting decision. In contrast, network-level RTI is associated with higher variability in passenger loads. Since passengers incorporate time-dependent information in their decision process, passenger activity has a more uneven pattern. The case of scenario DR3 is different since the main function of RTI is to inform passengers about the exceptional riding times on the Blue Line. The large numbers of passengers choosing to alight at (F) are the reason for the decrease in the coefficient of variation indicator. Similarly, the lower levels of variability under the DF scenarios are due to the accumulation of passengers during the long service headways. The coordination scenario described above (noted by C3) reduced the variability of boarding volumes compared with the base case. With the current timetable, RTI at the network-level leads to more uneven passenger patterns at (F) depending on the respective expected arrival time. However, when transfer at (F) is coordinated, passengers consistently choose to transfer there when this information is available.
7.4 CASE STUDY II: STOCKHOLM RAPID TRANSIT NETWORK

7.4.1 DESCRIPTION AND DESIGN

The previous case study of Stockholm Metro system investigated the potential time savings from providing RTI on service disruptions. The analysis of RTI impacts was limited to a single OD pair. A follow-up case study extended the analysis to the network-wide impacts of Stockholm rapid transit system under regular operational conditions. The simulation model was applied to the Stockholm’s rapid transit system which consists of: seven Metro line branches (lines 10-11; 13-14; 17-19), four high-demand trunk bus lines (lines 1 to 4) and a LRT line (Tvärbanan, line 22). The complete network of these lines was coded in BusMezzo with the real-world timetables, vehicle schedules and walking distances between platforms and stops. The network is shown in Figure 7.7. During the morning peak-period (6:00-9:00) there are approximately 700 service trips carried out by over 200 vehicles. Since the peak-period headway of each of the transit lines is in the range of 5 to 10 minutes, all travelers are assumed to arrive randomly at stops. Passengers have prior knowledge of planned headways and timetable IVT. The three different transit modes have different vehicle types, operating speeds, travel time variability and are operated with different holding control strategies. Each mode was assigned with its own dwell time function structure obtained from (TCRP, 2003a) with the respective boarding regime and number of doors. Unlike
the super-network approach of Carlier et al. (2003), there is no need for artificial links for intra-modal or inter-modal transfers.

Figure 7.7: Stockholm’s rapid network as displayed by BusMezzo

To allow a warming up period of transit supply, passenger demand was simulated only for the peak hour (7:30-8:30 am). There are approximately 150,000 passenger trips that are initiated during the peak hour. The passenger demand was extracted from data collected at entrance barriers at Metro stations, passenger counts at transfer locations and LRT stations (SL, 2009) and automatic passenger counts on trunk buses. Using these data, the passenger stop-to-stop OD matrix was obtained by applying a proportional trip distribution procedure.

For this network, the model generated 99,270 alternative paths that are used in the construction of choice-sets throughout the simulation. The utility function for path
alternatives remained as given by Equation 7.5. Trip fare is fixed for the extended network and hence does not affect passenger path decisions.

The case study considers passenger information systems with the same four scenarios as in the previous case study (Section 7.3): No-RTI, Stop-RTI, Cluster-RTI and Network-RTI. The different scenarios imply the availability of RTI at different trip stages. Stop-RTI supports boarding decisions. Cluster-RTI provides information that can also be used in making connection decisions between stops. In addition, the network-level RTI may be used pre-trip (initial connection) and in making alighting decisions.

For each scenario, 10 simulation runs were conducted for a three hours period with a uniform passenger demand during the peak-hour. This number of replications yielded a maximum allowable error of less than 1% for the average passenger travel time. The execution time for a single run was less than 1 minute on a standard PC.

7.4.2 RESULTS AND DISCUSSION
Table 7.4 presents the average travel times and their components in the different scenarios. The results presented are for the entire demand in the network, and only for trips that have both origin and destination within the inner city boundaries (about half of the total demand). In the no-RTI scenario, the average passenger journey time is almost 39 minutes for the entire network and 23.3 minutes for inner city trips. The waiting times, which include waiting times at the origin stop and at transfer stops, account for 26% of the total travel time in the inner city. Walking times refer to walking between stops either when making a transfer or in order to reach a nearby stop in the beginning or the end of passenger's trip. Walking times account for 11% of the travel time in the inner city. In the other scenarios, the more informed passengers are, the more their journey time decreases. The travel time in the network-RTI case is 4.5% lower compared with the no-RTI case. This trend is expected as RTI enables passengers to make more informed decisions and improve the coordination of downstream transfers. The reduction is largest in waiting times, which decreased by up to 7.5%. In contrast, walking times did not decrease and even increased by up to 1.9%. IVT decrease only in the case of network-RTI when passengers may choose to wait longer in order to take a path which is overall shorter. Interestingly, also the number of boardings per trip decreases from 1.60 to 1.45. This decrease is obtained already when RTI is provided at the stop level, suggesting that uninformed travelers make more unnecessary transfers instead of waiting longer for a direct service.
Table 7.4: Average passenger journey attributes (entire network / inner-city, upper line) and the relative change compared with the no-RTI scenario (in percentages)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>No-RTI (base)</td>
<td>2323/1399</td>
<td>1621/885</td>
<td>545/362</td>
<td>157/152</td>
<td>1.69/1.60</td>
<td>511</td>
</tr>
<tr>
<td>Stop-RTI</td>
<td>2317/1359</td>
<td>1619/876</td>
<td>540/333</td>
<td>158/149</td>
<td>1.61/1.44</td>
<td>529</td>
</tr>
<tr>
<td>Cluster-RTI</td>
<td>2312/1354</td>
<td>1616/880</td>
<td>536/321</td>
<td>160/152</td>
<td>1.60/1.42</td>
<td>544</td>
</tr>
<tr>
<td>Network-RTI</td>
<td>2265/1334</td>
<td>1583/849</td>
<td>522 /334</td>
<td>160/150</td>
<td>1.61/1.45</td>
<td>556</td>
</tr>
</tbody>
</table>

Providing RTI at each stop requires substantial investment in the installation of digital displays and communication systems. However, it may be possible to obtain a large share of the time savings by concentrating on the major transfer stops with a considerably smaller investment. The marginal gain from extending RTI at the stop-level to cover nearby stops in order to support connection decisions was found to be negligible for this network in the case of no service disruptions. The topology of Stockholm’s transit network plays an important role in limiting the opportunities to gain from more informed connection decisions. The network is designed so that there are only few cases where interchangeable rapid lines do not serve the same stop – interchangeable metro lines use the same platform and overlapping bus lines use the same stops.

The analysis suggests that providing network-level information through personal mobile devices can yield considerable time savings and substantially different time-dependent passenger loads. The additional information available in network-RTI yielded a considerable reduction in waiting times and total travel times. Access to such information supports more informed boarding and connection decisions as it addresses all downstream waiting times for path alternatives that involve transfers. Moreover, it enables passengers to make alighting decisions based on RTI at potential transfer locations. Developing applications to personal mobile devices is an inexpensive dissemination channel that can obtain substantial time savings. The rapidly increasing penetration rate of smartphones enable to provide RTI that can be customized to traveler needs and presented in an attractive way to a wide audience.
As passengers are more informed, passenger loads are subject to more fluctuation due to the adaptive passenger path choice process. When path shares and passenger loads are aggregated for the entire peak period, the aggregate values remain stable under different RTI provision scenarios. At the same time, passenger travel times are consistently shorter as passengers become more informed. A time-dependent analysis reveals that access to RTI affects passenger assignment results considerably. This is in line with results from an analytical study by Hickman and Wilson (1995).

In order to illustrate this important consequence of RTI provision, passenger loads on peak-hour trips of two lines are presented in Figure 7.8. Each curve corresponds to a chain of consecutive trips with scheduled departure times from the terminal between 7:30 and 8:00. The shadowing marks the transition between successive trips. It is evident that the passenger loads exhibit substantially different pattern under the no RTI and the network-level RTI scenarios. In the case of the LRT line 22 (upper figure), a more fluctuating pattern emerges in case of network-RTI compared with a clearer pattern for the no-RTI scenario. The higher fluctuation in passenger load under network-RTI is due to higher temporal variability as well as spatial variability. Sharp changes in passenger loads occur at transfer stops to metro lines. As passengers have access to RTI, the actual system conditions are subject to more fluctuations and passenger decisions deviate from expectations that are based on average values. The inner-city part of metro line 14 shows a similar trend with higher temporal variability and higher extreme passenger loads in the network-RTI case compared to the no-RTI case.

The fluctuations in passenger loads increase the probability of on-board and at-stops discomfort and capacity concerns. Moreover, it may negatively affect service regularity as uneven passenger loads contribute to the bunching phenomenon. This trend is reflected in the increase in the average standing time per passenger shown in Table 7.4, which is a measure of crowding that is used as a proxy for the level of comfort. It was calculated by the summation of Equation 6.9 over lines. More uneven loads on transit trips lead to an increase of 9% in the average standing time per passenger.
Figure 7.8: Passenger load on rush hour trips of the eastbound LRT line 22 (up) and southbound Metro line 14 (down, inner-city part)
7.5 CONCLUSIONS

The evaluation of RTI requires the dynamic modeling of transit supply and demand. A framework for modeling RTI was presented and implemented in BusMezzo, a transit simulation model. The simulation of individual transit vehicles and travelers enables to model the generation, dissemination and influence of RTI on passenger choices. Each traveling decision is based on the anticipated attributes of path alternatives. Travelers' anticipation depend on the information that is available to the passenger when making the decision either from location-based displays or individual access to RTI through personal mobile devices.

This model was used as an evaluation and analysis tool for case studies based on Stockholm network. The CSGM composed all reasonable paths and the DPCM processed passenger decisions under various operational conditions and RTI provision scenarios. The results indicate that providing more comprehensive RTI has the potential to lead to path choice shifts and time savings. The analysis also suggests that significant benefits can be achieved by simple improvements in transfer coordination.

The results from both case studies point out to substantial differences between the time-dependent passenger assignments at the individual vehicle run level under different RTI scenarios. Hence, it is important to model the level of information that is available to passengers at different stages of their trip in order to capture their path choice reactions. This is necessary in order to capture the implications of RTI on the time-dependent system conditions. Note that the simulation of a sub-network that consists only of rapid services hindered the potential adaptive reactions to RTI. Therefore, time savings and variations in path choices could be underestimated. The current analysis considered the provision of RTI with respect to transit arrival times. A future study may analyze the interdependency between passenger decisions and the potential impacts of RTI regarding on-board crowding conditions. A stated-preferences survey by Kim et al. (2009) suggested that this may have considerable affect on passenger route choice decisions.

The analysis of RTI impacts can be used as part of an economic assessment of RTI system installation. The evaluation can support decision makers in prioritizing locations and attributes of the displayed information since the coverage of this systems is typically limited to certain stations or services. Furthermore, the model can be used as a test-bed for various methods to generate RTI based on transit performance
predictions. The arrival prediction model used in this study could be enhanced to incorporate real-time predictions of downstream traffic conditions and passenger volumes.

Finally, behavioral and cognitive factors follow systematic patterns that influence passengers’ decision making process. A more elaborate model of the mental process involved with RTI provision may consider the expected costs (monetary, time, cognitive effort) and gains associated with acquiring more information (Richardson, 1982). This is of particular importance in the context of accessing RTI through personal mobile devices which depend on traveler’s initiative. Sun et al. (2009) proposed an information acquisition model that defines the expected value of information as the expected increase in the utility function value due to the additional information. The model is developed within the context of rescheduling decisions. Its approach is in line with BusMezzo as the sequence of decisions is represented in terms of a decision tree with uncertain outcomes that depends on individual’s decisions and system conditions.

Ben-Elia and Shiftan (2010) demonstrated how impacts of learning and experience that are derived from Prospect Theory and Reinforced Learning can be estimated within the framework of random utility models. Notwithstanding, the portrait of an economic rational decision maker could be enhanced by considering computational limitations, network-related knowledge and limited adaptation (e.g. Prato et al. 2011).
8. DISCUSSION

8.1 RESEARCH CONTRIBUTION

The aim of this thesis was to develop a tool for analyzing transit performance in the era of APTS. The dynamic transit operations and assignment model that was developed in this thesis is part of a new generation of transit models which consider the effects of APTS on transit operations and passenger decisions.

This thesis contributes to the domain of transit modeling in the following respects:

- Providing a framework for representing the interactions between the main components of the transit system – namely, traffic dynamics, transit operations and traveler behavior. This framework considers the evolution of the transit system as the result of inter-dependent dynamic processes that can adapt to changing system conditions.

- The implementation of the modeling framework in a mesoscopic traffic simulation model which enables to evaluate transit performance and level-of-service at the network-level. An agent-based approach was adapted for emulating how the system evolves over time.

- The development of a two-stage semi-compensatory dynamic transit path choice model. The dynamic path choice model (DPCM) represents traveler’s trip as a process of successive travel decisions. Each decision is concerned with choosing among alternative actions rather than path alternatives while taking into consideration the complete trip from traveler’s current location to the final destination. Hence, the model implies an adaptive and dynamic choice process without implying a myopic choice strategy.

- The model represents individual travelers and account for variation in preferences and perceptions. Each travel decision involves the evaluation of alternative actions based on the respective path alternatives and their anticipated downstream attributes. The anticipated values of path attributes depend on travelers’ prior-knowledge and real-time information provision. Travel decisions are modeled within the framework of discrete random utility models.
• A non-compensatory rule-based choice-set generation model (CSGM) for transit path alternatives was developed. The model generates a set of logical paths and then filters unattractive paths by applying a set of behavioral and dominancy rules at the single-path and the choice-set levels. Path alternatives are merged based on common lines and stops to avoid potential overlapping. The model results in a master-set that could be adapted based on time-dependent filtering rules.
• The estimation problem for the non-compensatory CSGM was formulated. The problem involves the estimation of activation and threshold parameters which correspond to filtering rules. The problem was formulated for the general case along with simplified versions.
• An analysis of choice-set composition properties and path choice decisions based on a web-based revealed- and stated-preferences survey that was designed and conducted as part of this thesis. The survey data was used for the estimation of the CSGM and the path alternative utility function.
• The development of a transit operations simulation model that captures the various sources of uncertainty including passenger arrival process, dwell times, capacity constraints, travel times and vehicle scheduling. A preliminary validation of the transit operations model was conducted by comparing simulation results with real-world data. The capabilities of the model were demonstrated in several case studies.
• The analysis of transit performance under various real-time control strategies from both passenger and operator perspective. The results of the case study, the analysis of AVL data and an investigation of the current operations practices were then synthesized into guidelines for a field trial of the proposed control strategy. An experiment that implemented the proposed strategy took place on October 2011.
• The formulation of the time point layout problem as a multi-objective optimization problem that simultaneously considers the number and location of time points. The optimization process consisted of iterative BusMezzo simulation runs which assessed the performance of candidate solutions.
• The development of a framework for analyzing real-time information provision which includes the generation of such information and the incorporation of
information into travelers’ decision process. The framework was implemented in BusMezzo. The dynamic operations and assignment model was applied for evaluating the impacts of various levels of real-time arrival information on passengers flows in the rapid transit network of Stockholm.

- BusMezzo is an open-source simulation model that is available upon request. It can either be used as an external part of a larger planning model or a standalone model for a joint traffic and transit operations analysis. Moreover, it has a modular structure that enables further developments and its use as a laboratory tool for testing various modeling features.

8.2 LIMITATIONS AND FURTHER RESEARCH

The transit analysis tool that was developed in this thesis can be used as part of a decision support system for transit operations and management. Such a system will simulate, analyze and evaluate alternative real-time based strategies (e.g. holding, vehicle allocation, scheduling) by considering their downstream effects. The system may assess alternative strategies with respect to various level-of-service aspects and subject to operational constraints. As was illustrated in the holding case study, the model can be used as an evaluation tool within an optimization process. For example, BusMezzo can assess alternative timetables, vehicle scheduling or control schemes as part of an interactive process of system design.

The transit simulation model is yet to be validated with a system-wide case study and real-world data. Automatic data collection (ADC) methods such as smartcard data, AVL and APC data can facilitate the estimation and validation of both transit operations and passenger decision modeling components. The integration and synthesis of these data sources can be further enhanced by data from personal mobile devices and travel journey inquiries.

The following outlines additional potential future research directions that would address the current limitations of the model:

- DPCM estimation – the decision making process needs to be estimated by considering the sequence of path decisions and the respective attributes of the alternative decisions. The current implementation of the DPCM suggests a highly-adaptive decision maker. The level of adaptation exercised by transit
users is an important behavioral modeling issue due to potentially important habitual patterns (Nagel and Marchal, 2003). An imperfect adaptation can be embedded into the DPCM in several ways. For example, a tendency to avoid deviations from previous decisions can be incorporated into the choice model by introducing an inertia factor or by determining a certain threshold for making a new decision.

- **Steady-state conditions** – currently, the model represents only within-day adaptation with a single-shot simulation run. Modeling the accumulated experience of travelers through iterative simulation runs would facilitate the specification of anticipated trip attributes’ values by carrying on previously experienced values. The day-to-day learning would result in steady-state conditions that could be equivalent under certain conditions to equilibrium conditions (Nagel and Marchal, 2003).

- **Anticipated values of trip attributes** - the evaluation of alternative actions could be enriched through the inclusion of additional path attributes in the utility function. Anticipated level of comfort and mode-specific coefficients are two potential factors to be included in the choice model specification. The expected values of quality-of-service measures as comfort and service reliability could be derived from their steady-state values based on a large number of simulation runs (e.g. the probability of denied boarding, schedule adherence distribution per line and stop).

- **Choice model structure** - the successive travel decision process could be enhanced by accommodating alternative random utility choice models. In particular, it is important to capture the correlations among path alternatives. For example, the cross-nested Logit (CNL) model can be applied in the transit path context. This implies the definition of path alternatives at the higher level and legs at the lower level where each leg can belong to more than a single path and each path includes several legs.

- **Mixed high and low frequency services** - the model was developed with the intention to study high-frequency service. However, most networks consist typically on a mixture of high and low frequency services. Moreover, trips are often composed of both services without a clear division. Hence, the model could be enhanced to represent a larger range of service frequencies by considering
various passenger arrival patterns and travelers' anticipated waiting time. Furthermore, the representation of low-frequency services may require the consideration of trip departure choice as an additional choice level.

- Improved multi-modality capabilities – These applies for both supply and demand aspects. On the supply side, the characteristics of various transit modes can be modeled in greater detail. These include the better representation of the level of interaction with car traffic and TPS schemes. Capturing the effects of TPS would also contribute to the design of more comprehensive and proactive control schemes. On the demand side, the model could be extended to consider multi-modal trips as park-and-ride and bike-and-ride. The joint car and transit model can facilitate in the future an additional modeling layer of mode choice which takes into account the individual agent characteristics instead of the current independent demand per mode.

In addition, the following are two potential directions in which the simulation model could be developed:

- Demand responsive transit (DRT) - DRT schemes are designed to provide an adaptive transit service based on real-time conditions. Currently, DRT systems are typically used in rural areas or for people with mobility difficulties. However, it may prove beneficial for both passengers and operators with the increasing use of APTS and personal mobile devices. The dynamic modeling of both transit supply and demand and the adaptive strategies that both operators and passengers can undertake in reaction to real-time conditions provides the foundations for considering DRT schemes. This includes dynamic scheduling and routing of transit services that employ proactive strategies to satisfy anticipated travel demand under a set of operational constraints.

- Freight modeling – the traffic simulation model could be extended to represent the operations of urban freight traffic. The modeling of freight operations would benefit from noticeable similarities that it has with transit operations – the importance of trip chaining, loading/unloading times and real-time control strategies. In particular, the evaluation of rerouting and rescheduling strategies is in line with DRT modeling.
REFERENCES


Fu, L. and X. Yang (2002). Design and implementation of bus-holding control strategies with real-time information. Transportation Research Record, 1791, 6-12.


Iseki H. and Taylor B.D. (2009). Not all transfers are created equal: Towards a framework relating transfer connectivity to travel behaviour. Transport Reviews, 29 (6), 777-800.


Munizaga M., Palma C. and Fischer D. (2011). Estimation of a disaggregate multimodal public transport OD matrix from passive smart card data from Santiago,
Chile. Proceedings of the 90th Transportation Research Board Annual Meeting, Washington DC.


PTV (2009). VISUM 11.0 basics. PTV AG, Karlsruhe, Germany.

Puong A. (2000). Dwell time model and analysis for the MBTA red line, MIT's Open Course Ware Project, course number 1.258, http://ocw.mit.edu/


dynamic schedule-based network. Transportation Research Part B, 33, 107-121.

University Press.

estimation in a transit smart card automated fare collection system. Journal of
Intelligent Transportation Systems, 11, 1-14.

headway control strategies for transit systems. Transportation Research Record, 746,
25-29.

Review, 79 (4), 281-299.

reliability measurement using automated fare card data. Application to the London
underground. Transportation Research Record, 2143, 92-99.


to improve reliability. Transportation Research Record, 2112, 70-76.

short headway transit services: schedule-based and headway-based holding strategies.
Proceedings of the 89th Transportation Research Board Annual Meeting, Washington
DC.


Vovsha P. and Bekhor S. (1998). Link-nested Logit model of route choice,
Overcoming route overlapping problem. Transportation Research Record, 1645, 133-
142.

Vu N.H. and Khan A.M. (2010). Bus running time prediction using a statistical
pattern recognition technique. Transportation Planning and Technology, 33 (7), 625-
642.


Weidmann U. (1994). Der Fahrgastwechsel im öffentlichen Personenverkehr. IVT No. 99, Zürich. [In German]


APPENDIX A: OBJECT-ORIENTED FRAMEWORK

The simulation model is developed with an object oriented programming (OOP) framework. The figure below is a simplified presentation of only the transit-related classes included in the simulation-model. The notations are based on unified modeling language (UML) which was designed to be compatible with object-oriented developments. The classes Vehicle, Action, Link and Route are part of the general traffic simulation. The classes that are necessary for dynamic route choice modeling are marked with a yellow background.

Figure A.1: Object-oriented framework for the transit simulation structure
APPENDIX B: INPUTS AND OUTPUTS OF THE SIMULATION MODEL

The input data includes information about the transit system and passenger demand. Table B.1 summarizes the necessary inputs for defining the transit system. It consists of three layers: fundamental network - the most basic layer of data that refers to the relevant network of roads and railways; transit network configuration - the second layer that includes network typology and timetables and; fleet assignment. The different levels of demand representation have different input requirements. A set of parameters has to be specified when performing dynamic loading. The input files are given in a text file format. A user manual to the input files format is updated on a regular basis in order to assist simulation users.

Table B.1: Transit-related input of BusMezzo

<table>
<thead>
<tr>
<th>Type</th>
<th>Information included</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fundamental network</td>
<td>• Nodes: coordinates, type (origin, destination or junction)</td>
</tr>
<tr>
<td></td>
<td>• Links: origin node, destination node, distance, speed-density function</td>
</tr>
<tr>
<td></td>
<td>• Turning movements: node, connecting links, stochastic server</td>
</tr>
<tr>
<td>Transit network configuration</td>
<td>• Stops: location, length, type, availability of real-time information and walking distances to nearby stops</td>
</tr>
<tr>
<td></td>
<td>• Routes: origin node, destination node and list of links</td>
</tr>
<tr>
<td></td>
<td>• Lines: route, list of stops, set of control stops, type and parameters of holding control strategies</td>
</tr>
<tr>
<td></td>
<td>• Trips: frequency or dispatching times, scheduled arrival times at stops</td>
</tr>
<tr>
<td>Fleet</td>
<td>• Vehicle types: length, number of seats, capacity, type and parameters of the dwell time function</td>
</tr>
<tr>
<td></td>
<td>• Vehicles: vehicle type and the chain of trips to be carried out</td>
</tr>
</tbody>
</table>
BusMezzo generates detailed output on transit performance, in addition to the general traffic output generated by the traffic simulation model. Table B.2 lists the transit output data that is generated by the simulation model and is available at various levels of aggregation. The raw data is based on two sources: an output record that is generated every time that a transit vehicle visits a stop and; a passenger decision report that is generated every time a passenger makes a travel decision.

Table B.2: Transit-related output of BusMezzo

<table>
<thead>
<tr>
<th>Type</th>
<th>Information included</th>
<th>Levels of aggregation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stop visit report</td>
<td>- Identification data: line, trip, vehicle, stop</td>
<td>Stop visit event</td>
</tr>
<tr>
<td></td>
<td>- Times: arrival time, scheduled time, dwell time, holding time, exit time, headway at departure and arrival.</td>
<td>Stop</td>
</tr>
<tr>
<td></td>
<td>- Passenger counts: number of boarding and alighting passengers, passenger load and the number of passengers left behind.</td>
<td>Line</td>
</tr>
<tr>
<td></td>
<td>- Performance measures: scheduled deviation, on-time performance (shares of early, on-time and late arrivals), the need to enforce capacity constraints, dwell time variability, headway variability, average and 90th percentile of the total trip time.</td>
<td>Network</td>
</tr>
<tr>
<td></td>
<td>- Total passenger-times: total passenger-waiting/riding/dwell/holding time, weighted passenger-time objective function.</td>
<td></td>
</tr>
</tbody>
</table>
As shown in the table above, the transit operational measures are aggregated at various levels to enable the evaluation and comparison of scenarios and strategies. Moreover, the simulation model can be used as part of an external optimization procedure. A general objective function that is a function of passenger-time components can be internally calculated by specifying the weights associated with the respective factors. The vehicle trajectory data is a useful tool for analyzing the way the system evolves over both time and space and detecting the bunching phenomenon. With respect to the passenger decisions database, an internal processing generates an output file that contains the complete description of chosen paths (stops and lines) as well as path choice summary files. This allows a straight-forward analysis and comparison of assignment results from different scenarios.

All the output files are given at a plain text format and can be copied directly to any data analysis software as Microsoft Excel or MATLAB for further analysis.
APPENDIX C: SCREENSHOTS FROM THE WEB-BASED SURVEY

The survey reported in Chapter 5 was constructed using the web-platform of “Qulatrics”. Below there are selected screenshots. The survey was conducted in Hebrew.

Figure C.1: Choice-set selection (“Please check all trip alternatives that you find reasonable when travelling from the Bay CBS to the Technion”)

Figure C.2: Choice-set rating (“Grade each of the following trip alternatives from the Bay CBS to the Technion”)

278
Figure C.3: Instructions on the dynamic path choice section
Figure C.4: An example of the format of the dynamic path choice questions which was given as part of the instructions preceding this section
Figure C.5: Transfer choice attributes ranking