

AVL-BASED TRANSIT OPERATIONS CONTROL

By

Aichong Sun

A Dissertation Submitted to the Faculty of the

DEPARTMENT OF CIVIL ENGINEERING & ENGINEERING MECHANICS

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

WITH A MAJOR IN CIVIL ENGINEERING

In the Graduate College

THE UNIVERSITY OF ARIZONA

2005

THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Dissertation Committee, we certify that we have read the dissertation
prepared by (Aichong Sun)
entitled (AVL-Based Transit Operations Control)
and recommend that it be accepted as fulfilling the dissertation requirement for the
Degree of (Doctor of Philosophy)

(Mark Hickman) Date: 02/14/2005

(Kevin Lansey) Date: 02/14/2005

(Simon Washington) Date: 02/14/2005

(Wei Lin) Date: 02/14/2005

(Pitu Mirchandani) Date: 02/14/2005

Final approval and acceptance of this dissertation is contingent upon the candidate's
submission of the final copies of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and
recommend that it be accepted as fulfilling the dissertation requirement.

Dissertation Director: (Mark Hickman) Date: 02/14/2005

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at The University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgment of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: _____

ACKNOWLEDGEMENTS

My sincere thanks and gratitude go to my supervisor, Dr. Mark Hickman, for his continued support, guidance and understanding, and for his encouragement and belief in my potential throughout my pursuit of Ph.D. I would also like to express my appreciation to other members of my committee, Dr. Lansey, Dr. Lin, Dr. Mirchandani, and Dr. Washington, for their valuable suggestions on my dissertation research.

I truly enjoyed the time to study and work with a number of graduate students, Zawwar, Alejandro, Yuhwa, Dogyeong, Chris, Ashish, Wen, Felipe and many others. Thank you for your invaluable friendship and support, and thank you for bringing a fresh perspective and a few laughs.

My deepest gratitude also goes to my parents. Without their dedication to my academic career, I can never have got so far. I also offer my warmest regards to my uncle Biao and my brother Aizhuang for their perennial friendship and support. Finally, My wife Aiying, you are the one of the utmost importance to me. My appreciation to you has gone far beyond any words of human being, thank you so much and so much for making me smile again and bringing sunshine back to my life.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS	9
LIST OF TABLES	10
ABSTRACT	12
CHAPTERS	
1. BACKGROUND	14
1.1. Introduction	14
1.2. Current Practice of Operations Control	16
1.3. AVL	18
1.4. Study Objectives and Scope	20
1.5. Glossary of Terms	23
2. HOLDING CONTROL AT MULTIPLE HOLDING STATIONS	26
2.1. Introduction	26
2.2. Problem Formulation and Solution	33
2.2.1. Model Formulation	33
2.2.2. Proposed Heuristic	47
2.3. Simulation Study	62
2.3.1 Scenario Design	62

TABLE OF CONTENTS - Continued

2.3.2 Simulation Design	70
2.3.3 Simulation Results	72
2.3.3.1. Optimal Holding Station(s)	72
2.3.3.2. Strategy Performance Comparison	77
2.4. Conclusions and Discussion	87
3. STOP-SKIPPING CONTROL	89
3.1. Introduction	89
3.2. Problem Formulation & Solution	97
3.2.1. Problem Nature	98
3.2.2. Assumptions and Variable Definitions	99
3.2.3. Objective Function	103
3.2.4. Stop-Skipping Control and Vehicle Trajectory Evolution Dynamics.....	105
3.3. Simulation Study	118
3.3.1. Simulation Logic	118
3.3.2. Simulation Study Design	120
3.3.3. Simulation Results	124
3.3.4. Capacity Constraint Analysis	138

TABLE OF CONTENTS - Continued

3.4. Conclusions	140
4. VEHICLE DISPATCHING WITH SWAPPING AT A TRANSFER TERMINAL	144
4.1. Introduction	144
4.2. Vehicle Dispatching Control Strategies	152
4.2.1. Holding	153
4.2.2. Vehicle Swapping	154
4.2.3. Holding and Swapping	155
4.3. Vehicle Dispatching Process & Model Formulation	156
4.3.1. Assumptions and Vehicle Dispatching Process	156
4.3.2. Model Formulation	158
4.3.3. Solution Method	164
4.4. Simulation and Sensitivity Analysis	166
4.4.1. Simulation Logic	166
4.4.2. Assumptions	171
4.4.3 Test Scenarios and Basic Parameters	174
4.5. Simulation Results and Sensitivity Analysis	177
4.5.1. Sensitivity to Operating Factors	177

TABLE OF CONTENTS - Continued

4.5.2. Sensitivity to Uncertainty of Passenger Boardings	186
4.5.3. Characteristics of Swapping-Prone Scenarios	188
4.6. Conclusions and Future Study	190
5. AVL DATA ANALYSIS	194
5.1. Introduction	194
5.2. AVL Data	203
5.3. Regression Methodology and Parameter Estimation	208
5.3.1. Assumption 1: Trip-Specific Vehicle Operating behavior	209
5.3.2. Simulation Analysis of Sampling	214
5.3.3. Assumption 2: Day-Specific Vehicle Operating Behavior	220
5.3.4. Assumption 3: Combined Vehicle Operating Behavior	227
5.3.5. Comparison of the Methodologies	233
5.4. Conclusion	236
REFERENCES	238

LIST OF ILLUSTRATIONS

1.1.	Schematic of an AVL System Used in a Transit Agency	19
1.2.	Study Objectives and Scope	20
2.1.	Chapter Flowchart (Holding Control)	32
2.2.	Comparison of Model Trajectory and Real Trajectory	49
2.3.	Typical Transit Route with Multiple Holding Stations (Holding Control).....	55
2.4.	Simulation Flowchart – Holding Problem	64
2.5.	Test Passenger Boarding and Loading Profiles (Holding Control)	69
2.6.	Illustration of Vehicle Trajectories under a Variety of Conditions (Holding Control)	76
3.1.	Chapter Flowchart (Stop-Skipping Control)	97
3.2.	Typical Route Segmentation for Stop-Skipping Control.....	106
3.3.	Passenger Loading Profiles (Stop-Skipping Control)	123
3.4.	Passenger Boarding Profiles (Stop-Skipping Control)	123
3.5.	Overall Passenger Waiting Time Reductions (Stop-Skipping Control)	131
3.6.	Comparison of Skipping Segment Length and Location.....	132
3.7.	Comparison of Passenger Cost Reduction between Policy 1 and Policy 2 (Stop-Skipping Control)	133
4.1.	Chapter Flowchart (Vehicle Dispatching with Swapping)	152
4.2.	Logic Diagram of Simulation Study for Terminal Swapping.....	168
4.3.	Simulation Flowchart – Vehicle Dispatching with Swapping Problem	170
4.4.	Sensitivity Analysis for Terminal Swapping	182
5.1.	Example of vehicle locations	200
5.2.	Route 8 of Sun Tran in Tucson, AZ	204
5.3.	Three Typical Situations for Running Speed Derivation	222
5.4.	Example for Estimating the Composite Difference of Running Time	231

LIST OF TABLES

2.1.	Other Operating & Simulation Parameters (Holding Control)	70
2.2.	Holding Strategy Performance Comparison (Profile 1 & Headway 5 Min)....	78
2.3.	Holding Strategy Performance Comparison (Profile 1 & Headway 6 Min)....	78
2.4.	Holding Strategy Performance Comparison (Profile 1 & Headway 8 Min)....	78
2.5.	Holding Strategy Performance Comparison (Profile 1 & Headway 10 Min).....	79
2.6.	Holding Strategy Performance Comparison (Profile 2 & Headway 5 Min)....	79
2.7.	Holding Strategy Performance Comparison (Profile 2 & Headway 6 Min)....	79
2.8.	Holding Strategy Performance Comparison (Profile 2 & Headway 8 Min)....	80
2.9.	Holding Strategy Performance Comparison (Profile 2 & Headway 10 Min)....	80
2.10.	Holding Strategy Performance Comparison (Profile 3 & Headway 5 Min)....	80
2.11.	Holding Strategy Performance Comparison (Profile 3 & Headway 6 Min)....	81
2.12.	Holding Strategy Performance Comparison (Profile 3 & Headway 8 Min)....	81
2.13.	Holding Strategy Performance Comparison (Profile 3 & Headway 10 Min)....	81
2.14.	Minimum Passenger Waiting Time with Varying Headways	85
2.15.	Comparison of Total Passenger Cost after Holding with Minimum Total Passenger Cost	86
2.16.	Comparison of Total Passenger Cost without Holding with Minimum Total Passenger Cost	86
3.1.	Other Operating & Simulation Parameters	124
3.2.	Relative Performance Comparison of Policy 1 and Policy 2.....	136
3.3.	Comparison of Simulation Results with/without a Capacity Constraint	139
4.1.	Transit Operation Parameters in Simulation (Terminal Swapping)	175
4.2.	Sensitivity to Swapping Group Size	178
4.3.	Sensitivity to Swapping Threshold Cost	178
4.4.	Sensitivity to Total Number of Routes	178
4.5.	Sensitivity to Travel Time Variation	179

LIST OF TABLES - Continued

4.6.	Sensitivity to Originating Passenger Boardings	179
4.7.	Sensitivity to Average Route Headway	179
4.8.	Sensitivity to Slack Time at Terminal	180
4.9.	Originating Passenger Cost Reduction vs. Originating Passenger Variation	187
4.10.	Statistical comparison of Swapping-Prone/Non-Swapping-Prone Scenario Groups in the Baseline Case	189
5.1.	AVL Data Format of after Preliminary Processing	208
5.2.	AVL Data Form for Regression (Assumption 1)	210
5.3.	Transformed AVL Data Form for Regression	211
5.4.	Regression Results – Vehicle Running Speed (Assumption 1, 7:15 AM Trip)	212
5.5.	Regression Results – Passenger Boarding Rate (Assumption 1, 7:15 AM Trip)	213
5.6.	Average Vehicle Running Speeds and Passenger Boardings (10 min vehicle headway)	216
5.7.	Regression Results – Passenger Boardings (Simulation Example)	217
5.8.	Regression Results – Vehicle Running Speed (Simulation Example).....	217
5.9.	Travel Time Using Derived Parameters and Using Actual Parameters.....	219
5.10.	Regression Data Form (Assumption 2).....	225
5.11.	Regression Results – Passenger Boarding Rate (Assumption 2, 7:15 AM trip).....	226
5.12.	Regression Results – Vehicle Running Speed (Assumption 1, 7:05 AM Trip)	229
5.13.	Regression Results – Passenger Boarding Rate (Assumption 1, 7:05 AM Trip)	229
5.14.	Regression Results – $\eta_{j-1,j}$ (Assumption 3, 7:15 AM Trip)	232
5.15.	Model Performance Comparison	235

ABSTRACT

This dissertation studies three public transit operations control strategies with automatic vehicle location (AVL) data available. Specifically, holding control, stop-skipping control and vehicle dispatching with swapping are investigated. Moreover, AVL data from Tucson, Arizona are employed to investigate the methodologies for deriving vehicle operating parameters.

The problem of holding vehicles at multiple holding stations can be modeled as a convex mathematical programming problem which can be solved to near optimality by a proposed heuristic. A simulation study on the holding problem suggests that holding control based on the proposed problem formulation can effectively reduce the total passenger cost. Also, multiple holding stations may offer more opportunities to regularize vehicle headways so that holding vehicles at multiple stations can further reduce the passenger cost compared to holding vehicles only at a single station.

Stop-skipping is investigated to respond more rapidly to vehicle disruptions occurring in the middle of a route. Based on a preliminary analysis of the basic stop-skipping policy, a policy alternative is constructed. The stop-skipping strategy is formulated separately for both policies as a nonlinear integer programming problem. The problem solution relies on an exhaustive search method. Another simulation study is conducted to examine how the performance of the two policies change with the passenger distribution pattern, the vehicle disruption location and length, and the vehicle travel time variability. The simulation result suggests selective superiority of the two policies.

The vehicle dispatching problem investigates the potential of integrating real-time swapping into the vehicle dispatching strategies at a transit transfer terminal. With a hypothetical study design, simulation is employed again to evaluate the significance of real-time swapping by comparing the performance of a swapping-holding combined strategy with the holding-only strategy. A sensitivity analysis is also employed to compare these two strategies among key transit operating factors.

Finally, using three different understandings (assumptions) of vehicle operating behavior, regression methods are proposed for using AVL data to derive the vehicle running speeds and passenger boarding rates, which serve as inputs to the operations control models. The regression results show that the day-specific operating behavior may not be appropriate, and that operating behavior combining both trip-specific and day-specific effects seems to be slightly superior to the trip-specific behavior overall.

CHAPTER 1 BACKGROUND

1.1. Introduction

In the past decade, public transit has played a growing role within the multi-modal transportation system in many urban areas. An increasing number of commuters resort to the public transit system for commuting to and from work/school. The capacity of the public transit system is mostly defined by the network design, route layout and the operating schedule. However, the effectiveness of the system capacity and the system operational performance are largely determined by the system operating conditions. In a majority of the urban areas in the United States, traffic congestion has become increasingly common, especially in central urban areas. For instance, according to the 2001-2025 Regional Transportation Plan developed by the Pima Association of Governments (PAG, 2001), the vehicle-miles traveled are expected to increase by 75 percent from 2000 to 2025 while vehicle-hours traveled are anticipated to increase by 87 percent; travel under heavily congested roadway conditions is expected to increase from 22 percent in 2000 to 31 percent in 2025; and travel under severe congestion is expected to increase from about 5 percent in 2000 to 23 percent in 2025. As a result, public transit service has become more subject to the traffic environment, and public transit operations has become more variable, which may lower the system reliability and add extra travel time and inconvenience to transit riders. In the long run, this may result in the loss of ridership.

Operations control has been long employed by transit agencies in United States for a variety of purposes, such as to improve the transit system performance, to restore the

service from disturbance and/or disruption, as well as to reduce the system cost from the perspectives of both operators and transit passengers. For a fixed-route transit service, service frequency may imply different passenger arrival processes at the stops. Passengers are more likely to consult with the published schedule to ride the bus coming every half hour than to take the bus coming every five minutes. In the Transit Capacity and Quality of Service Manual (TCRP, 1999), the frequent transit service is referred to as a service with an operating headway less than or equal to 10 minutes on average. For frequent transit service, the passengers do not need to consult a schedule. Correspondingly, the transit service with the average operating headway larger than 10 minutes is called infrequent transit service.

For frequent transit service, Welding (1957) showed that passengers arriving at stop can be assumed to be a random process, and the average passenger waiting time at the stop can be expressed as follows:

$$E(w) = \frac{1}{2} \left(E(h) + \frac{V(h)}{E(h)} \right) \quad (1-1)$$

Wherein,

$E(w)$: Average passenger waiting time;

$E(h)$: Expected headway between successive vehicles; and,

$V(h)$: Variance of headways.

Equation (1-1) implies that, for frequent transit service, the passenger waiting time is not only defined by the predetermined service headway but also by the operating stochasticity.

Due to the different passenger arrival processes, operation control strategies may function differently for the transit services of different frequencies. For example, holding control is one of the most commonly used transit operations control strategies by transit agencies. It involves delaying a vehicle's departure time at the specified point(s) along a transit route. For the infrequent service, since passengers arrive at stops according to the schedule, holding control is often implemented to delay a vehicle running ahead of schedule until the scheduled departure time, so as to avoid passengers missing the vehicle. On the contrary, for a frequent transit service, holding control may be employed to regularize the vehicle headway distribution, whether or not the vehicle is currently operating on schedule. Holding the vehicle just because it arrives slightly earlier than scheduled may not make much sense, since the following vehicle will come shortly anyway for the frequent transit service, and the passengers missing the vehicle can catch the next one without incurring unacceptable extra waiting time. As indicated by equation (1-1), for the frequent transit service, regularizing the vehicle headways, or equivalently reducing the vehicle headway variance at the stops downstream of the control point, is of the utmost importance for the operations control to improve the system performance.

1.2. Current Practice of Operations Control

Determining the optimal transit operation control strategies can be seen as a decision-making process, and the information used to input to the decision-making process can be roughly grouped into two categories:

Static Information: the route layout, timetables (schedules), average passenger demand at stops, etc.; and,

Dynamic information: vehicle location, vehicle running time, vehicle dwell time, headway distribution, schedule adherence, etc.

In a public transit system, the route layout is rarely subject to change, and the operating timetable is only updated on relatively infrequent basis, such as monthly or seasonally. In addition, the passenger demand at stops is essentially determined by the land use within the vicinity of the stops, so it can be basically seen as a constant during a specified time period, e.g. peak hours. In view of the steady nature of these factors, we consider them and the like as the static information, and generally assume that they are perfectly known, or can be easily collected or derived, for the operations control decision-making process.

However, the static information can only function as part of the basis on which the operations control decision is made. Dynamic information, e.g. vehicle running time, dwell time, headway variance, schedule adherence and many other factors that change with traffic conditions, are also indispensable for evaluating the transit operating performance, and in turn, judging whether or not an operation control strategy is advantageous. Furthermore, both headway variance and schedule adherence are defined by the vehicle arrival time or departure time at stops, which makes the estimation of the vehicles' arrival time or departure time crucial for the operations control decision-making. This will be further discussed in detail in the following chapters, and a specific study of vehicle operating parameters and travel time forecasting is introduced in Chapter 5.

1.3. AVL

Before the 1990's, there were very few technologies being able to collect, process and distribute the dynamic information efficiently for the transit system. Consequently, the public transit operations control decisions were often made locally with fairly limited information. It may not be reasonable to expect that operations control decisions made in such a way can bring anything more than service improvement at a local level.

As an echo to the calls for more advanced technology to facilitate public transit fleet management, in 1991, the Federal Transit Administration (FTA) launched the Advanced Public Transportation Systems (APTS) program to coordinate all federally-sponsored transit Intelligent Transportation System (ITS) programs and initiatives, with Automatic Vehicle Location (AVL) systems as one of the core technologies. In *Advanced Public Transportation Systems: The State of the Art Update 2000* (Federal Transit Administration, 2000), AVL systems are defined as the computer-based vehicle tracking systems that function by measuring the real-time position of each vehicle and relaying the information back to a central location. A typical AVL system employed by transit agencies is illustrated in Fig. 1.1.

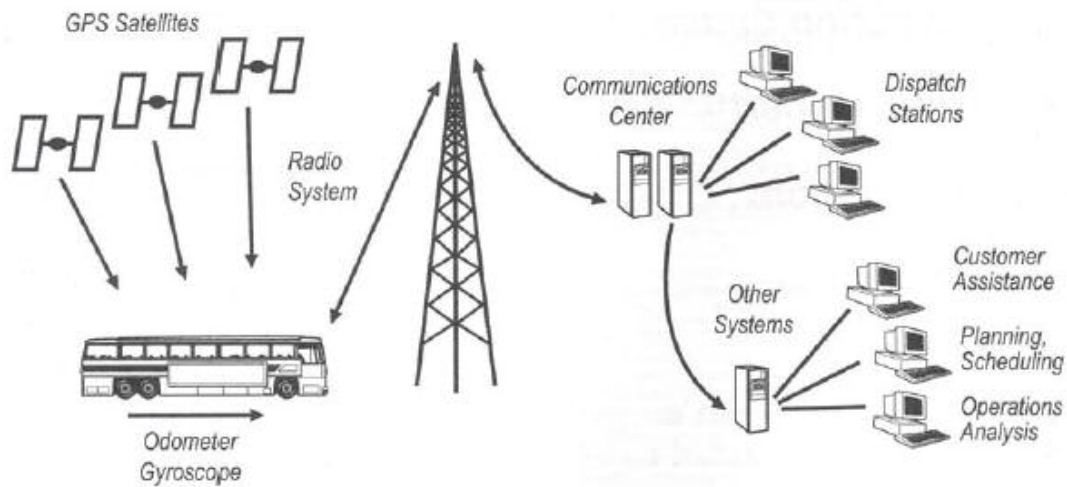


Fig. 1.1. Schematic of an AVL System Used in a Transit Agency (Federal Transit Administration, 2000)

For transit agencies, the AVL system makes it possible to manage the transit fleet in a real-time manner. Vehicle locations, which may be translated into vehicle trajectories, can be collected, stored and processed into a specific form at the vehicle dispatching center, and can provide a wealth of information to examine the propriety of the timetable, to evaluate the transit service performance in terms of either schedule adherence or operating headway regularization, and specifically to predict vehicles' future trajectories on a real-time basis.

Furthermore, with the predicted vehicle trajectories and the transit system operating performance, along with the continuously-updated real-time vehicle location information, the AVL system can provide many advantages for operations control decision making. The latest vehicle location information continuously contributes to the decision making process. Also, rather than the vehicle operators, the central dispatch, with more complete knowledge of the overall transit system operation, decides the operations control. This

can make the control strategy mostly free of the deficiencies of the decision made locally with limited information.

1.4. Study Objectives and Scope

The study objectives and scope of this dissertation is illustrated in Fig. 1.2.

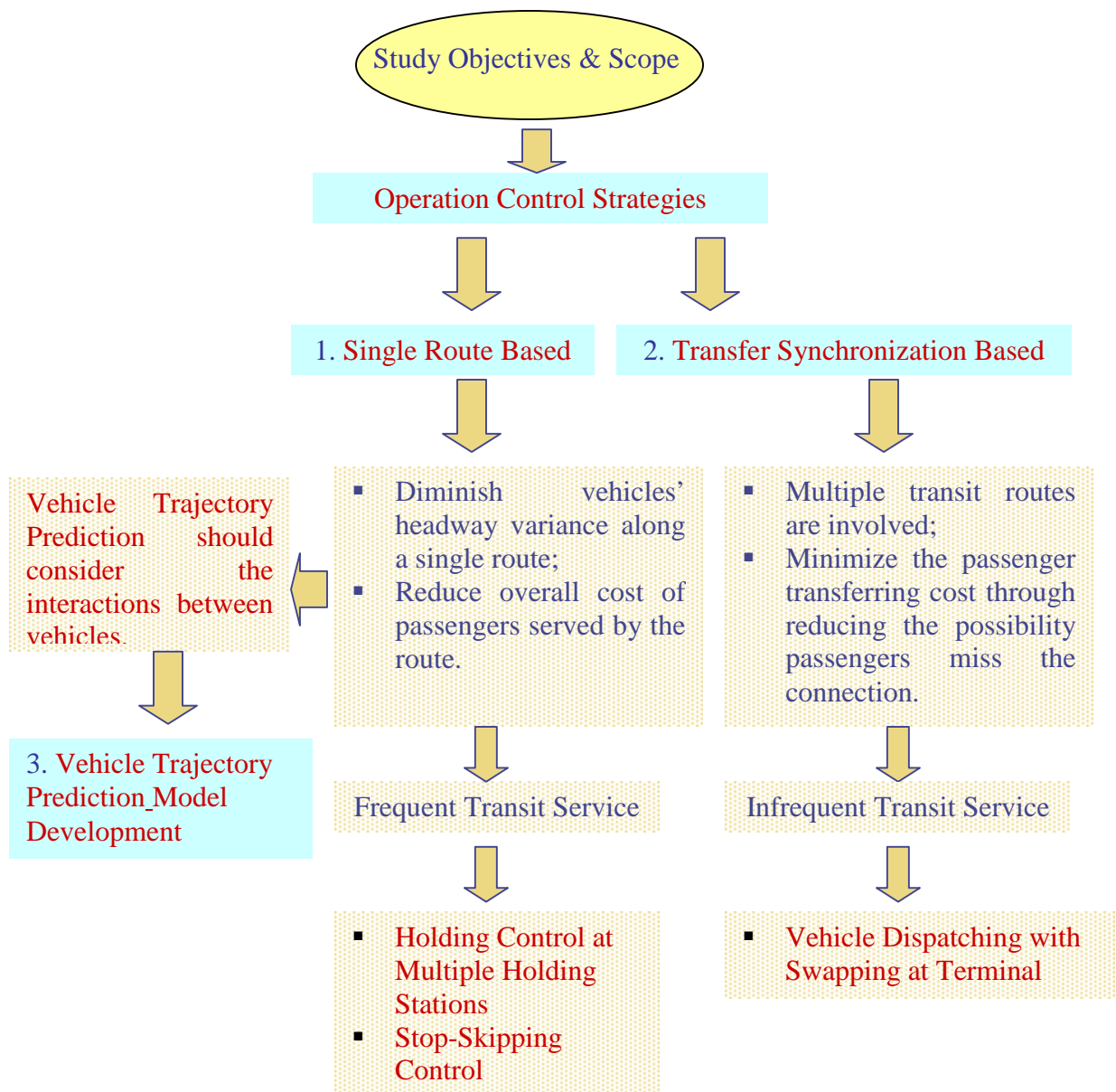


Fig. 1.2. Study Objectives and Scope

As shown in Fig. 1.2, the operation control strategies can be loosely divided into two categories in terms of the purpose and function: namely, single route-based control strategies, and transfer synchronization-based control strategies. Single route-based strategies aim to diminish the vehicles' headway variance along a single transit route so as to reduce the overall cost (in terms of waiting time) to the passengers served by the route. In comparison, multiple transit routes are involved for the transfer synchronization-based strategy. This strategy is applied to enhance the passenger transfer coordination at a transfer point so as to minimize the passenger transferring cost through reducing the possibility that passengers miss the connection. Relatively, the detailed operating characteristics of each individual route are not of major concern for the transfer synchronization-based strategy. This study deals with both single route-based and transfer synchronization-based operations control strategies, though no intention is paid to be exhaustive. Specifically, this study formulates the problems for two single route-based operations control strategies that are most commonly used in practice by transit agencies, namely the holding control problem and the stop-skipping control problem. Furthermore, the control strategies' performance is examined within the context of real time implementation through simulation studies. Also, this study investigates the potential of swapping vehicles operating on multiple routes that connect at a common terminal for facilitating passenger transfers.

As also shown in Fig. 1.2, though operations control can be employed for transit services of different frequencies, the holding control and stop-skipping control problem studied in this dissertation focuses on only the frequent transit service, with the vehicle

headway regularization along a single transit route as the major concern. Furthermore, to differentiate from the similar studies on the holding control problem and the stop-skipping control problem, the holding control problem studied in this dissertation focuses on holding transit vehicles at multiple holding stations, rather than at a single holding station. The stop-skipping control problem emphasizes on the real time application of stop-skipping control with different policies in that the stop-skipping control can be applied whenever it is needed to respond more rapidly to the vehicle disruption. In the past studies, stop-skipping control is generally treated as a vehicle dispatching strategy at the terminal, and the policy guiding how stops are skipped has obvious shortcoming since it may need to force some onboard passengers to get off the vehicle before their destinations.

For the transfer synchronization-based operations control problem, the past research has indicated that only infrequent transit service can benefit from the operations control. This will confine the study of the problem of vehicle dispatching with swapping to be applied only to the relatively infrequent transit services. Furthermore, this study on the problem of vehicle dispatching with swapping makes a significant contribution in the area of transit operation control in that it is the first to investigate vehicle swapping within the context of a real-time application.

Moreover, consistent with a majority of the previous studies on operations control for frequent transit service, the studies on the single route based operations control, specifically holding control and stop-skipping control in this dissertation, again assume that the vehicle downstream trajectories can be precisely predicted. This dissertation also

develops methods to predict the vehicle trajectories, using a recursive vehicle trajectory evolution to explicitly model the interactions between different vehicles, as shown in Fig. 1.2. Furthermore, the methodologies proposed in this dissertation focus on deriving the transit vehicle operating factors using AVL data collected specifically from a low-level AVL system, which does not provide directly the measures of the vehicle operating parameters, e.g. vehicle running speeds and passenger boarding rates.

The remainder of this dissertation is organized into four chapters, addressing the holding control problem, the stop-skipping control problem, the vehicle swapping problem, and the AVL data analysis, respectively. Though the study design, problem formulation and notation within the three operation control problems may share some common parts, all chapters are relatively independent of each other, and each of them constitutes a stand-alone study. Furthermore, for the sake of maintaining the consistency of style throughout the dissertation, the following three chapters for the operations control problems all start with the literature review on the pertinent problem, followed by problem introduction, formulation and solution, then a simulation study to examine the performance of the control strategy, and conclusions. The AVL data analysis is much different, but the core part of the chapter is still the methodology and mathematical formulation.

1.5. Glossary of Terms

For the sake of clarity and uniformity, some major terms that appear in later chapters are defined here.

Public Transit: A type of service which is delivered by public transit agencies and operates on established schedules along fixed routes so as to move relatively large ridership between designated stops.

On-Call Service: A type of transit service in which the transit vehicle only stops to serve the stops/stations in response to calls from passengers, either for boarding or alighting, or both.

Dwell Time: The time incurred by a transit vehicle for decelerating to a stop, for passenger boarding and alighting at the stop, for accelerating from the stop, and for re-entering the traffic stream.

Vehicle Trajectory: The transit vehicle's course along the transit route, which can be depicted both temporally and spatially with the vehicle's location and the associated time when the vehicle's location is recorded.

Vehicle Running Time: Vehicle travel time over a route segment(s), excluding the dwell time at stop(s), but including the intersection delay time.

Vehicle Travel Time: The time incurred to a transit vehicle for traveling certain distance, including both running time over the route segment and the dwell time at stop(s).

Vehicle Running Speed: The distance traversed divided by the vehicle running time.

Vehicle Operation Speed: The distance traversed divided by the vehicle travel time.

Vehicle Headway: The time difference between two successive vehicle departures at one specific stop.

Headway Variance: The variance of the vehicle headways observed at a specific stop during the time period of interest.

Schedule Adherence (or Schedule Deviation): The difference between a vehicle's actual departure time and the scheduled departure time at a stop.

Holding: A strategy to hold a vehicle at a stop/station for a specified amount of time to reduce the passenger waiting time and improve the service reliability downstream.

Threshold-Based Holding: A type of headway-based holding strategy with a headway threshold. A vehicle is required to be held to make its leading headway at least as large as the predefined threshold value.

Stop Skipping (also called vehicle expressing): A strategy to improve the vehicle operation by skipping several stops so as to catch up the schedule or improve the headway distribution downstream.

Vehicle Swapping: A strategy to swap the vehicles among different routes so that the service reliability can be improved, and the overall passenger delay as well as the probability that passengers miss the connection at a transfer terminal may be reduced.

Vehicle Overtaking: The following vehicle overtakes the leading vehicle.

Automated Passenger Counting (APC) Device: A device used to count passengers getting on and off a vehicle.

Automatic Vehicle Location (AVL) System: A computer-based vehicle tracking system that measures the real-time location of a vehicle and relays this information to a central location.

CHAPTER 2 HOLDING CONTROL AT MULTIPLE HOLDING STATIONS

2.1. Introduction

Among all the transit operations control strategies currently employed in practice by transit agencies in the U.S., holding control is undoubtedly the most commonly used one. Holding control involves keeping a vehicle at a station for a period of time, in order to improve the service performance (minimizing passenger waiting time).

Barnett (1974) developed a model for holding a vehicle at a chosen control point. He proposed a solution algorithm for constructing an approximately optimal dispatching strategy at the control point in terms of minimizing the delay for both at-stop and in-vehicle passengers. This strategy is a threshold-based holding control. In specific, a threshold headway is determined. At the control point, if the vehicle headway is less than the threshold, the vehicle is held until the threshold. If the vehicle headway is greater than the threshold, the vehicle is dispatched immediately. Barnett's algorithm was tested using data from a Boston subway line to propose service improvements. Abkowitz and Tozzi (1986) conducted a study for evaluating the sensitivity of headway-based holding control to various boarding and alighting profiles, headways, and other characteristics of route operations. They found that profiles with passengers boarding at the middle and alighting at the end of a route produce the most significant passenger waiting time savings with holding control. Also, the increase in the initial headway variation and amount of parking permitted along a route are likely to deteriorate route reliability and thus improve the effectiveness of the holding strategy. At about the same time, Abkowitz et al. (1986) investigated the effects of a threshold-based holding control strategy on reducing the

headway variation at stops downstream of the control point. Their simulation results indicated that the headway variation does not increase linearly along a route. Also, the study results showed that it is preferable to locate the control point just prior to a group of stops where many passengers are boarding; and, the threshold headway is sensitive to the number of passengers onboard the bus at the control point. In addition, this study concluded that the optimal holding control could result in a 3-10 percent reduction in total passenger waiting cost. Later, Abkowitz and Lepofsky (1990) conducted a before-after study for evaluating the effectiveness of the threshold-based holding strategy on several bus routes chosen from the MBTA in Boston. The results from this study were not conclusive; however, it still appeared that certain route segments might have benefited from the holding actions. O'Dell and Wilson (1999) developed a deterministic model of a rail system and mixed integer programming formulations for the holding and short-turning problems. Three holding strategies, holding each train at any station, holding each train at the first station it reaches after the disruption occurs, and holding each train at an optimally chosen station, were considered and formulated. Study results based on the MBTA Red Line showed that passenger waiting time can be significantly reduced by applying the controls.

With the advent of the AVL and APC (automatic passenger counting) technologies, the real-time vehicle location information is incorporated by many researchers into their studies. Furth (1995) developed a strategy to deal with a vehicle operating behind schedule, given the existence of an intelligent system providing information about vehicle location, vehicle load, and number of passengers waiting at stops. In his study, the

problem is formulated as a constrained non-linear optimization problem, to decide: how many vehicles following the initially delayed vehicles should be held; the location at which each vehicle should be held; and, the amount by which each vehicle should be held. Study results show that the optimal solution is a gradual increase in the overall headway from the first vehicle, whose headway is short, until the last vehicle with headway returning back to the base headway. Ding and Chien (2001) formulated a real-time operational control model in which the vehicle departure time at each stop is optimized so that the headway variance weighted by passengers at each stop can be reduced. The proposed real-time control model was tested by simulation based on a high frequency light rail transit route in the city of Newark, New Jersey, and the simulation results demonstrated the average passenger waiting time can be significantly reduced by applying the proposed control model.

Hickman (2001) presented an analytical model for optimizing the holding time at a given control point in the context of a stochastic vehicle operations model. In this study, the single vehicle holding problem is a convex quadratic program in a single variable, and is easily solved using gradient or other line search techniques. Eberlein et al. (2001) also formulated an analytic model using a rolling-horizon approach, using AVL information. The problem can be effectively solved by a proposed heuristic. The study results showed significant reductions of passenger waiting time at stops. Fu and Yang (2002) investigated both the threshold-based holding control model and an optimal holding control model by considering both a vehicle's preceding and following headways, with the assumption that the future bus arrival time at the control stop can be predicted

with real-time location information. Based on a simulation, the study results indicated that: the control point should be placed at the bus stop with high demand and located close to the middle of the route; two control points are more preferable than only one; holding control is fairly robust with respect to the control parameter, control strength or headway threshold; and, real-time bus location information can help reduce passenger in-vehicle time and bus travel time when a number of control points are used.

Zhao et al. (2001) present a distributed control approach based on multi-agent negotiation (between a bus agent and a stop agent) for the holding problem. The negotiation in this study is conducted based on the marginal cost and marginal benefit of a hold, negotiated between a vehicle and the set of stops on the route. Also, the comparison between the negotiation algorithm and other commonly used strategies was conducted through simulation, and study results indicated that the negotiation algorithm is robust to different transit operating environments.

From the literature review above, one may see that it is commonly concluded that holding can undoubtedly improve the performance of transit service by diminishing the vehicle headway variance and schedule deviation, and hence can reduce passenger waiting time, if the control location is judiciously selected. However, some of the previous studies also pointed out either explicitly or implicitly that the transit operating stochasticity still plays a role in the vehicles' trajectories downstream of the control point after holding is applied. Based on the equations developed in their study, Abkowitz et al. (1986) concluded that:

The reduction in headway variation at points downstream of the control point is not uniform. The maximum benefits of the control strategy

are accrued by passengers at stops immediately downstream of the control point. Stops that are far from the control point may not be impacted significantly. (Abkowitz et al. 1986 , p. 78-79)

Furthermore, Turnquist and Blume (1980) showed that there might be multiple points qualifying as holding point candidates along the route. Though not clearly indicated in the study, choosing one qualified location as the control point does not imply that the others cannot still be qualified as control points, while some correlation certainly exists between the potential holding points. Abkowitz and Tozzi (1986), Abkowitz et al. (1986) and Fu and Yang (2002) all presented desirable conditions to deploy a control point to hold a vehicle(s). However, it can often be seen that a route may have favorable conditions on separate segments, which might justify multiple control points.

Moreover, it has been assumed by a majority of the previous studies that the transit vehicle trajectories downstream of the holding station can be predicted precisely with the currently available information, typically from AVL technology; or, the vehicle arrival/departure time at the stop can be depicted by the best-fit probabilistic distribution built on the historical data, if they are subject to random variation. However, in reality, as the transit vehicle's running time and dwell time may be both subject to significant variability, it becomes fairly difficult, if not impossible, to precisely predict vehicle trajectories far downstream of the holding station.

Seneviratne and Loo (1986) have analyzed the vehicle travel time data from two transit routes in Halifax, Nova Scotia, Canada, and found out that fundamental to a realistic analysis of a bus route is proper segmentation; that is, routes may be broken into route segments within which operations is fairly homogeneous.

In summary, it appears possible that holding control can be implemented at multiple stations, especially when the transit route is relatively long with many stops. This conclusion is based on the premise that separate route segments may need separate operations control actions.

The problem of holding control at multiple holding stations is to decide multiple vehicles' holding times at multiple holding stations. Therefore, holding vehicles at multiple holding stations can essentially be seen as a three-dimensional decision problem: the vehicle holding time at a particular stop is one dimension, the control vehicle is the second dimension, and the holding station is the third dimension. Eberlein et al. (2001) have presented efforts to compare the benefits from holding vehicles at multiple holding stations and from holding at only one single holding station. Their study concluded that holding the vehicles at more than one holding station did not show any significant advantages, using a numerical example based on a real-life transit route. However, the observation may be not sufficiently conclusive due to the limitations of the tested passenger loading/boarding profiles.

Although the holding control at multiple stations involves three dimensions of the problem, this study only examines two dimension of the problem, namely to determine the holding times for multiple vehicles at a given set of holding stations. In the remainder of this chapter, the problem of holding vehicles at multiple holding stations is formulated as a mathematical programming problem within the context of a deterministic service model to optimize the vehicle holding times so as to minimize the total passenger cost. A heuristic is proposed to solve for the optimal holding times through decomposing the

overall problem into smaller problems and solving the smaller problems iteratively to achieve global solution. Furthermore, a CRN (Common Random Number) simulation study is conducted to compare the holding control at a single holding station and the holding control at multiple holding stations in terms of the total passenger cost. The basic structure of this chapter is also depicted in Fig. 2.1.

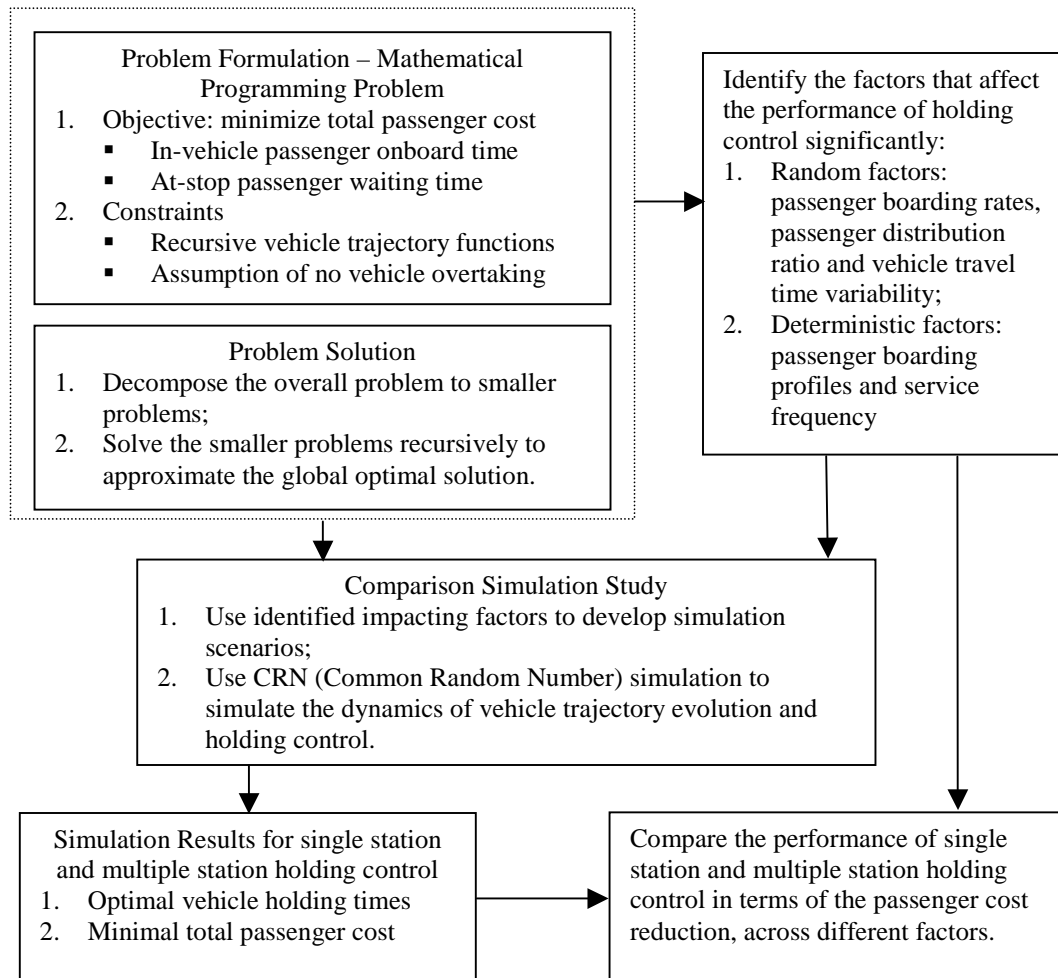


Fig. 2.1 Chapter Flowchart (Holding Control)

Fig. 2.1 shows that the problem formulation can help identify the important factors that may affect the performance of holding control. These identified impacting factors

can be further used to construct simulation scenarios to conduct a CRN based comparison simulation study (under each simulation scenario, use the holding problem formulation and solution method to seek optimal vehicle holding times and total passenger cost) to compare how holding control at a single holding station and holding control at multiple holding stations relatively perform in terms of the total passenger cost reduction. Specifically, the remainder of this chapter is organized into three sections. Section 2.2 formulates the general holding problem with either a single holding station or multiple holding stations. A heuristic based on an analytical model is also described in this section. Section 2.3 provides a hypothetical simulation study designed to demonstrate the effectiveness of the algorithm developed in Section 2.2, and to compare the performance of holding vehicles at multiple stations and at only a single station in terms of the passenger cost reduction being achieved from the control. Finally, Section 2.4 concludes this chapter and presents the direction for future research on the holding problem.

2.2. Problem Formulation and Solution

As argued in Eberlein et al. (2001), the holding control problem can be formulated in the context of a deterministic model of transit operations. In a similar manner, the problem formulation for holding control at multiple stations will be again presented using a deterministic model in this chapter.

2.2.1. Model Formulation

For the sake of simplifying the analysis that follows, several assumptions are made:

- Vehicle overtaking is not a major concern;

- The passenger arrival rate at any one particular stop and vehicle average travel time between adjacent stops are given during the time period of interest;
- The number of alighting passengers at a stop is proportional to the number of passengers onboard; and,
- Vehicle capacity is not considered.

One may argue with the first assumption of no vehicle overtaking, but this assumption can be justified when: transit service is provided at high frequency, but the average headway is still relatively large, e.g. larger than 5 minutes; and, when traffic conditions do not change abruptly during the time period of interest, so that vehicle running times only differ randomly from one trip to another.

Therefore, the assumption of no vehicle overtaking may hold in the situations likely to satisfy the conditions above. Furthermore, holding control at multiple holding stations helps regularize vehicle trajectories for multiple times, which may greatly reduce the chance for vehicle overtaking to occur. This will be further discussed later in the chapter.

Before the holding control problem is formulated, the major variables are defined below.

$h_{i,j}$: Leading headway for the i^{th} vehicle at stop j ;

$d_{i,j}$: Departure time for the i^{th} vehicle at stop j ;

$a_{i,j}$: Arrival time for the i^{th} vehicle at stop j ;

$l_{i,j}$: Onboard passengers of the i^{th} vehicle when it departs from stop j ;

H_{j,v_k} : Holding time for the j^{th} vehicle at holding station v_k ;

$B_{i,j}$: Passengers boarding the i^{th} vehicle at stop j ;

$A_{i,j}$: Passengers alighting from the i^{th} vehicle at stop j ;

λ_j : Passenger arrival rate at stop j ;

$r_{i,j}$: Running time between stop j and stop $j+1$ for vehicle i ;

q_j : Passenger alighting proportion at stop j ;

α, β_1, β_2 : Parameters defining the passenger boarding process. α represents the vehicle acceleration, deceleration, door open and close, and clearance time; β_1 is the average passenger boarding time; and, β_2 represents the average passenger alighting time at a bus stop;

v_k : Index of the k^{th} holding station as a stop;

b_k : Index of the earliest dispatched vehicle among those operating on the segment (v_{k-1}, v_k) ;

e_k : Index of the latest dispatched vehicle among those operating on the segment $(v_{k-1}, v_k]$;

M : Total number of holding stations;

N : Total number of stops on the route;

O : Total number of vehicles operating within the time period of concern;

V : Holding station set $\{v_1, v_2, \dots, v_M\}$.

$X_{i,n,j}$: The ratio of the number of passengers boarding at stop n and alighting at stop j to the total number of passengers boarding at stop n for vehicle i . This ratio is originally derived from the historical data, but it is also subject to change when a vehicle

passes a stop and the real number of passengers boarding and alighting at the stop is obtained.

$\bar{B}_{i,j}$: In contrast to $B_{i,j}$, $\bar{B}_{i,j}$ represents the real number of passengers boarding vehicle i at stop j ;

$\bar{A}_{i,j}$: In contrast to $A_{i,j}$, $\bar{A}_{i,j}$ represents the real number of passengers alighting from vehicle i at stop j ;

$P_{i,j}$: The probability that stop j is skipped by vehicle i ;

$Q_{i,j}$: The probability that stop j is served by vehicle i ; and,

$D_{i,n,j}$: The expected number of passengers boarding vehicle i at stop n and alighting at stop j .

As described in Chapter 1, determining the optimal control strategies is a decision-making process, which is essentially adaptive in nature due to the continuously changing traffic conditions and other operation factors. For such an adaptive decision-making process, a deterministic model is appropriate to formulate the problem of holding vehicles at multiple stations. Within an entirely deterministic context, it is meaningless to consider holding one vehicle at all holding stations within one decision-making cycle, because all effects resulting from holding vehicles at downstream holding stations can be mostly achieved by holding the vehicle at the first holding station it will arrive. More specifically, with M holding stations available, the transit route can be divided into $M+1$ segments either bounded by two consecutive holding stations as (v_k, v_{k+1}) , or by a terminal and a holding station as $(1, v_1)$ or (v_M, N) . The vehicles $[b_k, e_k]$ operating on

each segment (v_{k-1}, v_k) make up of the control vehicle group for the holding station, and all vehicles dispatched later are called the impacted vehicles whose trajectories will be affected by the operations control implemented on the control vehicles. It is further assumed that all vehicles within the control group are only considered to be held at the holding station v_k in one decision-making cycle. Obviously, those vehicles operating on the segment $(v_M, N]$ are not subject to control. More concisely, the multiple holding problem can be described as:

At any point in time a decision is made, determine the holding times for vehicles at only the immediate downstream holding station, while multiple holding stations are available.

Specifically, the holding problem formulated in this study is to decide the holding times of the control vehicles $[b_k, e_k]$ at the holding station v_k , given the currently available vehicle locations collected by AVL technology. In the real adaptive decision-making process, the problem formulation will be used again and again to determine the optimal holding strategies based on the latest vehicle operations information. Obviously, the control vehicles' trajectories from their current locations to the holding station to which they belong are not affected by the holding control decision. Accordingly, there are two separate stages to make a holding decision. The first stage is to predict the ready-for-departure times (holding time is not included) for the control vehicles $[b_k, e_k]$ at the holding station v_k , and another stage is to decide the control vehicles' planned departure times (including holding time) at the holding station. For both stages, how to model the vehicle trajectory evolution dynamics is of the utmost importance. The vehicle trajectory

evolution dynamics for the two stages can be modeled either identically or in different manners.

Furthermore, for real-time holding control, two issues merit particular concern. One is the proximity of the modeled vehicle trajectory evolution dynamics and the real dynamics; and, probably more importantly, another one is the effectiveness of the problem solution method. The first stage, namely to predict the control vehicles' trajectories up to the corresponding holding station, does not depend on the solution. Hence, a more detailed and accurate vehicle trajectory evolution dynamics model can be formulated. On the contrary, a simpler vehicle trajectory evolution dynamics model is desirable for the second stage to ensure the problem can be solved effectively.

Stage 1: Predict the trajectories of the control vehicles and impacted vehicles up to the corresponding holding station

As defined in the first chapter, for an on-call transit service, whether or not the vehicle stops to serve a particular stop depends on whether there are passengers calling for boarding or/and alighting at the stop. This may be represented as the probability of serving or bypassing the stop, and is used in the operations control problem formulation to predict vehicles' trajectories. Kikuchi and Vuchic (1982) analyzed the optimal number of stops and vehicle operation policy for transit route operations. In the study, the number of actual vehicle stops is estimated by assuming Poisson passenger arrival pattern for on-call service. Also, the probability of a vehicle bypassing a stop is determined as the probability of no passengers calling for either boarding or alighting at the stop. Banks (1984) improved the representation of the Poisson model to estimate the number of

stoppings for a vehicle, by differentiating the passenger arrival rates at different stops. Powell and Sheffi (1983) employed a binomial distribution to estimate the number of passengers alighting at a stop, given the number of onboard passengers and the probability of a randomly chosen passenger on the bus alighting at this stop. They also used a Poisson process to compute the number of boarding passengers at the stop, given the passenger arrival rate. Furth (1986) treated the passenger demand (including both boarding and alighting demand) at a particular stop as Poisson random variables. Hickman (2001) derived the number of passengers alighting and boarding at a stop based on the properties of binomial and Poisson random variables respectively. To be consistent with these past studies, Poisson and binomial distributions are used in the holding problem and the stop-skipping control problem in next chapter. However, the binomial distribution will be used in a slightly different manner as introduced in following parts of this chapter.

At each time instant when a holding control decision is made, each vehicle trajectory between the current vehicle's position and the downstream holding station can be predicted using the equations (2-1) through (2-10) as follows.

Throughout the dissertation, the passenger arrivals at each stop are assumed to be a Poisson process, and more specifically, all passengers boarding at any particular stop will be seen as coming from the same family with certain probability of alighting at each specific downstream stop. As a result, a binomial distribution can be appropriately employed to estimate the probability of each passenger boarding at stop n and alighting at

stop j . Along with the assumed Poisson process for passenger arrivals, the probability of the vehicle i skipping stop j can be determined as below.

$$P_{i,j} = \left(\prod_{n=1}^{j-1} (1 - X_{i,n,j}) \right)^{B_{i,n}} \cdot \exp[-\lambda_j \cdot (a_{i,j} - d_{i-1,j})] \quad (2-1)$$

The first term on the right hand side of (2-1) is the probability that no onboard passenger calls for alighting at stop j , and the second term represents the probability that no passenger is waiting at stop j for boarding when vehicle i approaches the stop.

The probabilities of vehicle serving a stop and skipping a stop are supplemental to each other.

$$Q_{i,j} = 1 - P_{i,j} \quad (2-2)$$

The vehicle arrival time at each stop would be $r_{i,j-1}$ later than the departure time at the upstream adjacent stop.

$$a_{i,j} = d_{i,j-1} + r_{i,j-1} \quad (2-3)$$

The vehicle departure time at each stop includes the vehicle dwell time, which consists of the passenger boarding time, alighting time, as well as the constant part accounting for door open/close and so on at the stop, whose expected value is also determined by the probability of a vehicle serving that particular stop. Mixed use of simultaneous and sequential passenger boarding and alighting process have been found in previous studies. However, one may also see that the sequential passenger boarding and alighting process has been adopted by a majority of the previous studies to formulate the operations control problem due to its relatively simple mathematical form. For the sake of being consistent with the previous studies and being able to compare the results on a

common basis, the holding control problem in this dissertation also formulates the vehicle dwell time based on the sequential passenger boarding and alighting process, which gives equation (2-4) as follows,

$$d_{i,j} = a_{i,j} + \alpha \cdot Q_{i,j} + \beta_1 \cdot B_{i,j} + \beta_2 \cdot A_{i,j} \quad (2-4)$$

In equation (2-4), the departure time $d_{i,j}$ of vehicle i at stop j is the total of vehicle arrival time $a_{i,j}$ and the vehicle dwell time at stop j , which includes the average constant part of the vehicle dwell time $\alpha \cdot Q_{i,j}$, passenger boarding time $\beta_1 \cdot B_{i,j}$ and passenger alighting time $\beta_2 \cdot A_{i,j}$.

Vehicles are not allowed to overtake each other. Therefore, the vehicle i cannot pull into stop j until the vehicle $i-1$ leaves the stop.

$$d_{i-1,j} \leq a_{i,j} \quad (2-5)$$

The number of passengers alighting at a particular stop is simply the total of all passengers boarding from upstream stops and, at the same time, having the destination at this stop.

$$A_{i,j} = \sum_{n=1}^{j-1} D_{i,n,j} \quad (2-6)$$

The average number of passenger boardings is the product of the vehicle headway and the average passenger arrival rate.

$$B_{i,j} = \lambda_j \cdot (d_{i,j} - d_{i-1,j}) \quad (2-7)$$

Passenger distributed between stops is a proportion ($X_{i,n,j}$) of the passengers boarding at the origin stop.

$$D_{i,n,j} = B_{i,n} \cdot X_{i,n,j} \quad (2-8)$$

For all of the stops that have been passed by the vehicles, all the variables defined by (2-1) through (2-7) are assumed known. This adds the following two equations (2-9) and (2-10). The $D_{i,n,j}$ can also be modified with the observed passenger boardings at stop n and an algorithm can be used for updating $X_{i,n,j}$. In the algorithm, the decision-making process keeps track of the number of passengers which boarded ($\bar{B}_{i,n}$) at all upstream stops and that have alighted ($\bar{A}_{i,n}$) previously. The control system also assumes the passengers actually alighting at a stop come proportionately from those passengers remaining on board. When the vehicle passes a stop j , the decision-making process obtains immediately the number of passengers $\bar{A}_{i,j}$ alighting at the stop j , and in turn updates the passenger O/D distribution ratio $X_{i,n,j}$ for $n < j$. Herein, j represents the stop just passed. In turn, for all downstream stops $n > j$, each $X_{i,j,n}$ is updated proportionately in terms of its own magnitude and the net change of $X_{i,n,j}$ ($n < j$).

$$B_{i,j} = \bar{B}_{i,j} \quad (2-9)$$

$$A_{i,j} = \bar{A}_{i,j} \quad (2-10)$$

Certainly, the methodology employed in this holding control problem and in the stop-skipping control problem in the next chapter can also accommodate the case without passenger counting devices to collect the real number of passengers boarding and alighting at each stop. For such a case, one may simply substitute the expected values for the real numbers in equations (2-9) and (2-10).

One may wonder how to acquire the $X_{i,n,j}$, which could be a prohibitive task for the transit agency. Besides directly using historical data, a simple way to generate $X_{i,n,j}$ is to use the passenger boarding and alighting data from a ride-check survey, which is relatively easy to conduct. With these data, one may estimate the ratio using the origin-destination (O/D) estimation methods, e.g. the maximum entropy method with the number of boarding passengers as the production and the number of alighting passengers as the attraction. For example (the simplest case), $X_{i,n,j}$ can be expressed as $\frac{\bar{A}_{i,j}}{\sum_{l=1}^{j-1} \bar{B}_{i,l}}$, and

this means that all passengers boarding the vehicle at the upstream stops have the equal probability of alighting at stop j .

Stage 2: Decide the control vehicles' departure times at the corresponding holding stations

As argued at the beginning of the problem formulation, the problem solution is another major concern. An over-sophisticated problem formulation may add difficulty to the problem solution. A tradeoff point may need to be achieved to balance the amount of details included in the problem formulation and the ease of the problem solution. For the sake of simplifying the problem solution, it is specifically assumed that the passenger boarding time dominates passenger alighting time at most stops/stations along route, so that the passenger boarding time can be used as the vehicle dwell time. This assumption is often true due to the fact that passengers usually take a longer time to board a vehicle than to get off a vehicle, and it is particularly true for the downtown-oriented transit service with few passengers alighting at the stops along the route before the stops/stations

in downtown. This assumption yields equation (2-11) with the last term removed from the right hand side of equation (2-4).

$$d_{i,j} = a_{i,j} + \alpha \cdot Q_{i,j} + \beta_1 \cdot B_{i,j} \quad (2-11)$$

With the predicted control vehicles' ready-for-departure times at the corresponding holding stations, for a transit route with M holding stations numbered in ascending order with ' I ' representing the one closest to the dispatching terminal, the holding problem can be formulated as follows.

$$\begin{aligned} \text{Minimize} \quad Z = & \frac{1}{2} \cdot \sum_{k=1}^M \sum_{i=b_k}^{e_k} \sum_{j=v_k}^N \lambda_j \cdot (d_{i,j} - d_{i-1,j})^2 + \frac{1}{2} \cdot \sum_{i=e_1+1}^{O+1} \sum_{j=v_1}^N \lambda_j \cdot (d_{i,j} - d_{i-1,j})^2 \\ & + \sum_{k=1}^M \sum_{i=b_k}^{e_k} l_{i,v_k-1} \cdot (1 - q_{v_k}) \cdot H_{i,v_k} \end{aligned} \quad (2-12)$$

In this objective function, the first two components represent the total passenger waiting time at stops, and the third term defines the delay experienced by the onboard passengers at the holding stations. In the second component, since the vehicles from the first non-dispatched vehicle e_1+1 to O are assumed to be the impacted vehicles not subject to control, their departure times at each stop will be determined solely by the vehicle trajectory evolution dynamics. Also, the vehicle $O+1$ is a dummy vehicle which serves to avoid biasing the objective reduction due to the possibly earlier-than-scheduled arrival times of the vehicle O at each stop. Herein, the vehicle $O+1$ is assumed to operate exactly on schedule. Though not salient, it can be seen in the objective function that the departure times of vehicles $[b_k, e_k]$ at each holding station v_k are the decision variables.

Each vehicle's departure time at any stop other than the holding station to which it 'belongs' (e.g. $[b_k, e_k]$ belongs to holding station v_k) is entirely determined by its

arrival time and dwell time at the stop. The dwell time in turn is essentially defined by the time the leading vehicle departed as well as the passenger arrival rate at the stop.

$$\text{If } j \notin V \text{ or } (j = v_k \in V \text{ and } i \notin [b_k, e_k])$$

$$d_{i,j} = (d_{i,j-1} + r_{i,j-1} + \alpha - \beta_1 \cdot \lambda_j \cdot d_{i-1,j}) / (1 - \beta_1 \cdot \lambda_j) \quad (2-13)$$

Equation (2-13) can be directly derived from the relationship below:

$$d_{i,j} = d_{i,j-1} + r_{i,j-1} + \alpha + \beta_1 \cdot \lambda_j \cdot (d_{i,j} - d_{i-1,j}) \quad (2-14)$$

Equation (2-14) is to say that vehicle i 's departure time at stop j is the total of its departure time at stop $j-1$, the vehicle running time $r_{i,j-1}$ between stops $j-1$ and j , and the vehicle dwell time at stop j , which includes a constant vehicle dwell time α and the passenger boarding time $\beta_1 \cdot \lambda_j \cdot (d_{i,j} - d_{i-1,j})$. To be noted here is that the number of passenger boardings for vehicle i at stop j is the product of vehicle i 's headway $(d_{i,j} - d_{i-1,j})$ and the average passenger boarding rate λ_j at stop j .

Otherwise, i.e. for those currently operating vehicles (other than the non-dispatched vehicles) at the corresponding holding stations, the vehicle holding time will define the vehicle's departure time. Specifically, the vehicle departure time after holding is a holding time $H_{i,j}$ later than the ready-for-departure time calculated with equation (2-13).

$$d_{i,j} = (d_{i,j-1} + r_{i,j-1} + \alpha - \beta_1 \cdot \lambda_j \cdot d_{i-1,j}) / (1 - \beta_1 \cdot \lambda_j) + H_{i,j} \quad (2-15)$$

However, any currently operating vehicle $i \leq e_1$ cannot be held as late as the time when the vehicle $i+1$ arrives to avoid vehicle overtaking. The vehicle $i+1$'s arrival time at stop j can be expressed as $d_{i+1,j-1} + r_{i+1,j-1}$.

$$d_{i,j} \leq d_{i+1,j-1} + r_{i+1,j-1} \quad (2-16)$$

Vehicle holding time must be positive.

$$H_{i,j} \geq 0 \quad (2-17)$$

The number of onboard passengers when a vehicle departs from a stop is determined by the number of passengers boarding and alighting at the stop and the number of onboard passengers as the vehicle arrives at the stop.

$$l_{i,j} = l_{i,j-1} + B_{i,j} - A_{i,j} \quad (2-18)$$

The number of passengers boarding a vehicle is the product of the average passenger arrival rate and the vehicle's leading headway.

$$B_{i,j} = \lambda_j \cdot (d_{i,j} - d_{i-1,j}) \quad (2-19)$$

The number of passengers alighting from a vehicle is assumed to be proportional to the number of onboard passengers. q_j can be calculated based on the estimated $X_{i,n,j}$ values and is assumed fixed and identical for all vehicles.

$$A_{i,j} = l_{i,j-1} \cdot q_j \quad (2-20)$$

Equations (2-18) through (2-20) can be combined into one single equation as,

$$l_{i,j} = l_{i,j-1} \cdot (1 - q_j) + \lambda_j \cdot (d_{i,j} - d_{i-1,j}) \quad (2-21)$$

In the model formulation, the decision variable can be either the vehicle holding times (for vehicles $i \leq e_1$) or equivalently the vehicle departure times at the corresponding holding stations, due to the linear relationship between them. From now on, in this chapter, the decision variable is arbitrarily chosen as the departure times of vehicles $[b_k, e_k]$ at each holding station v_k , and accordingly the equation (2-15) is modified into the inequality below to remove the holding time variable $H_{i,j}$.

$$d_{i,j} \geq (d_{i,j-1} + r_{i,j-1} + \alpha - \beta_1 \cdot \lambda_j \cdot d_{i-1,j}) / (1 - \beta_1 \cdot \lambda_j) \quad (2-22)$$

In the objective function, the holding time H_{i,v_k} can be replaced by:

$$H_{i,v_k} = d_{i,v_k} - (d_{i,v_k-1} + r_{i,v_k-1} + \alpha - \beta_1 \cdot \lambda_{v_k} \cdot d_{i-1,v_k}) / (1 - \beta_1 \cdot \lambda_{v_k}) \quad (2-23)$$

Equations (2-13) through (2-23) together define the feasible region for each decision variable. Specifically, inequalities (2-22) and (2-16) together set the lower and upper bounds respectively for the decision variables.

The problem solution proposed in the subsequent sub-sections is focused on the problem formulation for the second stage of the holding problem, given that the control vehicles' ready-for-departure times at the corresponding holding stations have already been successfully predicted.

2.2.2. Proposed Heuristic

With the problem definition and formulation in the previous sub-section, one may see that the departure time of a vehicle within the control group $[b_k, e_k]$ at the stops on the downstream segment $[v_k, v_{k+1}]$ is only determined by a subset of the decision variables as follows.

$$d_{b_k,j} = f(d_{b_k,v_k}) \text{ if } v_k \leq j < v_{k+1} \quad (2-24)$$

$$d_{b_k+i,j} = f(d_{b_k,v_k}, d_{b_k+1,v_k}, \dots, d_{b_k+i,v_k}) \text{ if } v_k \leq j < v_{k+1} \text{ and } b_k + i \leq e_k \quad (2-25)$$

Herein, $f(\bullet)$ is a linear function of the decision variables. Furthermore, the departure times of vehicles $[b_k, e_k]$ at the stops further downstream of the subsequent holding station, say $v_k + m$, will be determined by more decision variables as follows.

$$d_{b_k+i,j} = f(d_{b_k,v_k}, d_{b_k+1,v_k}, \dots, d_{b_k+i,v_k}, d_{[b_{k+1}, e_{k+1}], v_{k+1}}, \dots, d_{[b_{k+m}, e_{k+m}], v_{k+m}}) \quad v_{k+m} \leq j < v_{k+m+1} \quad (2-26)$$

In equation (2-26), if v_{k+m+1} does not exist, it will be substituted by the end terminal of the route.

With the transformation using (2-24) through (2-26), and substituting the linear expression of the vehicle departure times for the counterparts in the objective function, the holding control problem formulation has a general form of:

$$\text{Minimize} \quad Z = F(\bullet) + f(\bullet) \quad (2-27)$$

$$\text{S.T.} \quad g_j(\bullet) \leq C_j \quad \forall j$$

Herein, $g_j(\bullet)$ is a linear function of decision variables; $F(\bullet)$ is a quadratic function, which consists of two quadratic functions, the squared form of a linear function of decision variables; $f(\bullet)$ again is a linear function of the decision variables; C_j is constant; and, j varies from 1 up to twice the number of currently operating vehicles upstream of the most downstream holding station, since each decision variable is subject to two constraints with the form of the inequalities (2-22) and (2-16). Therefore, this problem formulation is a convex problem with a convex objective function and a set of linear constraints. Such a problem can be solved to optimality by many classical techniques. However, the scale of the problem is not necessarily small when the route is long with many stops and many vehicles operating at the same time.

A solution algorithm is developed by decomposing the overall problem, namely the problem of holding vehicles at multiple holding stations, into several two-dimensional

problems. Furthermore, the two-dimensional problem is further decomposed into one-dimensional problems, which eventually can be solved analytically.

Before getting into the details of the algorithm, a proposition about the vehicle overtaking is presented to have more insights about how vehicle overtaking may affect the performance of the holding control decided with the problem formulation without considering vehicle overtaking explicitly.

Proposition 1:

Given h_2 and h_3 as the real headways of vehicle 2 (the control vehicle's first following vehicle) and 3 (second following vehicle) at stop j , respectively, if $h_2 \geq h_3 \cdot \beta_1 \cdot \lambda_j / (1 - \beta_1 \cdot \lambda_j)$ holds, the real objective value is always less than the model objective value on the route segment downstream of where the vehicle overtaking occurs.

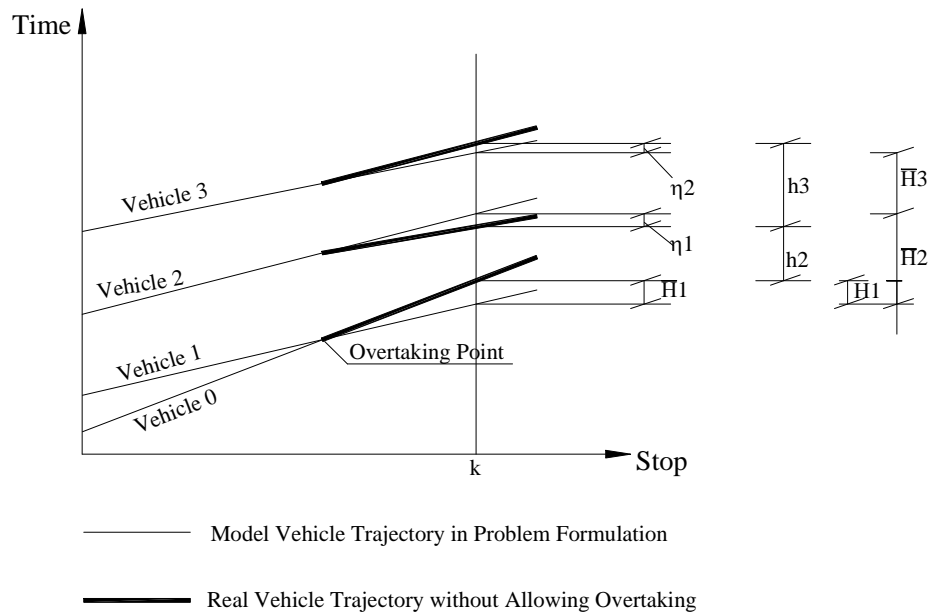


Fig. 2.2. Comparison of Model Trajectory and Real Trajectory

Proof of Proposition 1:

Though the problem formulation for holding control at multiple holding stations cannot avoid vehicle overtaking, it essentially represents the vehicle overtaking as a negative headway. However, the negative headway still contributes positively to the objective since the headway term is always squared in the objective function. On the other hand, the vehicle overtaking may not be allowed in practice. For such a case, if the following vehicle catches up to the leading vehicle, the trajectory of the following vehicle will overlap the leading vehicle's trajectory. On the other hand, if overtaking is allowed, these two vehicles may overtake each other alternately without ever getting very far apart, and thus the two vehicles' trajectories can still be seen as overlapping. Based on these two aspects, a difference exists between the vehicle trajectories as formulated and the real vehicle trajectories when vehicle overtaking does occur.

As shown in the Fig. 2.2, as vehicle 1 (the control vehicle) overtakes vehicle 0 (the boundary vehicle), the trajectory of vehicle 1 will follow the thin line after the overtaking point according to the model. However, the solid line represents the real vehicle trajectories if overtaking is not allowed. Accordingly, $\bar{H}_1, \bar{H}_2, \bar{H}_3$ are defined as the vehicle headways derived from the model formulation, and, in contrast, h_1, h_2, h_3 as the real vehicle headways ($h_1 = 0$ due to the trajectory overlap). For each stop k downstream of where the overtaking occurs, such headway patterns and the magnitude of η_1, η_2 (the difference between the model trajectory and real trajectory) can be easily seen and derived by mathematical induction based on equation (2-13) as:

$$\eta_1 = \beta_1 \cdot \lambda_j / (1 - \beta_1 \cdot \lambda_j) \cdot \bar{H}_1 \quad \text{and} \quad \eta_2 = \beta_1 \cdot \lambda_j / (1 - \beta_1 \cdot \lambda_j) \cdot \eta_1 \quad (2-28)$$

If only vehicle 0, 1 and 2 are considered, it can be obviously seen graphically that the real objective value with only the three vehicles is less than the model value. As vehicle 3 is included in, the model objective value can be expressed as:

$$\overline{H}_1^2 + (h_2 + \eta_1 + \overline{H}_1)^2 + (h_3 - \eta_1 - \eta_2)^2 = \varpi + h_2^2 + h_3^2 + 2h_2 \cdot (\eta_1 + \overline{H}_1) - 2h_3 \cdot (\eta_1 + \eta_2) \quad (2-29)$$

Herein, ϖ is a positive value. As can be seen directly from equation (2-28),

$$2h_2 \cdot (\eta_1 + \overline{H}_1) - 2h_3 \cdot (\eta_1 + \eta_2) = 2h_2 \cdot (\eta_1 + \overline{H}_1) - 2h_3 \cdot (\eta_1 + \overline{H}_1) \cdot \beta_1 \cdot \lambda_j / (1 - \beta_1 \cdot \lambda_j) \quad (2-30)$$

Therefore, if

$$h_2 \geq h_3 \cdot \beta_1 \cdot \lambda_j / (1 - \beta_1 \cdot \lambda_j), \quad (2-31)$$

It is always true that the model objective value is larger than the real objective value. $\beta_1 \cdot \lambda_j / (1 - \beta_1 \cdot \lambda_j)$ is actually a very small number generally on the order of 0.1 or less. h_2 in the equation is essentially the departure time difference between vehicle 2 and vehicle 0 at stop j . Therefore, unless the vehicle trajectory pattern is extreme, inequality (2-31) holds, and it is true that the model objective value is larger than the real objective value for the four vehicle case. Based on the same argument, it can be easily inferred that even when more vehicles are included into consideration, the proposition is still true. ■

Since the model formulation proposed in this chapter does not explicitly include overtaking, this proposition states that a solution to the model formulation will have a larger (or higher) objective value than would occur if overtaking were included. In this way, the model formulation for the problem of holding vehicles at multiple stations as

proposed is more conservative, in that it will recommend a larger overall objective value from holding actions than would be seen if overtaking were included explicitly. This may imply if one sees the service improvement from the holding control, the improvement is even more significant than what can be seen from the problem solution if vehicle overtaking occurs.

The following sub-sections start with the simplest problem, holding a single vehicle at a single holding station, then gradually add complexity to the problem to achieve the full problem solution for the problem of holding multiple vehicles at multiple stations.

Holding a Single Vehicle at a Single Holding Station (PSS)

The complexity of the holding problem lies in the fact that any adjustment to the departure time of one particular vehicle at a stop will in turn change this vehicle's trajectory downstream of the stop, and also affect many following vehicles' trajectories. Therefore, while considering holding one particular vehicle, it is also necessary to account for the following vehicles (impacted vehicles), as well as one leading vehicle, which functions as a boundary vehicle on the solution. If we expand the impacted vehicles up to the vehicle O , all vehicles upstream of any one of the holding stations, including all non-dispatched vehicles within the time window of interest, can be simply categorized into two groups:

Holding Group: the vehicles within this group will be considered for holding; and,

Non-Holding Group: the vehicles within this group will not be held, but they are assumed to be affected by the holding control decision for the holding group.

For the problem of holding one vehicle at a single holding station, only one control vehicle is within the holding group, and the non-holding group consists of all other impacted vehicles up to vehicle O , including the boundary vehicle immediately ahead of the control vehicle. Accordingly, the *PSS* can be seen as a one-dimensional problem due to the single decision variable.

Though presented for the overall problem, problem formulation (2-27) and the general vehicle departure time equations (2-24) through (2-26) can still apply to the *PSS* problem. Obviously, all impacted vehicle trajectories downstream of the holding station can be derived with the equations of the same form as (2-26). An univariate convex problem can be easily solved by many techniques. However, since the *PSS* problem solution is the core of the overall heuristic, an analytical solution is employed to solve the *PSS* problem in this study. The problem is solved exactly using an analytical method. The analytical solution can be obtained with the differential equation regarding the control vehicle's departure time at the holding station, and the global optimal solution to *PSS* is either at the analytical solution, or at one of the extreme points.

Holding Multiple Vehicles at a Single Holding Station (PMS)

As more than one vehicle is included into the holding group for a single holding station, the holding problem upgrades to the *PMS* problem. For a particular holding station v_k , $[b_k, e_k]$ constitutes the holding group, and all vehicles following the vehicle e_k up to the vehicle O , along with the vehicle $b_k - 1$, make up the non-holding group.

Equation (2-22) basically says that the decision variables are dependent of each other ($d_{i,j}$ is dependent on $d_{i-1,j}$). Therefore, for the general form of the problem (2-27),

each of the linear constraints may include multiple decision variables. With no intention beyond the purpose of making the concepts clearer and of simplifying the problem solution, some special treatment of the transit holding station is implemented.

By inspecting equation (2-15), theoretically, the holding control can be realized by either postponing the vehicle departure time for $H_{i,j}$ at the holding station, or delaying the vehicle arrival time by an equivalent amount of time $H_{i,j} \cdot (1 - \beta_1 \cdot \lambda_j)$.

If the holding control is considered as a means to delay the vehicle's arrival time, the holding problem becomes an equivalent problem of how to optimize the vehicle arrival time at the holding station. However, delaying one vehicle's arrival time at a stop would not affect the arrival times of other impacted vehicles. To clarify this idea, a simple treatment on the route and station is made by introducing a dummy stop to separate the vehicle arrival process and departure process at each holding station. This dummy stop is inserted just upstream of the holding station to represent the vehicle arrival process, and will function as the surrogate of the original holding station, as shown in Fig. 2.3.

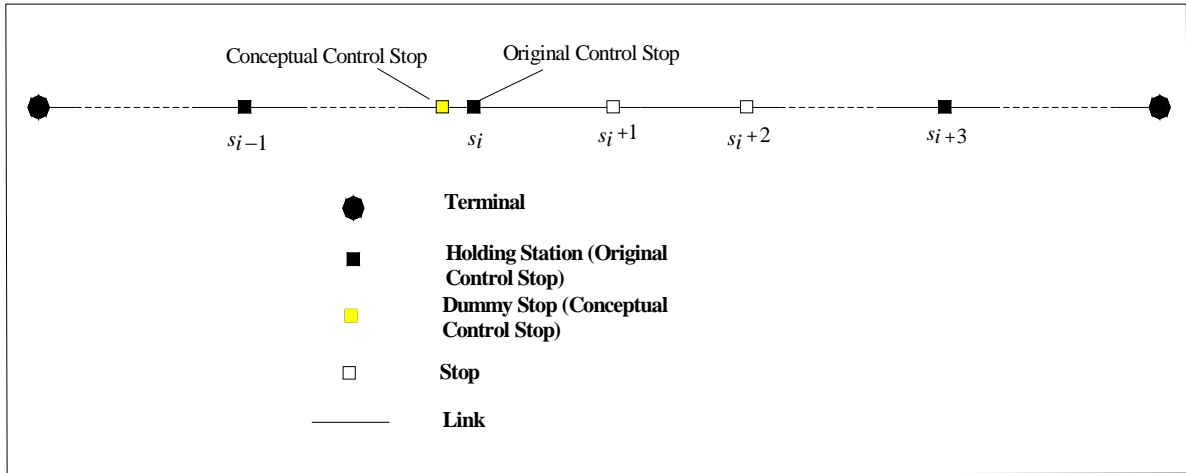


Fig. 2.3. Typical Transit Route with Multiple Holding Stations (Holding Control)

With this ‘physical’ treatment,

- The original holding station becomes a regular stop. Furthermore, it is assumed that all passenger boarding and alighting still occurs at the original control stop, with none at the dummy stop. The dummy link connecting the dummy stop and the original holding station has a length of zero;
- The dummy stop becomes the holding station, at which the vehicle arrival times are exactly identical to the departure times if no control is implemented. The vehicle ready-for-departure times (i.e. arrival times) at the dummy stop are thus independent of each other;
- The transit route operating process (the process of propagating arrival and departure times at downstream stops) remains the same as would be without any treatment; and,
- The control vehicles’ holding times are independent of each other, since no boarding and alighting occurs at the dummy stop and the interdependency of the

holding times has to be realized through the passenger boarding and alighting process, as one may see from equation (2-14).

However, it must be pointed out that the fourth observation only holds when vehicle overtaking will not occur, because the dummy stop treatment still cannot prevent the occurrence of the vehicle overtaking at the original holding stations. The dummy stop treatment itself does not actually change the nature of the problem, but adds a little more conceptual clarity. If the holding control at the dummy stop does not lead to vehicle overtaking at the original holding station, the holding times are essentially independent of each other at the original holding station even without the dummy stop treatment. However, as argued in Proposition 1, the vehicle overtaking rarely occurs in the given problem context.

With all treatments introduced above, the *PMS* problem still possesses a convex objective function with linear constraints. However, with the dummy stop treatment, the decision variables are independent of each other within the constraints. With this additional characteristic, a solution algorithm for the *PMS* problem is developed.

The solution algorithm basically employs an iterative process to decompose the *PMS* problem into separate *PSS* problems. The algorithm tries to hold only one vehicle which can reduce the overall objective value the most at each iteration. It finally converges at the point that no holding control over any single vehicle can reduce the objective value further.

In more detail, the solution algorithm (H1) can be elaborated into following steps.

Step 1: Initialization.

Set a threshold for algorithm convergence;

Predict the current departure times at the holding station for all vehicles in the holding group using equation (2-4), and set these current departure times as the *Departure Time Lower Limit (DTLL)*. At the same time, the DTLL will also function as the *Departure Time Upper Limit (DTUL)* for the preceding vehicles;

Set the current departure times as the solution I ;

Compute the total passenger cost (objective value) based on solution I , and set this passenger cost as the *Previous Passenger Cost (PPC)*;

Set $n = 2$.

Step 2: For iteration n :

Optimize the departure time for each individual vehicle within the holding group $[b_k, e_k]$ by solving the PSS problem analytically for each vehicle sequentially in the order that the vehicle is dispatched, with all other vehicles' departure times remaining the same as in solution $n-1$.

Step 3: If all optimized vehicle departure times from step 2 are earlier than or the same as in solution $n-1$, go to step 5; otherwise, identify the vehicle whose optimized departure time leads to the minimum total passenger cost among all optimized departure times;

Update the corresponding vehicle departure time in solution $n-1$ with this new vehicle departure time; and, set the minimum total passenger cost as the *Current Passenger Cost (CPC)*;

Step 4: Check the proximity of CPC to PPC . If CPC is within the convergence threshold of PPC , go to Step 5; otherwise, $PPC = CPC$, $n = n+1$, and go to Step 2;

Step 5: Stop.

Following these steps, in each iteration, each vehicle's departure time is optimized conditionally on other vehicles' departure times inherited from the last iteration, and H1 captures the most efficient vehicle's departure time in terms of the overall passenger waiting time reduction to conclude the iteration. The interacting behavior between all control vehicles' departure times is hence realized by consecutive iterations.

Based on the algorithm above, proposition 2 is developed.

Proposition 2:

H1 can solve the problem PMS to optimality.

Proof of Proposition 2:

- First, the solution from the algorithm is a KKT (Karush-Kuhn-Tucker) point.

For the minimization problem, as the solution satisfies the following conditions, it is a KKT point.

$$\frac{\partial f(\bar{x})}{\partial x_i} + \sum_{j=1}^m \lambda_j \cdot \frac{\partial g_j(\bar{x})}{\partial x_i} = 0 \quad (2-32)$$

$$\lambda_j \cdot [C_j - g_j(\bar{x})] = 0 \quad j = 1, \dots, m$$

The H1 algorithm stops at where the objective cannot be improved by changing the value of any single decision variable x_i . Based on this, there exist two cases: (1) non-binding constraints, and (2) binding constraints regarding x_i .

For case (1), $\frac{\partial f(\bar{x})}{\partial x_i}$ must be zero. Otherwise, the objective can always be improved by changing x_i . The KKT condition is satisfied by setting all multipliers λ_j for the constraints with x_i as zeros.

For case (2), $\frac{\partial f(\bar{x})}{\partial x_i}$ can be either negative or positive. Independence of the decision variables in the constraints through the dummy stop treatment implies only two constraints, say n^{th} and l^{th} constraints, in the general form of the problem formulation as presented by (2-17) can have and only have x_i ,

$$x_i \leq b_n, \text{ and} \quad (2-33)$$

$$-x_i \leq -b_l \text{ Herein } b_n > b_l \quad (2-34)$$

b_n, b_l are both positive and only one of the two constraints above can be binding. If (2-33) is binding, then $\frac{\partial f(\bar{x})}{\partial x_i}$ has to be negative. Otherwise, the objective improvement

can be achieved by decreasing x_i . Meanwhile, $\frac{\partial g_j(\bar{x})}{\partial x_i}$ is 1 only when $j = n$, otherwise,

it equals zero. Therefore, the KKT conditions can be satisfied by simply setting λ_n as

$-\frac{\partial f(\bar{x})}{\partial x_i}$, which is a positive value, and λ_l as zero. If (2-34) is binding, it can be proved

similarly that the KKT condition can be satisfied by setting λ_l as $\frac{\partial f(\bar{x})}{\partial x_i}$ and λ_n as zero.

- KKT conditions are necessary and sufficient for local optimality.
- For a convex problem, the local minimum is also the global minimum.

Therefore, proposition 2 is true and the solution from H1 can solve the PMS problem to optimality.

Holding Multiple Vehicles at Multiple Holding Stations (PMM)

As a final extension of the previous two problems, the full version of the holding problem is to hold multiple vehicles at multiple holding stations (*PMM*). As introduced earlier, holding multiple vehicles at multiple holding stations does not consider holding each vehicle at all downstream holding stations in one decision-making cycle; instead, each vehicle is only considered to be held at the immediate downstream holding station. However, even with such a simplification, the problem becomes more complicated since the departure time (d_{e_k, v_k}) of the last control vehicle (e_k) of the downstream holding station (v_k) is always dependent on the departure time ($d_{b_{k-1}, v_{k-1}}$) of the first control vehicle (b_{k-1}) of its immediately upstream holding station (v_{k-1}), and vice-versa. In considering this fact, a heuristic (H2) is developed to search for a solution that can approximate the global optimum to the full problem.

This heuristic decomposes the overall problem into *PMS* problems first, then again uses iterations to mimic the interaction among the control vehicles (b_{k-1} and e_k) at different holding stations. In more detail, the heuristic (H2) is described below.

Step 1: Initialization.

Set a threshold for algorithm convergence;

Check all en-route operating vehicles. Set $[b_k, e_k]$ as the *Holding Group* and all following vehicles up to vehicle O , along with vehicle e_{k+1} , in the *Non-Holding Group*, for each holding station v_k ;

Predict all en-route vehicles' trajectories without holding, and set all vehicles' departure times at the corresponding holding stations together as the solution I ;

Compute the total passenger cost (objective value) based on solution I , and set it as the *Previous Passenger Cost (PPC)*;

Set $n = 2$;

Step 2: For iteration n .

for $i = M$ *to* 1

Solve the single holding station problem *PMS* by using H1 for holding station v_k , based on the solution $n-1$.

Update the corresponding parts in the solution $n-1$ with the new optimized departure times for $[b_k, e_k]$ at holding station v_k .

end

Step 3: Solution $n =$ Solution $n-1$;

Compute the total passenger cost based on the solution n , and set it as the *Current Passenger Cost (CPC)*;

Compare *CPC* and *PPC*. If *CPC* is within the convergence threshold of *PPC*, go to Step 4; otherwise, $PPC = CPC$, $n = n+1$, go to Step 2.

Step 4: Stop.

Always starting with the most downstream holding station in each iteration at Step 2, the heuristic solves the *PMS* problem for each holding station sequentially in descending order. Though it has been represented in the heuristic description, it is useful to point out again that, when the heuristic solves the *PMS* problem for a particular holding station v_k , all trajectories of the control vehicles belonging to all its upstream holding stations will function as impacted vehicles. Certainly, the trajectories of the boundary vehicle(s) and impacted vehicles affect the solution of the *PMS*, and the revision of these trajectories is just the essence of the iterative process that the heuristic H2 employs. The heuristic eventually converges at the point at which the objective cannot be improved significantly by changing any vehicle's departure time at the corresponding holding station.

Proposition 3:

If no vehicle e_k 's ($i = 1, \dots, M-1$) trajectory is bound by the immediately following vehicle's arrival time, algorithm H2 solves the overall problem to optimality.

Proposition 3 can be proved using a similar method as for Proposition 2.

2.3. Simulation Study

Besides simulating the basic dynamics of the holding control and the vehicle trajectory evolution, and revealing the discrepancy between the real problem and the problem formulated in this chapter, the primary task of the simulation study is to evaluate how the holding strategies applied at a single holding station and multiple holding stations perform.

2.3.1. Scenario Design

Prior to presenting the details of the simulation design, a flowchart (Fig. 2.4) is given to depict how the holding control simulation study is conducted.

As shown in Fig. 2.4, this simulation study for holding control starts by identifying the key factors which may affect the holding control performance significantly. The key factors (both random and deterministic) are further employed to construct simulation scenarios ($R \times D$). For each of these scenarios, a number (N) of common random factor sets are randomly generated. These scenarios are then used together with the holding problem formulation and solution to determine the optimal holding decisions for both the single-station and multiple-station cases. The simulation results, in terms of passenger waiting time, are collected to compare the performance of the two holding control strategies on an aggregate level. More detail of the simulation is given below.

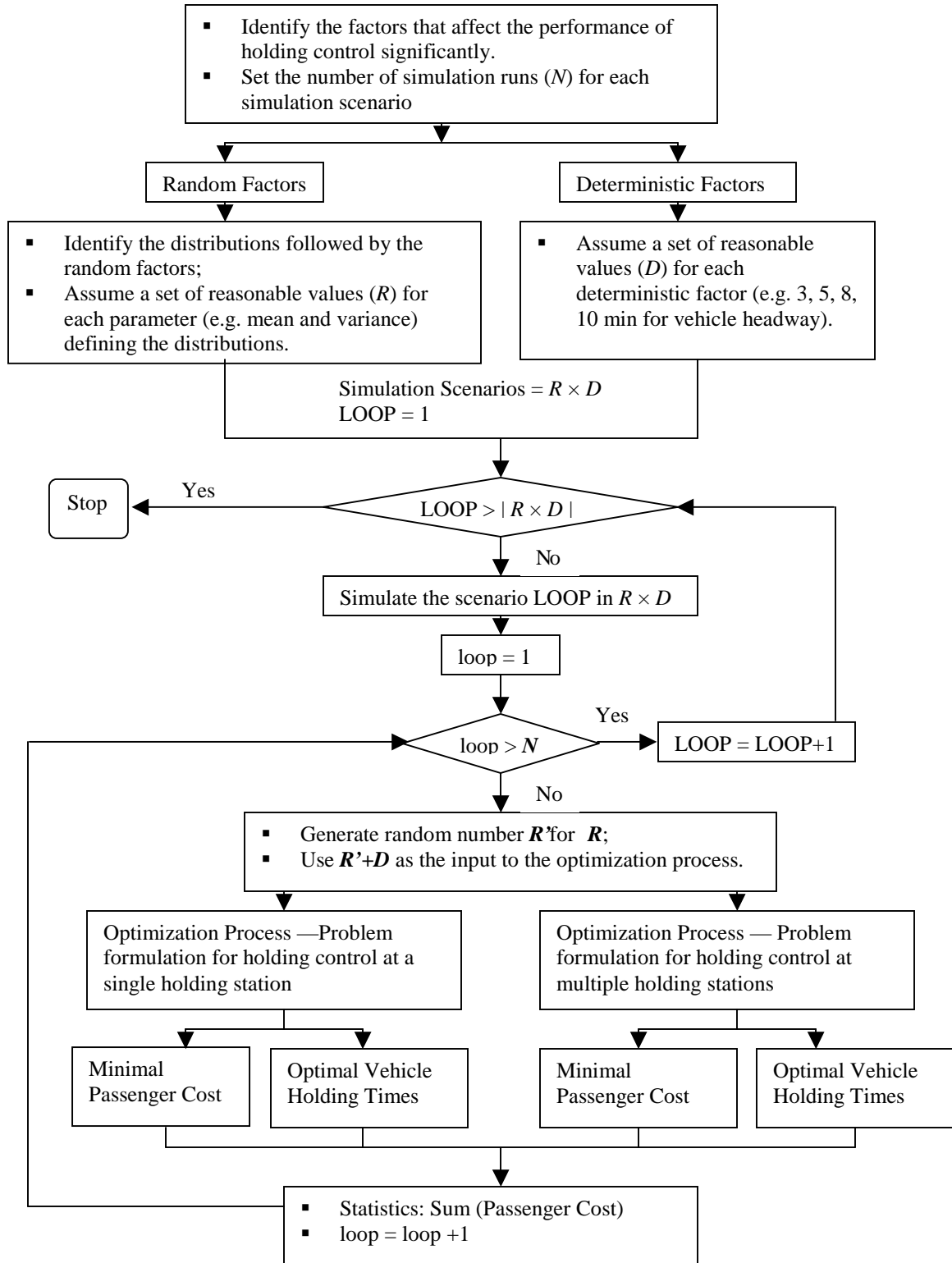


Fig 2.4. Simulation Flowchart – Holding Problem

The vehicle headway variability along a transit route can be attributed to two factors, namely the variability of the vehicle travel times between adjacent stops and vehicle dwell times at stops. The travel times between adjacent stops are generally assumed independent of each other, and are also independent of the vehicle dwell times at stops. Accordingly, the vehicle headway variability contributed solely by the vehicle between-stop travel time variation is additive. However, the portion of the vehicle headway variability resulting from the vehicle dwell time, or more accurately, the passenger boarding and alighting random process, is essentially affected by the variation of travel times. Besides, there is fairly strong correlation between the vehicle dwell time variability at different stops, in that the variability of the vehicle dwell times at upstream stops can result in larger variability at the downstream stops. Furthermore, such a relationship is defined by the service frequency. A service with higher frequency is more likely to have more intense passenger boarding/alighting at the stops, which in turn can result in more vehicle dwell time variability at those stops, and, more importantly, increase the cumulative dwell time variability and travel time variability downstream. For instance, at a certain stop, an arrival lateness of 2 minutes may result in a significantly larger amount of passenger boarding time for a service with a 5 min headway than for a 10 min headway, given that the passenger arrival rate for the former service is approximately twice as much as that for the latter service. Therefore, a service with higher frequency is more likely subject to overall system operation variability than a service with lower

frequency. Accordingly, one may expect that holding control plays a more important role for the service with higher frequency.

Based on these arguments, it is of particular interest in the simulation study of the holding control problem to compare the performance of the multiple-station holding strategy and the single-station holding strategy based on a wide range of scenarios, spanning different service frequencies and other significant factors, including the passenger boarding profiles as well as vehicle travel time variability.

Test Transit Route

The test transit route within this simulation study is designed to have a total of 41 stops (including terminals) approximately evenly distributed, with one-way vehicle trips time of approximately 58 minutes, including the vehicle acceleration, deceleration and door open/close times and other vehicle dwell time components.

Service Frequency

As introduced in the first chapter, this dissertation focuses on the relatively frequent transit service. Therefore, only services with a headway less than or equal to 10 minutes are of the interest in this simulation study. Specifically, the services with four different vehicle headways are chosen: 5, 6, 8 and 10 minutes respectively.

Passenger Boarding Profile Design

As indicated in many previous studies, the desirable location of the control point and the performance of holding control could be affected significantly by the passenger boarding profiles, especially by the location where the peak passenger boarding occurs. Accordingly, to investigate the effects of the passenger boarding profiles on the

performance of holding strategy, three discrete passenger boarding profiles are also designed as follows, with the peak passenger boarding occurring at different locations along the route.

- Pattern 1: Passenger boarding peaks at stops 13 through 15, together accounting for 15 percent of the overall passenger boarding;
- Pattern 2: Similar to passenger boarding profile 1, but exchanging the passenger boarding rates at stops 13 through 15 with those at stops 20 through 22; and,
- Pattern 3: Similar to passenger boarding profile 1, with exchanging the passenger boarding rates at the stops 13 through 15 with those at the stops 29 through 31.

The passenger boarding and loading profiles for the service with a headway of 10 minutes are illustrated in Fig. 2.5. Herein, the on-board passenger loading profiles are depicted assuming the vehicle headways are perfectly even (i.e. exactly 10 minutes) at any point along the route, given an approximately normally distributed passenger O/D pattern between stops.

For the sake of evaluating the effects of the service frequency on the performance of the holding strategy on a common basis, the passenger boarding rate at each stop for the services with the headway other than 10 minutes is inversely proportional to the headway. For instance, for the service with the headway of 5 minutes under passenger boarding profile 1, the passengers boarding rate at each stop will double the corresponding rate in the first diagram of Fig. 2.5. This makes sense in that it basically underlies how the operation headway is determined in practice for a relatively frequent

transit service. Relatively low/high passenger demand is served with relatively long/short headways. Accordingly, based on such a passenger boarding profile design, for all services of different frequencies, the total passenger waiting time will be the same, provided the vehicle dispatching headways can be maintained perfectly along the route. Furthermore, this study design also ensures identical passenger loading profiles for all services of different headways, as shown in Fig. 2.5.

Furthermore, the values shown in Fig. 2.5 are the expected values of probabilistic distributions. In the simulation, passenger arrivals are depicted as a Poisson process with these values as the mean.

Travel Time Variation

In this study, the holding problem is formulated within a context of deterministic service model, which can only approximate the transit operation in the real world. From this perspective, it becomes particularly valuable to also incorporate the inherent stochasticity within the transit operation into the simulation study, to evaluate the impacts on the performance of holding strategy from a variety of stochastic operating factors, among which the travel time variation is of primary interest.

Vehicle travel time has been reported to follow a normal distribution by Lesley (1975), and a lognormal distribution by Andersson and Scalia-Tomba (1981). The lognormal distribution proposed by Andersson and Scalia-Tomba (1981) can also be approximated by the normal and the gamma distributions for appropriate parameters. For the sake of convenience, the vehicle travel time is assumed to be normally distributed in this study. Furthermore, the coefficient of variation (COV) of vehicle travel time on each

segment connecting adjacent stops is assumed to vary between 0.1 and 0.3. Fu and Liu (2003) have employed a typical travel time variation (COV of 0.2), observed in the field, in their study.

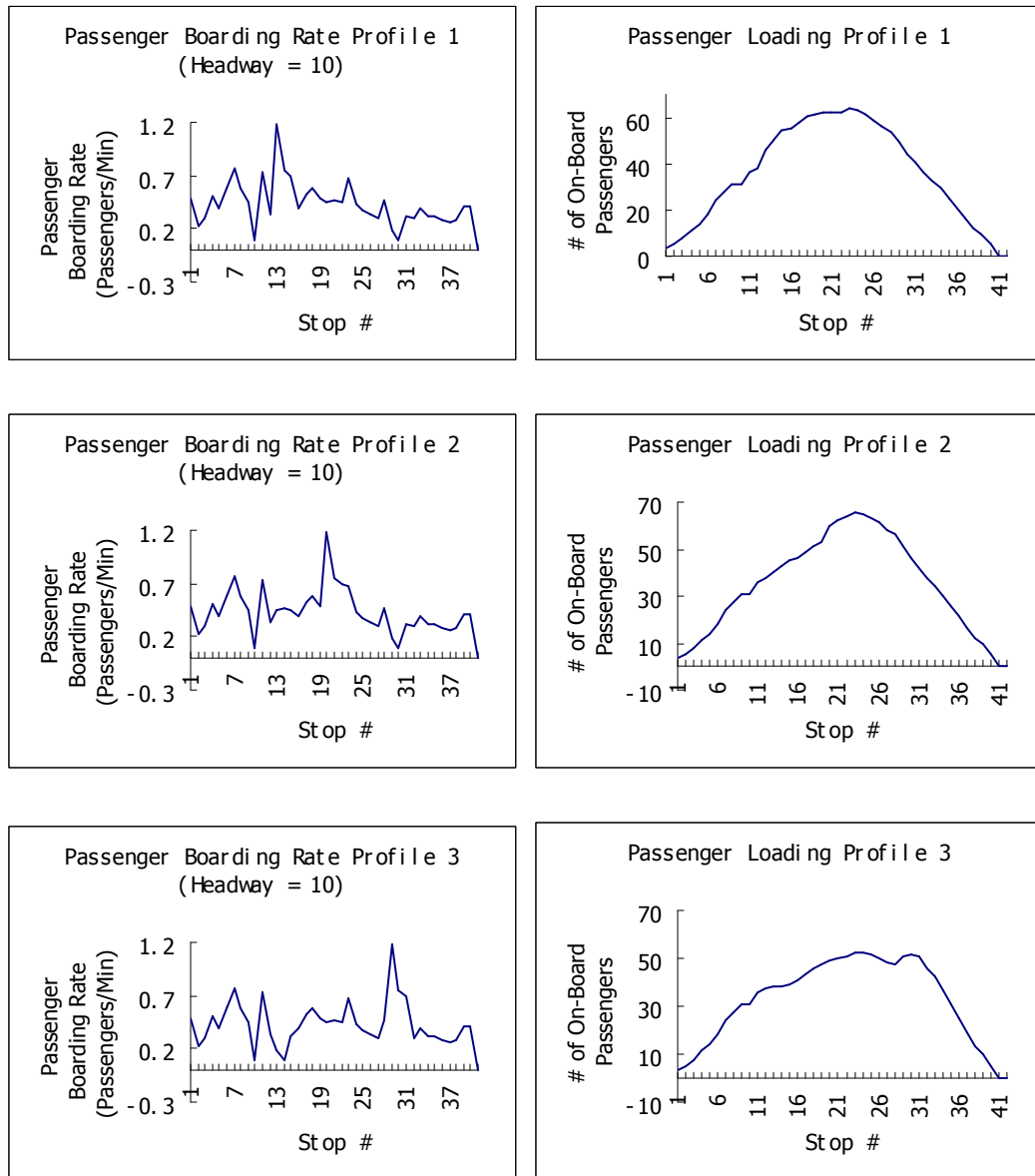


Fig. 2.5 Test Passenger Boarding and Loading Profiles (Holding Control)

Other Operating and Simulation Parameters

Other operating and simulation parameters are given in Table 2.1.

Table 2.1. Other Operating & Simulation Parameters (Holding Control)

Parameters	Values
AVL Data Polling Time Interval (sec)	40
α, β_1, β_2 (sec)	15, 3 and 1.5, respectively
Threshold Cost Value for <i>PMS</i> (Pass.-Min) ¹	20
Threshold Cost Value for <i>PMM</i> with <i>M</i> Holding Stations (Pass.-Min) ¹	20 * M
Time Period (Min) ²	120

Note: 1.Threshold cost values are set for the purpose of checking the convergence of algorithms H1 and H2
 2. It is assumed that the first vehicle is dispatched at time instant 0.

In table 2.1, the simulated AVL data frequency is set the same as in some real systems, e.g. AVL system employed by Sun Tran, Tucson, AZ. Also, in the following chapters in this dissertation, when a simulation study is conducted, the AVL data polling time interval is set as 40 sec. The constant part of vehicle dwell time at each stop, the average passenger boarding time and the average passenger alighting time are set as 15 sec, 3 sec and 1.5 sec respectively. The threshold cost value (for algorithm convergence purposes) for the *PMS* problem is set as 20 passenger-minutes, and $20 * M$ passenger-minutes for the *PMM* problem with *M* holding stations. Furthermore, all vehicle trips and the holding control applied on these trips within a two-hour time window are simulated.

2.3.2. Simulation Design

As introduced in the last sub-section, with 2 holding strategies (holding vehicles at multiple holding stations and at single holding station), 3 passenger boarding profiles, 4 types of services of different headways (5, 6, 8 and 10 minutes respectively), as well as 5 categories of travel time variation level (COV) varying from 0.1 through 0.3 at the increment of 0.05), a total of 120 cases are generated to represent all possible combinations of these four factors. For each specific case (or combination), 2500

simulation runs are conducted, and the average passenger cost reduction for each case is computed accordingly.

For each specific case (or combination), the overall computational burden is basically determined by the service headway, because each simulation run actually simulates all vehicle trips accomplished within the time period of interest, which is two hours. For instance, for the service with a 5 min headway, a total of 24 vehicle trips are simulated. On the contrary, for the service with the headway of 10 minutes, only 12 vehicle trips are simulated. Therefore, the computational effort of the simulation for the former service at least doubles that for the latter service.

As introduced at the beginning of this chapter, for the sake of comparing the performance of the two different holding strategies, namely holding control at multiple holding stations and at a single holding station, CRN-based simulation is conducted in the sense that, 2500 sets of passenger arrivals at each particular stop within the simulation time period and their destinations are generated randomly, and serve as the common input to the 2500 simulation runs for each specific case defined above. Aside from this, for all cases with the common travel time COV, the scenarios share an identical collection of randomly generated travel times between all pairs of adjacent stops. For instance, the scenario of holding at multiple holding stations under passenger boarding profile 2, with a 5 minute headway and a travel time COV of 0.2, shares a common set of travel times with the case of holding at single holding station under passenger boarding profile 3, with a 10 minute headway and a travel time COV of 0.2. With such a CRN-based simulation study, one may easily conclude that, given that the other factors are the same, the

difference in system performance results solely from the system configuration defined by two different holding strategies and other factors, namely the travel time variation, the passenger boarding profile and the service frequency.

Moreover, of particular note is that the passenger alighting is also explicitly incorporated into the simulation of the dynamics of the vehicle trajectory evolution, although it is not actually considered in calculating the vehicle dwell time when the holding control decision is made. In the simulation of the vehicle trajectory evolution dynamics, each passenger alighting incurs a time increment of β_2 to the vehicle dwell time at the stop.

2.3.3. Simulation Results

2.3.3.1. Optimal Holding Station(s)

For either the multiple-station holding strategy or the single-station holding strategy, the performance of holding control can be improved by judiciously choosing the holding station(s).

Optimal Single Holding Station

In this particular study, the optimal single holding station is determined by an exhaustive search, using the simulation to examine how the holding strategy performs at each stop, then choosing the one with the largest passenger cost reduction as the optimal holding station. Herein, 100 CRN-based simulation runs are conducted for each case for each stop, and the simulation results show that:

- For passenger boarding profile 1, the optimal holding station is consistently located at either stop 12 or stop 13 for all combinations of headway and travel time COV.
- For passenger boarding profile 2, the optimal holding station varies between stop 14 and stop 18 with different headways and travel time COVs, with the largest difference in the passenger cost reduction of no more than 5 percent.
- For passenger boarding profile 3, the optimal holding station varies between stop 15 and stop 19 with different headways and travel time COVs, with the difference in the total passenger waiting time saving of at most 4 percent.

In the simulation results introduced later, stops 13, 16 and 17 are employed as the optimal holding stations for passenger boarding profiles 1, 2 and 3, respectively.

According to the simulation results, the optimal holding position tends to move more downstream as the passenger demand peaks further downstream. This is primarily because the passenger waiting at the stops nearer to the holding station can benefit more from the holding control, and hence it is advantageous to place the holding point closer to the peak passenger boarding stops. However, this does not necessarily imply that the best location for deploying a control point is always immediately upstream of the peak passenger boarding stops. As indicated by the simulation results, the optimal holding station could be located significantly upstream of the peak passenger boarding stops for some cases. For instance, under boarding profile 3, the peak passenger boarding occurs at stops 29 through 31. However, the optimal holding station is at stop 17, which is 12 stops upstream of the nearest peak passenger boarding stop.

The observation above can be primarily concluded as the effect from an underlying tradeoff between the desirability of placing a holding station near a group of peak passenger boarding stops and the potential for the holding control to regularize the headway downstream. Headway variability tends to increase at downstream stops, and the larger headway variability tends to reduce the effects of the holding control. Therefore, although the passengers boarding at the peak stops desire that the holding station is placed immediately upstream of the peak stops, the increased system operation variability as vehicle proceeds to the peak stops may greatly discount the capability that the holding control reduces the headway variance. Furthermore, as the peak passenger boarding occurs more downstream, this tradeoff more likely influences the selection of the optimal holding station.

Moreover, the simulation results suggest that the location of the optimal holding point is not very sensitive to the service headway (varying from 5 through 10 minutes) and travel time variability (COV varies from 0.1 through 0.3). However, it can still be observed that the optimal holding point tends to be located more upstream for the longer headway case. This is simply because the vehicle headways are less variable (i.e. COV of headway is smaller) for a service of longer headway. The effects from a holding control can hence be maintained further downstream of the holding point without incurring any significant escalation of the headway variability. Accordingly, the objective, as in equation (2-12), generally drives the location of optimal holding station to be more upstream to regularize the downstream headways earlier so as to benefit more passengers at more downstream stops.

Fig. 2.6 gives a good illustration how the headways evolve along the route with and without holding control at different control points for the services with the headway of 10 minutes and 6 minutes respectively.

Basically, Fig. 2.6.1 and Fig. 2.6.4 show that, for the service with the headway of 6 minutes, it is generally more difficult to maintain even headways as vehicles proceed further downstream than for the service with a headway of 10 minutes. Furthermore, for the service with the headway of 6 minutes, holding control at a certain stop can only regularize the vehicle headways for a limited number of downstream stops. For instance, one may see from Fig. 2.6.3 that, even after the holding control has been implemented at stop 13, the vehicle headways can become fairly uneven when vehicles proceed to stop 20, where passenger boarding starts peaking. This underlies why the optimal holding position for this particular case is located at stop 17, as shown in Fig. 2.6.2, which can essentially lead to more regularized headways at the peak passenger boarding stops. On the contrary, for the service with the headway of 10 minutes, the effects of the holding control at stop 13 can be maintained at most of the downstream stops. One may see from Fig. 2.6.5 that the vehicle headways are still acceptable even at the route terminal, and this may further imply that, with relatively long headways, placing the holding point more upstream may result in more passenger cost reduction due to the increased number of downstream stops where the vehicle headway distribution can be improved by the holding control.

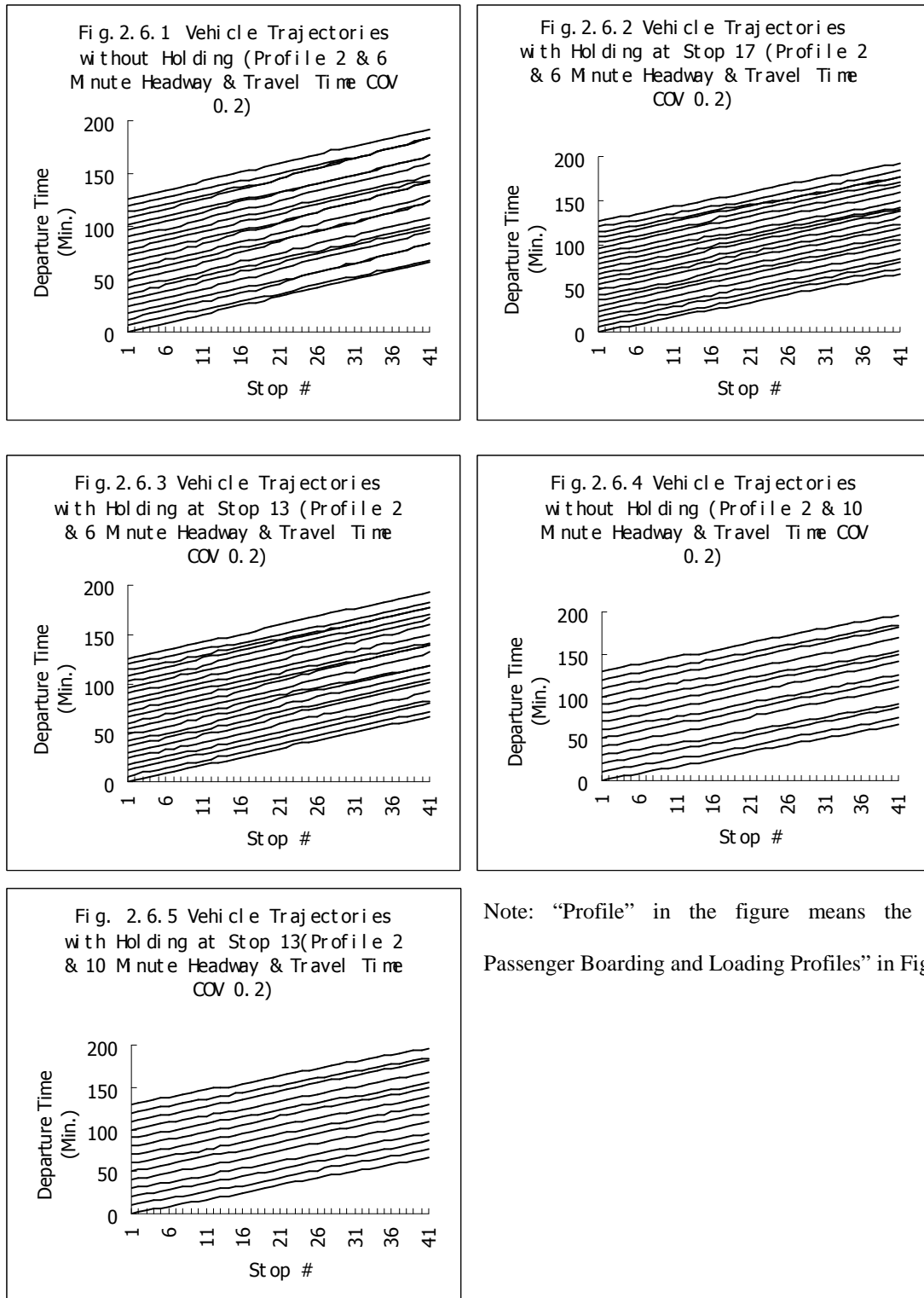


Fig. 2.6. Illustration of Vehicle Trajectories under a Variety of Conditions (Holding Control)

Multiple Holding Stations

For the case of multiple holding stations, no existing analytical method is available to determine the optimal location of the multiple holding stations. Hence, one can resort to an exhaustive search based on simulation. However, for a transit route of regular size, to determine the optimal location of multiple holding stations could be computationally prohibitive. For instance, for a transit route with 50 stops, to determine the optimal location of three holding stations implies a search among 117600 ($50 * 49 * 48$) different stop combinations. Therefore, in this study, three holding stations at stops 11, 21 and 31 respectively are arbitrarily chosen for the holding control across all cases in the simulation study. This forms a strong case when comparing the performance of holding control at multiple holding stations and holding control at one single optimal holding station, in the sense that, the holding control at an optimally chosen set of holding stations may lead to better strategy performance than suggested by the results presented here.

2.3.3.2 Strategy Performance Comparison

Implementing the holding control based on either multiple holding stations or a single holding station can result in a certain amount of passenger cost reduction. The strategy performance comparison between holding control at multiple holding stations and holding control at a single holding station, under different combinations of the passenger boarding profiles, the service frequency and the vehicle travel time variation, are shown in Tables 2.2 through 2.13.

Table 2.2. Holding Strategy Performance Comparison (Profile 1 & Headway 5 Min)

Travel Time COV	Total Passenger Cost Without Holding (Pass.Min)	Holding at Multiple Holding Stations		Holding at a Single Holding Station ¹		Passenger Cost Reduction Difference Between Strategies (Multiple – Single)	
		Passenger Cost Reduction (Pass.Min)	Percentage Passenger Cost Reduction	Passenger Cost Reduction (Pass.Min)	Percentage Passenger Cost Reduction	Absolute Difference (Pass.Min)	Percentage Difference
0.1	15323	3228	21.10%	2703	17.60%	525	20%
0.15	16387	3959	24.20%	3211	19.60%	748	24%
0.2	17510	4688	26.80%	3699	21.10%	989	27%
0.25	18519	5304	28.60%	4101	22.10%	1203	29%
0.3	19348	5776	29.90%	4407	22.80%	1369	31%

Note: 1. The single holding station in the table refers to the optimal holding station specifically to each different passenger boarding profile. Also, this applies to Tables 2.3 through 2.13.

Table 2.3. Holding Strategy Performance Comparison (Profile 1 & Headway 6 Min)

Travel Time COV	Total Passenger Cost Without Holding (Pass.Min)	Holding at Multiple Holding Stations		Holding at a Single Holding Station ¹		Passenger Cost Reduction Difference Between Strategies (Multiple – Single)	
		Passenger Cost Reduction (Pass.Min)	Percentage Passenger Cost Reduction	Passenger Cost Reduction (Pass.Min)	Percentage Passenger Cost Reduction	Absolute Difference (Pass.Min)	Percentage Difference
0.1	13580	1686	12.40%	1453	10.70%	233	16%
0.15	14327	2190	15.30%	1826	12.70%	364	20%
0.2	15150	2726	18.00%	2194	14.50%	532	24%
0.25	15885	3163	19.90%	2485	15.70%	678	27%
0.3	16479	3499	21.20%	2724	16.50%	775	28%

Table 2.4. Holding Strategy Performance Comparison (Profile 1 & Headway 8 Min)

Travel Time COV	Total Passenger Cost Without Holding (Pass.Min)	Holding at Multiple Holding Stations		Holding at a Single Holding Station ¹		Passenger Cost Reduction Difference Between Strategies (Multiple – Single)	
		Passenger Cost Reduction (Pass.Min)	Percentage Passenger Cost Reduction	Passenger Cost Reduction (Pass.Min)	Percentage Passenger Cost Reduction	Absolute Difference (Pass.Min)	Percentage Difference
0.1	12285	522	4.30%	459	3.70%	63	14%
0.15	12686	779	6.10%	668	5.30%	111	17%
0.2	13141	1063	8.10%	884	6.70%	179	20%
0.25	13576	1325	9.80%	1074	7.90%	251	23%
0.3	13952	1546	11.10%	1225	8.80%	321	26%

Table 2.5. Holding Strategy Performance Comparison (Profile 1 & Headway 10 Min)

Travel Time COV	Total Passenger Cost Without Holding (Pass.Min)	Holding at Multiple Holding Stations		Holding at a Single Holding Station ¹		Passenger Cost Reduction Difference Between Strategies (Multiple – Single)	
		Passenger Cost Reduction (Pass.Min)	Percentage Passenger Cost Reduction	Passenger Cost Reduction (Pass.Min)	Percentage Passenger Cost Reduction	Absolute Difference (Pass.Min)	Percentage Difference
0.1	11949	158	1.30%	140	1.20%	18	13%
0.15	12164	274	2.20%	237	1.90%	37	16%
0.2	12434	423	3.40%	358	2.90%	65	18%
0.25	12709	576	4.50%	481	3.80%	95	20%
0.3	12949	705	5.40%	576	4.50%	129	22%

Table 2.6. Holding Strategy Performance Comparison (Profile 2 & Headway 5 Min)

Travel Time COV	Total Passenger Cost Without Holding (Pass.Min)	Holding at Multiple Holding Stations		Holding at a Single Holding Station ¹		Passenger Cost Reduction Difference Between Strategies (Multiple – Single)	
		Passenger Cost Reduction (Pass.Min)	Percentage Passenger Cost Reduction	Passenger Cost Reduction (Pass.Min)	Percentage Passenger Cost Reduction	Absolute Difference (Pass.Min)	Percentage Difference
0.1	15416	3313	21.50%	2721	17.70%	592	22%
0.15	16512	4052	24.50%	3245	19.70%	807	25%
0.2	17675	4800	27.20%	3761	21.30%	1039	28%
0.25	18715	5427	29.00%	4180	22.30%	1247	30%
0.3	19567	5918	30.20%	4500	23.00%	1418	32%

Table 2.7. Holding Strategy Performance Comparison (Profile 2 & Headway 6 Min)

Travel Time COV	Total Passenger Cost Without Holding (Pass.Min)	Holding at Multiple Holding Stations		Holding at a Single Holding Station ¹		Passenger Cost Reduction Difference Between Strategies (Multiple – Single)	
		Passenger Cost Reduction (Pass.Min)	Percentage Passenger Cost Reduction	Passenger Cost Reduction (Pass.Min)	Percentage Passenger Cost Reduction	Absolute Difference (Pass.Min)	Percentage Difference
0.1	13678	1769	12.90%	1491	10.90%	278	19%
0.15	14431	2268	15.70%	1870	13.00%	398	21%
0.2	15246	2781	18.20%	2240	14.70%	541	24%
0.25	15993	3231	20.20%	2551	15.90%	680	27%
0.3	16614	3586	21.60%	2790	16.80%	796	29%

Table 2.8. Holding Strategy Performance Comparison (Profile 2 & Headway 8 Min)

Travel Time COV	Total Passenger Cost Without Holding (Pass.Min)	Holding at Multiple Holding Stations		Holding at a Single Holding Station ¹		Passenger Cost Reduction Difference Between Strategies (Multiple – Single)	
		Passenger Cost Reduction (Pass.Min)	Percentage Passenger Cost Reduction	Passenger Cost Reduction (Pass.Min)	Percentage Passenger Cost Reduction	Absolute Difference (Pass.Min)	Percentage Difference
0.1	12321	543	4.40%	455	3.70%	88	19%
0.15	12748	820	6.40%	678	5.30%	142	21%
0.2	13225	1116	8.40%	912	6.90%	204	22%
0.25	13685	1399	10.20%	1121	8.20%	278	25%
0.3	14081	1632	11.60%	1290	9.20%	342	27%

Table 2.9. Holding Strategy Performance Comparison (Profile 2 & Headway 10 Min)

Travel Time COV	Total Passenger Cost Without Holding (Pass.Min)	Holding at Multiple Holding Stations		Holding at a Single Holding Station ¹		Passenger Cost Reduction Difference Between Strategies (Multiple – Single)	
		Passenger Cost Reduction (Pass.Min)	Percentage Passenger Cost Reduction	Passenger Cost Reduction (Pass.Min)	Percentage Passenger Cost Reduction	Absolute Difference (Pass.Min)	Percentage Difference
0.1	11961	174	1.50%	139	1.20%	35	15%
0.15	12186	296	2.40%	239	2.00%	57	24%
0.2	12472	455	3.70%	373	3.00%	82	22%
0.25	12760	617	4.80%	504	4.00%	113	22%
0.3	13012	757	5.80%	608	4.70%	149	25%

Table 2.10. Holding Strategy Performance Comparison (Profile 3 & Headway 5 Min)

Travel Time COV	Total Passenger Cost Without Holding (Pass.Min)	Holding at Multiple Holding Stations		Holding at a Single Holding Station ¹		Passenger Cost Reduction Difference Between Strategies (Multiple – Single)	
		Passenger Cost Reduction (Pass.Min)	Percentage Passenger Cost Reduction	Passenger Cost Reduction (Pass.Min)	Percentage Passenger Cost Reduction	Absolute Difference (Pass.Min)	Percentage Difference
0.1	15494	3407	22.00%	2800	18.10%	607	22%
0.15	16625	4193	25.20%	3335	20.10%	858	26%
0.2	17796	4954	27.80%	3836	21.60%	1118	29%
0.25	18861	5599	29.70%	4256	22.60%	1343	32%
0.3	19722	6097	30.90%	4576	23.20%	1521	33%

Table 2.11. Holding Strategy Performance Comparison (Profile 3 & Headway 6 Min)

Travel Time COV	Total Passenger Cost Without Holding (Pass.Min)	Holding at Multiple Holding Stations		Holding at a Single Holding Station ¹		Passenger Cost Reduction Difference Between Strategies (Multiple – Single)	
		Passenger Cost Reduction (Pass.Min)	Percentage Passenger Cost Reduction	Passenger Cost Reduction (Pass.Min)	Percentage Passenger Cost Reduction	Absolute Difference (Pass.Min)	Percentage Difference
0.1	13728	1843	13.40%	1552	11.30%	291	19%
0.15	14525	2384	16.40%	1947	13.40%	437	22%
0.2	15384	2940	19.10%	2334	15.20%	606	26%
0.25	16160	3419	21.20%	2653	16.40%	766	29%
0.3	16807	3804	22.60%	2900	17.30%	904	31%

Table 2.12. Holding Strategy Performance Comparison (Profile 3 & Headway 8 Min)

Travel Time COV	Total Passenger Cost Without Holding (Pass.Min)	Holding at Multiple Holding Stations		Holding at a Single Holding Station ¹		Passenger Cost Reduction Difference Between Strategies (Multiple – Single)	
		Passenger Cost Reduction (Pass.Min)	Percentage Passenger Cost Reduction	Passenger Cost Reduction (Pass.Min)	Percentage Passenger Cost Reduction	Absolute Difference (Pass.Min)	Percentage Difference
0.1	12354	598	4.80%	503	4.10%	95	19%
0.15	12804	896	7.00%	743	5.80%	153	21%
0.2	13307	1221	9.20%	990	7.40%	231	23%
0.25	13778	1510	11.00%	1200	8.70%	310	26%
0.3	14186	1757	12.40%	1375	9.70%	382	28%

Table 2.13. Holding Strategy Performance Comparison (Profile 3 & Headway 10 Min)

Travel Time COV	Total Passenger Cost Without Holding (Pass.Min)	Holding at Multiple Holding Stations		Holding at a Single Holding Station ¹		Passenger Cost Reduction Difference Between Strategies (Multiple – Single)	
		Passenger Cost Reduction (Pass.Min)	Percentage Passenger Cost Reduction	Passenger Cost Reduction (Pass.Min)	Percentage Passenger Cost Reduction	Absolute Difference (Pass.Min)	Percentage Difference
0.1	11986	206	1.70%	164	1.40%	10	6%
0.15	12228	343	2.80%	277	2.30%	19	7%
0.2	12522	514	4.10%	416	3.30%	39	9%
0.25	12822	683	5.30%	551	4.30%	66	12%
0.3	13088	836	6.40%	666	5.10%	91	14%

According to Tables 2.2 through 2.13, a holding strategy based on either multiple holding stations or a single holding station can consistently reduce the total passenger cost, and the performance of the holding strategy is relatively independent of where the peak passenger boarding occurs.

Significantly, more passenger waiting time reduction from the multiple-station holding control over single-station holding control can be seen across all cases, especially as the service frequency is relatively high and vehicle travel time variation is large. For instance, under passenger boarding profile 3, for the case of a 5 minute headway and a travel time COV of 0.3 (Table 2.10), the holding control at multiple holding stations can result in 33 percent more passenger waiting time saving over the holding control at a single holding station.

Holding control can result in more passenger cost reduction under the condition where service frequency is high and travel time variability is large. Under passenger boarding profile 3, for the service with a 5 minute headway (Table 2.10), a total passenger cost reduction of more than 100 passenger hours can be achieved by holding vehicles at multiple holding stations, and more than 75 passenger hours by holding vehicles at a single holding station, with a travel time COV of 0.3. On the contrary, for the case of the passenger boarding profile 3, the service with a 10 minute headway and the travel time COV of 0.1 (Table 2.13), no more than 2.5 passenger hours can be saved by implementing the holding control at either multiple holding stations or a single holding station.

As service becomes more frequent, the average passenger waiting time increases substantially. This is simply because a more frequent service is more likely subject to larger system operation variability. For instance, by comparing Tables 2.2 and 2.5, one may see that, as the travel time COV is 0.1, the total passenger waiting time increases by 33 percent when the service headway is reduced from 10 minutes to 5 minutes, and by 40 percent, 46 percent, 52 percent and 55 percent as the travel time COVs are 0.15, 0.20, 0.25 and 0.30 respectively. This further demonstrates that a more frequent service is subject to the system variability. Here, it is useful to point out again that the passenger boarding profiles for different service frequencies are deliberately designed to ensure that the total passenger waiting time are the same for all service frequencies, given that the system operation is entirely free of stochasticity. In other words, if the vehicle dispatching headway can be maintained at all points along the route, the same total passenger waiting time can be observed for all cases with different service frequencies.

As the peak passenger boarding occurs further downstream, the total passenger waiting time tends to increase. This is simply because the headway variability also becomes larger at the downstream stops of the transit route. This, together with a relatively large proportion of the passengers waiting at these stops, will certainly escalate the total passenger waiting time. For example, as the peak passenger boarding stops move from stops 13, 14 and 15 (profile 1) downstream to stop 29, 30 and 31 (profile 3), for the case of a 5 minute headway (Tables 2.2 and 2.10), the total passenger waiting time is increased by 171, 238, 286, 342 and 374 passenger minutes for the travel time COVs of 0.1, 0.15, 0.2, 0.25 and 0.3, respectively. Even for the service with a 10 minute headway

(Tables 2.5 and 2.13), this passenger waiting time increase could be as high as 137 passenger minutes when the travel time COV is equal to 0.3. However, within a more variable transit operation environment, holding control can play an even bigger role to reduce the passenger waiting time. The escalation of the total passenger cost due to the passenger boarding peaking at downstream stops can be diminished by implementing holding control, especially by holding vehicles at multiple holding stations. For the case of a 5 minute headway and a travel time COV of 0.3, the total passenger cost after holding vehicles at multiple holding stations under passenger boarding profile 3 (Table 2.10) is only 53 passenger minutes larger than that under passenger boarding profile 1 (Table 2.2). This is basically negligible.

Furthermore, the problem formulated in this study is only an approximation of the real-world problem in the sense that, the problem formulated in this study uses a deterministic service model, and the passenger alighting process is also removed from the problem formulation with an assumption that it may have only negligible effects on the vehicle trajectory dynamics. Such an approximation could discount the effects of the holding control determined from the problem formulation. Herein, one may imagine a case that the transit operation is entirely free of stochasticity, so that each passenger would have experienced the waiting time of exactly one half of the vehicle headway on average. For such case, under passenger boarding profile 1, the minimum total passenger waiting cost is shown in Table 2.14.

Table 2.14. Minimum Passenger Waiting Time with Varying Headways

Vehicle Dispatching Headway (Min.)	Minimum Total Passenger Cost (Pass.Min)
5	11041
6	11129
8	11305
10	11482

As discussed earlier, scaling the passenger boarding rate to be inversely proportional to the operation headway ensures identical total passenger cost with different headways, provided the service is entirely deterministic. However, one may see that there is slight difference between the numbers in the second column of Table 2.14. This is because the total passenger cost also includes the passenger waiting time to the dummy vehicle $O+1$, which functions as the boundary vehicle operating immediately after the last vehicle within the peak time period. For the service with a 5 minute headway, this dummy vehicle incurs 1/24 of the total passenger cost increase, since there are, in total, 24 actual vehicles operating during the peak time period (two hours). Similarly, the dummy vehicle incurs 1/12 increase on the total passenger cost for the service with 10 minute headway. This is why a slightly larger passenger cost comes with larger vehicle headways.

The comparison of the actual total passenger cost with holding control at multiple holding stations with the minimum total passenger cost is shown in Table 2.15.

Moreover, one may compare the actual total passenger cost without holding control with the minimum total passenger waiting cost, as shown in Table 2.16.

Table 2.15. Comparison of Total Passenger Cost after Holding with Minimum Total Passenger Cost

Travel Time COV	Passenger Cost Difference (Actual Cost after Holding – Minimum Cost)							
	5 Min. Headway		6 Min. Headway		8 Min. Headway		10 Min. Headway	
	Absolute Difference (Pass.Min)	Percentage Difference (%)	Absolute Difference (Pass.Min)	Percentage Difference (%)	Absolute Difference (Pass.Min)	Percentage Difference (%)	Absolute Difference (Pass.Min)	Percentage Difference (%)
0.10	1054	9.6	765	6.9	457	4.0	308	2.7
0.15	1387	12.6	1008	9.1	602	5.3	408	3.6
0.20	1781	16.1	1294	11.6	772	6.8	528	4.6
0.25	2174	19.7	1593	14.3	944	8.4	650	5.7
0.30	2531	22.9	1852	16.6	1100	9.7	761	6.6

Table 2.16. Comparison of Total Passenger Cost without Holding with Minimum Total Passenger Cost

Travel Time COV	Passenger Cost Difference (Actual Cost without Holding – Minimum Cost)							
	5 Min. Headway		6 Min. Headway		8 Min. Headway		10 Min. Headway	
	Absolute Difference (Pass.Min)	Percentage Difference (%)	Absolute Difference (Pass.Min)	Percentage Difference (%)	Absolute Difference (Pass.Min)	Percentage Difference (%)	Absolute Difference (Pass.Min)	Percentage Difference (%)
0.10	4282	38.8	2451	22.0	979	8.7	467	4.1
0.15	5346	48.4	3198	28.7	1380	12.2	682	5.9
0.20	6469	58.6	4021	36.1	1835	16.2	952	8.3
0.25	7478	67.7	4756	42.7	2270	20.1	1227	10.7
0.30	8307	75.2	5350	48.1	2646	23.4	1467	12.8

From Tables 2.15 and 2.16, one may see that the passenger cost increase solely due to the system operation stochasticity can be significantly diminished by implementing holding control. For instance, for the case of the 5 minute headway in the above two tables, applying holding control at multiple holding stations can reduce the passenger cost increment over the ideal case (as shown in Table 2.14) by more than 70 percent (or up to 5776 passenger minutes). This passenger cost reduction is considerable, and can demonstrate the effectiveness of the problem formulation and solution method in this study. The remaining passenger cost increment (up to 30 percent) over the ideal case can be attributed to the fact that the holding control is only applied at a set of discrete points, between which the vehicle headway variability still exists. Similar conclusions can be

drawn for the strategy of holding vehicles at single holding station. However, holding vehicles at a single holding station is less effective than holding vehicles at multiple holding stations in terms of the total passenger cost reduction.

2.4. Conclusions and Discussion

To take the advantage of the AVL data, this chapter takes a significant leap from the previous studies on the holding control through applying holding control at multiple holding stations rather than a single holding station and comparing the performance of holding control at multiple holding stations and holding control at a single holding station through a simulation study.

The problem of holding multiple vehicles at multiple/single holding station(s) in this chapter is formulated as a convex problem with strictly convex objective function subject to linear constraints. Some classical techniques can solve this problem to optimality. However, this does not necessarily mean that the problem is small in scale. Therefore, heuristics are also developed in this study to solve this particular problem by decomposing the overall problem into sub-problems that can be tackled more easily. With the algorithms developed for *PSS*, *PMS* and *PMM* problems respectively, the *PSS* problem can be solved analytically; the proposed H1 algorithm can solve the *PMS* problem to optimality; and, the H2 heuristic can also find the optimal solution to *PMM* problem, if the assumption of no vehicle overtaking is strictly satisfied.

Though the model formulated in this study is only an approximation of the problem in the real world, the simulation results suggest that, based on this problem formulation, holding vehicles both at multiple stations and at a single station can effectively reduce the

total passenger waiting time, when transit operation is subject to a certain level of variability, typically when the service frequency is high and vehicle travel time variability is large. On the other hand, when the service frequency is low and the vehicle travel time variability is small, the operation system improvement in terms of the total passenger waiting time reduction resulting from the holding control becomes negligible. In addition, the performance of the holding strategy is relatively independent of where the peak passenger boarding occurs.

Strong evidence drawn from the simulation results suggest that, across all cases, multiple holding stations may offer more opportunities to regularize the vehicle headways, so that the overall passenger cost can be reduced further as compared to holding vehicles only at a single holding station. Furthermore, the extra total passenger waiting time savings resulting from holding vehicles at multiple stations over holding vehicles at single station can be augmented as the system operation is subject to larger variability, or equivalently as the frequency of service is higher and travel time variation is larger.

Moreover, for the strategy of holding vehicles at a single holding station, the position of the optimal holding station becomes less sensitive to where the passenger boarding peaks, as the peak passenger boarding occurs further downstream. This may suggest that the position of the optimal holding station is mostly determined by the route structure itself, rather than solely by where peak passenger boarding occurs. However, it may be desirable to conduct further study on this problem.

CHAPTER 3 STOP-SKIPPING CONTROL

As a continuation of the study on the single route based operations control problems, this chapter describes a study of the stop-skipping control problem in a similar manner to the holding control problem. In practice, stop-skipping control is also called vehicle expressing, in which a vehicle is requested to skip a number of stops/stations to avoid the associated dwell times, so that the vehicle operations can recover from a service disruption.

3.1. Introduction

As introduced in the previous chapter, those factors and effects contributing to the vehicle operation variability may also disrupt the planned service schedule. Specifically, major transit service disruptions often occur due to a variety of reasons, e.g. vehicle malfunction or a traffic incident, to name a few. As these disruptions may last more than a few minutes, this may strain the ability of the transit service to recover from these disruptions. For example, for a relatively frequent transit service, Abkowitz et al. (1986) have concluded that the vehicle headway variation does not increase linearly along a route. This may imply that, as traffic congestion prevails on the route, even a minor transit perturbation could eventually result in serious operational problems, if no operations control actions are implemented. Under such a circumstance, the previously discussed vehicle holding control may help to regulate vehicle headways and restore the transit service if the disruption is not serious. However, when a major disruption occurs, simply holding the following vehicles may finally lead to increasingly greater headway variation and schedule deviations.

In this context, vehicle stop-skipping control, as another commonly used control strategy, can be particularly helpful to this situation. In this strategy, a driver is instructed to skip a set of stops along a route. By skipping these stops, the dwell times at these stops are avoided, and the vehicle may recover after some schedule delay or disruption.

Vehicle stop-skipping can be applied both in the planning stage of timetabling as a strategy to equalize the vehicle load and to minimize the vehicle fleet size requirement, and also in the operational stage for the purposes of improving schedule adherence or of regularizing the vehicle headways so as to reduce the overall passenger cost. Obviously, stop-skipping results in the loss of service to a certain portion of passengers, and also results in inconvenience to some other passengers. Due to this, stop-skipping (or expressing) is usually considered more formally in the planning process during timetabling, rather than in operations control.

At the planning stage, so-called “zonal” route design has some similarities to the stop-skipping operation. In this planning activity, only those stops within consecutive zones specifically assigned to a particular route will be served regularly by the buses on that route. Other stops will be either skipped or partially served with boarding or alighting only. Jordan and Turnquist (1979) developed a model of an urban bus route incorporating measures of both reliability and average trip time, based on which zone structure can be optimized by using a dynamic programming model. Also, a case study has shown that zone scheduling can simultaneously improve route reliability and average passenger travel time, while the required bus fleet can be reduced too. Furth (1986) extended the past study on the zonal express service design for linear corridor to zonal

design for bi-directional local service, including light direction deadheading, and to branching as well as linear corridors. The results of the application of this study to a Boston area corridor showed considerable potential for reducing operator cost.

As examples of limited-stop and express services, Ercolano (1984) evaluated the limited-stop bus service in New York City's borough of Manhattan by comparing its performance characteristics and passenger use to those of local service on the same routes. This study found that the limited-stop bus service can offer considerable travel time savings, faster average operating speed, rider preference, and moderate operating cost savings. Also, two sets of bi-variate regression models and five warrants are presented in the study to serve as general sketch-planning guides for applying limited-stop bus service. More recently, Suh et al. (2002) introduced an express subway system planned for the Seoul, Korea, metropolitan area, in which the stop-skipping system is considered as an alternative to the express subway service. The total time saving of the stop-skipping schedule is predicted with a given O/D matrix, distance between stations, headway, and maximum link speeds. The results from a field study have shown that the total system time can be decreased by 7.1 to 7.8 percent, and in the maximum case, up to 9.8 percent.

Vuchic (1973) is among the first who have comprehensively described and evaluated stop-skipping in an operational context. This study concluded that the operational differences of stop-skipping operation are: scheduled speed is increased; the frequency of stopping is reduced; headways at stations which are not served by all vehicles are increased; there is no direct connection between those stations served by different vehicles; and, perhaps most obviously, service becomes more complicated. In

the 1990's, Wilson et al. (1992) systematically described the operation controls applied on the Massachusetts Bay Transportation Authority (MBTA) Green Line in Boston, among which the vehicle expressing is a very important component.

Though very few, some efforts have been devoted by researchers to investigate the real-time application of a stop-skipping strategy for operations control purposes. Li et al. (1991) formulated a 0-1 stochastic programming model to solve the real-time scheduling problem for stop-skipping control, with the objective function accounting for both schedule deviation and unsatisfied passenger demand. In their model, the stop-skipping problem is formulated as determining which stops to serve on a given route, prior to the vehicle dispatch from the terminal. The decision variables are 0-1, based on the decision to serve a particular stop. The numerical results from an application in Shanghai demonstrated the potential of the method for improving transit operation in practice. Lin et al. (1995) investigated the combined control strategy of holding and stop-skipping, and concluded that the combination of such tight controls may actually increase overall passenger waiting times and thus should be avoided. In Eberlein (1995), the stop-skipping problem is formulated as an integer nonlinear programming model (INLP) to decide which vehicle to skip which stops/stations. In this model, the decision problem is to determine the start and end stops of the segment to be skipped (the “express segment”). Some results for a light-rail line in Boston are also given. More recently, Fu and Liu (2003) described a new dynamic scheduling strategy that aims to optimally balance the benefits to operator and passengers. In this study, the problem is again formulated as a nonlinear 0-1 integer programming problem, similar to that defined by Li et al. (1991),

which can be optimally solved using an exhaustive search method, by taking the advantage of the relatively small scale of the problem for a single transit route. A sensitivity analysis is also conducted in this study with a real-life bus route, and the passenger demand, the bus travel time variation, and the headway at the dispatching terminal are chosen as the sensitivity factors. The simulation-based sensitivity analysis indicates that stop-skipping can be most effective in the context of high passenger demand and short headways. Also, it should only be used when appropriate route travel time variation exists.

The formulations of Li et al. (1991), Lin et al. (1995), Eberlein (1995), and Fu and Liu (2003) all formulated the real-time stop-skipping problem as a decision problem solved during vehicle dispatch from the terminal. In those studies, the underlying assumption is that, as soon as the vehicle is dispatched from the terminal, the prescribed skipping stops (or skipped segment) cannot be changed. In this sense, the application of stop-skipping control is not ‘thoroughly’ real-time, since the control decision is not essentially adaptive to the changes in operating conditions once the vehicle is on route. This greatly limits the advantages that can be offered by any of the advanced information technologies, e.g. automatic vehicle location (AVL) and monitoring systems, especially when the route operating conditions are subject to significant variations. Moreover, the methods in these studies cannot be easily implemented to respond to a vehicle disruption in a timely manner, since stop-skipping control can only be decided while the vehicles are being dispatched at the terminal.

With these arguments above as the motivation, this study differs from the previous studies by implementing the stop-skipping control strategy in a genuinely real-time and adaptive way, so that the control can be applied in more general situations and can adapt to the operation conditions by adjusting the number and locations of the skipped stops in a real-time manner.

Consistent with Eberlein (1995) and some stop-skipping practice, the stop-skipping control generally follows the following policy (hereafter the “basic” policy, or policy 1):

Basic Policy: The route skipping segment is defined by a start stop and an end stop, and all stops on this segment will be completely skipped by the control vehicle.

With this policy, the stop-skipping control strategy definitely adds the inconvenience to some onboard passengers: those destined for stops on the skipping segment will have to alight from the vehicle before reaching their destinations.

Again, as already discussed in Chapter 2, for an “on-call” service, it is not necessary for a vehicle to stop at all stops on the route due to the inherent stochasticity of the passenger arrival and distribution process. Therefore, if the vehicle only responds to passenger calls for alighting, the amount of stopping on the route can also be reduced, so that the vehicle operating speed can be increased. This somewhat resembles that underlies the so-called “zonal local service” in Furth (1986), in which the zonal local service requires that the inbound vehicle stops between its service zone and the CBD *only to allow passengers to alight*. This idea suggests another policy (hereafter the “policy alternative”, or policy 2), as a comparison to the basic policy introduced earlier.

Policy Alternative: The skipping segment is defined by a start stop and an end stop. However, it is not necessary for the control vehicle to skip all stops on this segment. Instead, it can drop off the passengers at stops in the skipping segment if their destinations happen to be in this segment. At those stops where alighting does occur, passenger boarding is also allowed.

With this policy alternative, one may see that some of the inconveniences incurred by the basic policy to the passengers can be avoided. More importantly, the stop-skipping control thus becomes appropriate for real-time application in the sense that the change of control decision does not incur any contradicted activity that the basic policy may bring, e.g. the destination stops of the passengers who are forced to get off before their destinations are actually served eventually.

Based on these two stop-skipping control policies, the objectives and the tasks of this study of the real-time stop-skipping control problem is to:

- Formulate the dynamics of each stop-skipping control policy; and,
- Evaluate and compare the performance of the two stop-skipping control policies in the context of real-time application through a simulation study.

In the remaining part of this chapter, the problem of stop-skipping control is formulated as a mathematical programming problem within the context of a deterministic service model to optimize the number and the locations of the stops to be skipped by the control vehicle, so as to minimize the total passenger cost in terms of both in-vehicle passenger onboard time and at-stop passenger waiting time. Binomial passenger alighting and Poisson passenger boarding are incorporated into the vehicle trajectory prediction to

estimate the probability that a stop is either served or bypassed by the vehicles under both basic policy and policy alternative. Much more detail of this will be given in later sections. To take advantage of the small scale of the problem for a single transit route, an exhaustive search method is employed to compare all feasible combinations of the stops to be skipped by the control vehicle and eventually to present the combination of stops resulting in minimal passenger cost as the optimal solution. Again, similar to the holding problem in Chapter 2, a CRN simulation study is conducted to compare how the two policies relatively affect the performance of the stop-skipping control in terms of the capability of reducing total passenger cost. The main flow of this chapter is also depicted in Fig. 3.1.

The comparison simulation study is conducted in this chapter is the same as presented in Fig. 2.2 in Chapter 2, except the impacting factors identified at the beginning of the flow chart (Fig. 2.2) and the systems to be compared in the simulation study. Therefore, a diagram similar to Fig 2.2 will not be given in this chapter. In this chapter, the systems to be compared in the simulation study are just the two policies, namely the basic policy and the policy alternative.

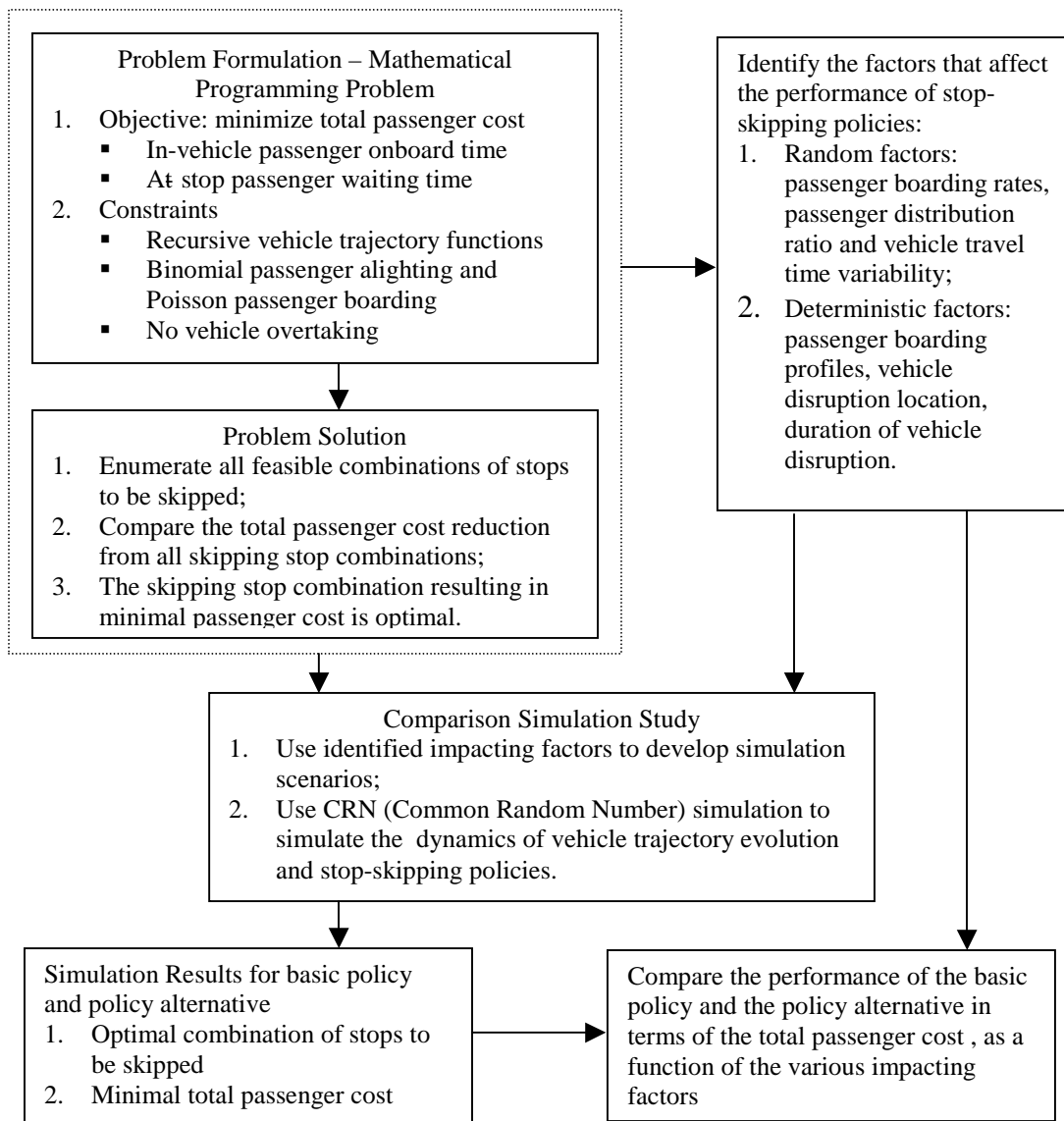


Fig. 3.1 Chapter Flowchart (Stop-Skipping Control)

Specifically, the remainder of this chapter is organized into three sections. Section 3.2 formulates the stop-skipping control dynamics based on both policies. Section 3.3 presents the simulation study results. Finally, Section 3.4 gives some conclusions.

3.2. Problem Formulation & Solution

3.2.1. Problem Nature

As indicated in the previous studies, the stop-skipping problem can be formulated as a nonlinear 0-1 integer programming problem, with the binary integer variables representing which stops to be skipped by the control vehicle. In this study, for the purpose of investigating the exact problem, the stop-skipping problem will be formulated again as a nonlinear 0-1 integer programming problem.

The formulation below is deterministic, in that the objective function uses expected values to determine the passenger cost. However, in order to consider policy 2, the problem formulation also uses the probability that a vehicle serves or bypasses a stop on the skipping segment. In this case, when the problem is formulated and solved in the decision-making process, only the expected values of passenger boardings and alightings are used to predict the vehicle trajectories and estimate the objective function value.

In this stop-skipping study, under both the basic and alternative stop-skipping policies, the decision problem is to determine the starting stop and the ending stop of the skipping segment. The stop-skipping problem with a relatively large number of stops (100 stops) and a substantial number of vehicles (6 vehicles) has been solved in an experimental study; even in this case, one stop-skipping control decision took only a few seconds to solve using full enumeration. This is entirely acceptable for a real-time application, given that most AVL systems provide vehicle location information at a lower polling cycle (e.g., from 20-30 seconds to 2-3 minutes). Therefore, an exhaustive search method is adopted as the solution method in this study.

3.2.2. Assumptions and Variable Definitions

For this study, it is assumed that a vehicle experiences a service disruption of a given length of time. When this disruption occurs, this vehicle experiences a delay equal to the length of the service disruption. At that end of this disruption, and based on the current location of the vehicle on the route and the operation status of other impacted vehicles, a decision is made on whether the vehicle should skip stops on the route, and, if so, one must determine the starting and ending stops of this skipping segment.

In terms of the stop-skipping control, the following assumptions are made:

- While considering a vehicle for stop-skipping, the following (subsequent) vehicles will not be eligible for any control actions; and,
- For the sake of simplifying the problem formulation and solution, vehicle capacity is not considered explicitly. However, practically, when a vehicle is involved in a disruption, vehicle capacity could easily become an issue due to the length of the following vehicle headway and the need to pick up passengers who are “forced off” the control vehicle. To examine this assumption, an additional set of scenarios is included in the simulation study, in Section 3.4, to explore the effects of a real-life capacity constraint.

For the basic policy, it is further assumed that:

- In order to make the basic policy more appropriate to adaptive decision-making, the control vehicle will pick up all passengers up until the stop just upstream of the first stop to be skipped, even those who are destined for stops on the

skipping segment, in case that the stops previously decided to be skipped are eventually served; and,

- Also in the basic policy, once the skipping segment has been decided, the control vehicle always drops off the passengers destined for stops in the skipping segment at the last stop before the skipping segment. In this sense, the control vehicle can never skip the stop immediately downstream of the decision-making location, since it may need to drop off some passengers destined along the skipping segment.

In terms of modeling features in the operations dynamics on the route, the following assumptions are made. For the sake of completeness, some other assumptions made in Chapter 2 are also duplicated below:

- For passengers boarding at each stop, the fraction alighting at each downstream stop can be derived from historical data and is assumed to be independent of the number of passengers boarding at the stop;
- The average passenger arrival rate at each stop is given;
- The average vehicle travel time between stops is given, and is the same for all vehicles during the time period of concern; and,
- For all stops already served by the vehicles, the number of passengers boarding and alighting is known, and the passenger distribution pattern derived from historical data can be adjusted based on the algorithm already introduced in the holding control problem study.

The variables are defined as follows. Again, some variables which have been described in the holding control problem study are defined below for completeness.

Inputs and Parameters:

$X_{i,n,j}$: The ratio of the number of passengers boarding at stop n and alighting at stop j to the total number of passengers boarding at stop n for vehicle i ;

λ_j : The passenger arrival rate at stop j ;

$d_{i,j}$: The departure time of vehicle i at stop j ;

$a_{i,j}$: The arrival time of vehicle i at stop j ;

$B_{i,j}$: The expected number of passengers boarding vehicle i at stop j ;

$A_{i,j}$: The expected number of passengers alighting from vehicle i at stop j ;

$\bar{B}_{i,j}$: In contrast to $B_{i,j}$, $\bar{B}_{i,j}$ represents the real number of passengers boarding vehicle i at stop j ;

$\bar{A}_{i,j}$: In contrast to $A_{i,j}$, $\bar{A}_{i,j}$ represents the real number of passengers alighting from vehicle i at stop j ;

$h_{i,j}$: The leading headway of vehicle i at stop j .

L_j : The number of passengers left over by the control vehicle at stop j ;

F : The passengers being forced to alight off the control vehicle at the stop immediately upstream of the first stop on the skipping segment (applies to policy 1 only);

$P_{i,j}$: The probability that stop j is skipped by vehicle i ;

$Q_{i,j}$: The probability that stop j is served by vehicle i ;

$r_{i,j}$: The vehicle running time between stop j and stop $j+1$ for vehicle i ;

$D_{i,n,j}$: The expected number of passengers boarding vehicle i at stop n and alighting at stop j ;

s_i : The stop immediately upstream of the current location of vehicle i when the control decision is made. For an adaptive decision-making process, when the first decision is made, s_1 is just the stop immediately upstream of the vehicle disruption location;

α, β_1, β_2 : The parameters representing the vehicle acceleration, deceleration, door open and close, and clearance time (α), average passenger boarding time (β_1) and average passenger alighting time (β_2) respectively at a bus stop;

w_1, w_2, w_3 : The weights for different passenger cost components in the objective function; and,

N, O : The total number of stops on route (N) and the total number of vehicles under consideration (O), respectively. Only for the purpose of convenience, the boundary vehicle (the vehicle preceding the control vehicle) is denoted as vehicle 0; the control vehicle is vehicle 1, and others from 2 through O are referred to as the impacted vehicles. Herein, the boundary vehicle 0 does not belong to the O vehicles under consideration. The definition of the boundary vehicle, the control vehicle and the impacted vehicles will be introduced later.

Decision Variables:

s_s : The first stop on the skipping segment that is skipped; and

s_e : The last stop on the skipping segment that is skipped;

Though the problem will be essentially formulated as a non-linear binary integer programming problem, the mathematics for the problem dynamics can have a wide variety of forms. As driven by the solution algorithm argued previously, and also for the sake of making the dynamics more readable, the problem formulation will be described in the manner that the solution algorithm works. For the solution algorithm, the skipping segment is defined by enumerating each feasible pair of start stop (s_s) and end stop (s_e) among all stops downstream of the decision location (s_1). Therefore, the critical issue in solving the problem is determining, for a given pair of start and end stops, the evolutionary dynamics of vehicle trajectories, while calculating the given costs in the objective function.

3.2.3. Objective Function

Since the two different stop-skipping policies result in different passenger boarding and alighting processes, the objective functions based on these policies are also different.

Basic Policy – Policy 1

$$\begin{aligned} \min Z = & w_1 \cdot \left(\sum_{i=1}^O \sum_{j=s_1+1}^N \frac{1}{2} \cdot \lambda_j \cdot h_{i,j}^2 + \sum_{j=s_s-1}^{s_e} L_j \cdot h_{2,j} \right) + w_2 \cdot \sum_{i=1}^O \sum_{j=l=\max(j+1, s_1+1)}^{N-1} \sum_{l=\max(j+1, s_1+1)}^N D_{i,j,l} \cdot (a_{i,l} - d_{i,j}) \\ & + w_3 \cdot F \cdot h_{2,s_s-1} \end{aligned} \quad (3-1)$$

In the objective function above, the first term shows that the passengers originating during any particular vehicle's leading headway will experience the waiting time of one half of the leading headway on average. However, for those passengers left over by the control vehicle 1, they will experience another full headway waiting for vehicle 2. The

second term indicates that each particular passenger will experience the in-vehicle travel time determined by the vehicle departure time at the stop at which they board on the vehicle and the vehicle arrival time at the stop at which they get off the vehicle. Finally, the third term counts the extra cost to those passengers that are forced to get off at stop $s_s - 1$, since their destination stops will be skipped by the control vehicle. Since in the first term of the objective function, L_j already includes F (as introduced later) at stop $j = s_s - 1$, the cost to the passengers being forced to get off before their destinations is counted by both the first and the third terms in the objective function. This formulation allows us to consider the additional inconvenience to those that are “forced off.” However, the weights w_1 and w_3 can be judiciously set to avoid any double counting problem.

Policy Alternative – Policy 2

For policy 2, since there is no passenger being forced to get off the vehicle before their destination stop, the third term in the objective function for policy 1 does not apply. Therefore, the objective function simplifies to:

$$\min Z = w_1 \cdot \left(\sum_{i=1}^O \sum_{j=s_1+1}^N \frac{1}{2} \cdot \lambda_j \cdot h_{i,j}^2 + \sum_{j=s_s}^{s_e} L_j \cdot h_{2,j} \right) + w_2 \cdot \sum_{i=1}^O \sum_{j=l=\max(j+1, s_1+1)}^{N-1} \sum_{l=j+1}^N D_{i,j,l} \cdot (a_{i,l} - d_{i,j}) \quad (3-2)$$

Obviously, for both policies, the stop-skipping control decision is just to determine which contiguous stops will be skipped by the control vehicle. Therefore, the decision variables are just the start stop s_s and the end stop s_e which together define the skipping segment. Though the decision variables are not shown explicitly in the objective

functions, they will define the stop-skipping control dynamics in the way as described below.

3.2.4. Stop-Skipping Control and Vehicle Trajectory Evolution Dynamics

As a concept, the impacted, control and boundary vehicles have been introduced in the holding control problem study in Chapter 2. Similarly, since the stop-skipping control is mostly applied when a major vehicle operating disruption occurs on the route, the impact of the operation disruption could propagate to many of the following vehicles, which are also defined as the impacted vehicles in this study. In contrast, the vehicle experiencing the disruption is defined as the control vehicle. Furthermore, the immediately preceding vehicle is defined as the boundary vehicle, whose trajectory is assumed to be known and will determine how the trajectories of the following vehicles evolve with or without the stop-skipping control. In this particular study, the number of impacted vehicles may vary with the route operating condition and the degree of the vehicle disruption. The rule of thumb in deciding the appropriate number of impacted vehicles is to ensure that the last impacted vehicle's trajectory can stay approximately the same when the vehicle disruption length varies from zero up to the average dispatch headway, provided that no operation control is applied.

When the stop-skipping control decision is made, the control vehicle is currently at the location immediately downstream of stop s_1 and upstream of stop $s_1 + 1$. The stops downstream of the vehicle disruption location may lie on any one of the three segments as defined below and shown in Fig. 3.2.

- Segment 1: between stop $s_1 + 1$ and $s_s - 1$, including both end stops;

- Segment 2: the skipping segment, between stop s_s and stop s_e , including both end stops; and,
- Segment 3: between stop $s_e + 1$ and stop N (end terminal), including these two stops.

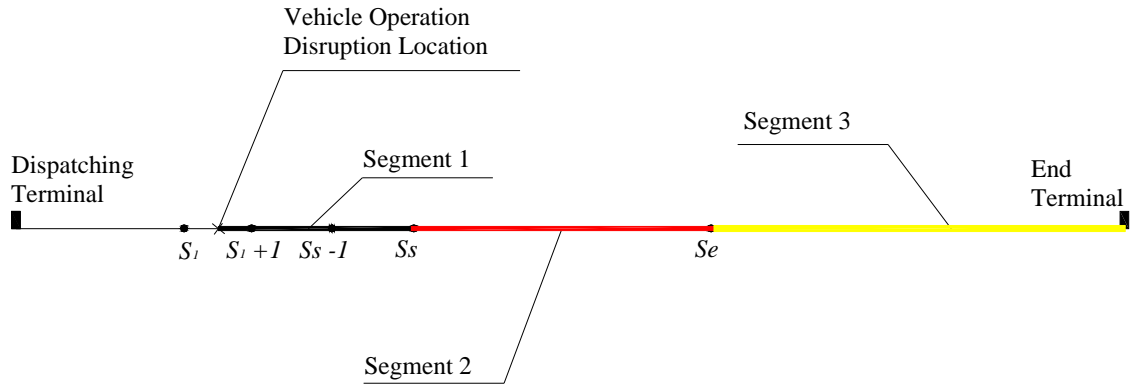


Fig. 3.2. Typical Route Segmentation for Stop-Skipping Control

The segment between the dispatching terminal and stop s_1 is not of concern in this study, since all vehicle trajectories on this segment either have materialized or cannot be changed by the operation control applied to the control vehicle.

At the stops on each of the route segments defined above, the vehicle trajectory evolves in different ways, which also depends on which stop-skipping control policy will be used.

For an on-call service, as introduced in Chapter 2 for holding control problem, besides the policies, whether or not the control vehicle serves a particular stop also depends on if there are passengers calling for boarding or/and alighting at the stop. Again, in the model, this will be represented as the probability of serving or bypassing

the stop, using Poisson and binomial distributions to represent passenger boarding and alighting processes respectively.

Though the two different stop-skipping policies affect the vehicle trajectory evolution in different ways, the basic vehicle dynamics for all vehicles at all stops are as the follows. First, the probability $P_{i,j}$ of a vehicle i serving a stop j and $Q_{i,j}$, the probability of a vehicle skipping a stop, sum to 1.0. However, calculating these probabilities is specific to the stop-skipping policy and to the specific route segment, and these will be introduced later.

The vehicle arrival time at each stop would be $r_{i,j-1}$ later than the departure time at the upstream adjacent stop.

$$a_{i,j} = d_{i,j-1} + r_{i,j-1} \quad \forall i \in O, j \geq 2 \quad (3-3)$$

The vehicle departure time at each stop should include the vehicle dwell time, which consists of the passenger boarding time, alighting time, as well as the constant part of vehicle dwell time at stop, whose expected value is also determined by the probability of a vehicle serving that particular stop. Previous studies of vehicle dwell times have used either simultaneous or sequential passenger boarding and alighting processes. However, the sequential passenger boarding and alighting process (alighting first, then boarding) is adopted by a majority of the previous studies to formulate the operations control problem due to its relatively simple mathematical form. For the sake of being consistent with the previous studies, this research formulates the vehicle dwell time based on the sequential passenger boarding and alighting process, which is given in (3-4) as follows:

$$d_{i,j} = a_{i,j} + \alpha \cdot Q_{i,j} + \beta_1 \cdot B_{i,j} + \beta_2 \cdot A_{i,j} \quad \forall i \in O, j \in N \quad (3-4)$$

Vehicles are not allowed to overtake each other.

$$d_{i-1,j} \leq a_{i,j} \quad \forall i \geq 1, j \in N \quad (3-5)$$

The detail of the equations (3-4) and (3-5) has been given in Chapter 2.

The number of passengers alighting at a particular stop is simply the total of all passengers boarding from upstream stops and, at the same time, having the destination at this stop.

$$A_{i,j} = \sum_{l=1}^{j-1} D_{i,l,j} \quad \forall i \in O, j \geq 2 \quad (3-6)$$

The number of passengers boarding at a particular stop and the passenger origin-destination (O/D) distribution between stops are much more complicated based on the specific policy and route segment.

For all of the stops which have been passed by the vehicles, all the variables defined by (3-2) through (3-6) are known. This adds the following equations:

$$B_{i,j} = \bar{B}_{i,j} \quad \forall i \in O, j \leq s_i \quad (3-7)$$

$$A_{i,j} = \bar{A}_{i,j} \quad \forall i \in O, j \leq s_i \quad (3-8)$$

Again, equations (3-7) and (3-8) also implies that the expected number ($B_{i,j}$) of passengers boardings and expected number ($A_{i,j}$) of passenger alightings on vehicle i at any one of the stops being passed by vehicle i is just the observed numbers $\bar{B}_{i,j}$ and $\bar{A}_{i,j}$ respectively.

Other elements of the vehicle dynamics are specific to the stop-skipping policy and the route segment. In the following, these are broken down by segment (1, 2 and 3, as defined previously in Fig. 3.2) and by policy (1 and 2, also defined previously). The

intent is to show the specific dynamics on these segments and under these policies, in order to calculate the objective functions in (3-1) and (3-2).

Furthermore, to cater to adaptive decision-making, the following equations are all based on the concept that the current location of each vehicle i ($\forall i \in O$) is just downstream of stop s_i , which is known from the AVL data. Therefore, for vehicle i at all stops upstream of s_i (i.e. $j \leq s_i$), the expected number of passenger boardings $B_{i,j}$ and alightings $A_{i,j}$ have been updated with the observed values $\bar{B}_{i,j}$ and $\bar{A}_{i,j}$ respectively, as presented by (3-7) and (3-8). Also, $X_{i,n,j}$ and $D_{i,n,j}$ will be adjusted accordingly. The method to adjust $X_{i,n,j}$ and $D_{i,n,j}$ has been introduced in Chapter 2.

Segment 1 ($s_1 + 1 \leq j \leq s_s - 1$):

Basic Policy -- Policy 1

Similar to that in holding control problem study, with the assumed Poisson process for passenger arrivals, the probability of the vehicle i serving stop j can be determined as below. Since the probability of skipping or serving a stop is the key in the stop-skipping control problem study, especially for the policy alternative, the probability estimation will be given much more detail below than in the holding control problem study in Chapter 2.

When $i > 2$ or $[(i = 1 \text{ or } i = 2) \text{ and } j \neq s_s - 1]$,

$$P_{i,j} = \left(\prod_{l=1}^{j-1} \left(1 - \frac{D_{i,l,j}}{B_{i,l}} \right)^{B_{i,l}} \right) \cdot \exp[-\lambda_j \cdot (a_{i,j} - d_{i-1,j})] \quad (3-9)$$

In equation (3-9), the term $\frac{D_{i,l,j}}{B_{i,l}}$ is another expression of the term $X_{i,n,j}$ in equation (2-1). Furthermore, although this is not done in the holding problem, to further clarify how the observed number $\bar{B}_{i,l}$ substitutes for the expected number $B_{i,l}$, the equation (3-9) can be rewritten as below:

$$P_{i,j} = \left(\prod_{l=1}^{s_i} \left(1 - \frac{D_{i,l,j}}{\bar{B}_{i,l}}\right)^{\bar{B}_{i,l}} \prod_{l=s_i+1}^{j-1} \left(1 - \frac{D_{i,l,j}}{B_{i,l}}\right)^{B_{i,l}} \right) \cdot \exp[-\lambda_j \cdot (a_{i,j} - d_{i-1,j})] \quad (3-10)$$

Equation (3-10) splits up the first term in equation (3-9) into two terms accounting for the stops that have been passed by the vehicle and the downstream stops respectively. Equations (3-9) and (3-10) are entirely equivalent since $\bar{B}_{i,l}$ and $B_{i,l}$ for all stops $l \leq s_i$ are exactly the same, as suggested by (3-7) and (3-8).

When $i = 1$ and $j = s_s - 1$, the passengers alighting at stop j also include those with their destination stops on the skipping segment (from stop s_s through stop s_e). Therefore, the probability that no onboard passenger calls for alighting at stop j is actually the probability that no onboard passenger calls for alighting at stops j through s_e .

$$P_{i,j} = \left(\prod_{l=j}^{s_e} \prod_{n=1}^{l-1} \left(1 - \frac{D_{i,n,l}}{B_{i,n}}\right)^{B_{i,n}} \right) \cdot \exp[-\lambda_j \cdot (a_{i,j} - d_{i-1,j})] \quad (3-11)$$

When $i = 2$ and $j = s_s - 1$,

$$P_{i,j} = \left(\prod_{l=1}^{j-1} \left(1 - \frac{D_{i,l,j}}{B_{i,l}}\right)^{B_{i,l}} \right) \cdot \exp[-\lambda_j \cdot (a_{i,j} - d_{i-1,j}) - L_j] \quad (3-12)$$

In the three equations above, the first multiplicative term is the probability of no passenger calling for alighting at stop, and second term is the probability that no calls for

boarding from the passengers originating during the headway and/or the passengers left over by the control vehicle. One may note that $\frac{D_{i,l,j}}{B_{i,l}}$ is used instead of $X_{i,l,j}$. This is because $\frac{D_{i,l,j}}{B_{i,l}}$ is not equivalent to $X_{i,l,j}$ for many situations in which passengers are left over due to the stop-skipping control. For example, without stop-skipping control to vehicle i , $X_{i,l,j} = \frac{D_{i,l,j}}{B_{i,l}} = 0.1$. However, vehicle disruption occurs to vehicle i , and the dispatch center needs vehicle i to skip stop j . For such case, $\frac{D_{i,l,j}}{B_{i,l}} = 0$ because stop j is skipped and thus has no passenger alighting from vehicle i , but $X_{i,l,j}$ is still equal to 0.1. Therefore, for the purpose of consistency, when computing the probability of serving a stop, the term $X_{i,l,j}$ is used directly. However, $X_{i,l,j}$ is exactly equivalent to $\frac{D_{i,l,j}}{B_{i,l}}$ if no passengers are left over.

At stop $s_s - 1$, the control vehicle I will drop off a portion of the onboard passengers (F) who boarded at stop 1 through stop $s_s - 2$ and have the destinations on the skipping segment between stop s_s and stop s_e .

$$F = \sum_{n=1}^{s_s-2} \left(\lambda_n \cdot (d_{1,n} - d_{0,n}) \cdot \sum_{l=s_s}^{s_e} X_{i,n,l} \right) \quad (3-14)$$

Also, the control vehicle I will not pick up some passengers waiting at stop $s_s - 1$ because they also have the destinations on the skipping segment. Together with the

passengers (F) being forced to get off the control vehicle at stop $s_s - 1$, the total number of passengers (L_{s_s-1}) left over by the control vehicle at stop $s_s - 1$ would be:

$$L_{s_s-1} = \sum_{n=1}^{s_s-1} \left(\lambda_n \cdot (d_{1,n} - d_{0,n}) \cdot \sum_{l=s_s}^{s_e} X_{i,n,l} \right) \quad (3-13)$$

The control vehicle will pick up all passengers originating during its headway at all stops except $s_s - 1$; at stop $s_s - 1$, those passengers having the destination on the skipping segment will be left at the stop.

$$B_{i,j} = \lambda_j \cdot (d_{i,j} - d_{i-1,j}) \quad \text{for } i = 1, j \neq s_s - 1 \quad (3-15)$$

$$B_{i,j} = \lambda_j \cdot (d_{i,j} - d_{i-1,j}) \cdot \left(1 - \sum_{l=s_s}^{s_e} X_{i,j,l} \right) \quad \text{for } i = 1, j = s_s - 1 \quad (3-16)$$

The vehicle immediately following the control vehicle will pick up both its own originating passengers and those left over by the control vehicle, if any.

$$B_{i,j} = \lambda_j \cdot (d_{i,j} - d_{i-1,j}) \quad \text{for } i = 2, j \neq s_s - 1 \quad (3-17)$$

$$B_{i,j} = \lambda_j \cdot (d_{i,j} - d_{i-1,j}) + L_j \quad \text{for } i = 2, j = s_s - 1 \quad (3-18)$$

Other impacted vehicles only pick up passengers originating during the preceding headway.

$$B_{i,j} = \lambda_j \cdot (d_{i,j} - d_{i-1,j}) \quad \text{for } i > 2 \quad (3-19)$$

Because some passengers' destinations lie on the skipping segment, they will change their destinations to stop $s_s - 1$ for the control vehicle. For other vehicles, the passenger distribution can be easily determined by the number of originating passengers and the distribution ratio $X_{i,l,j}$.

$$D_{i,n,j} = B_{i,n} \cdot \left(X_{i,n,j} + \sum_{l=s_s}^{s_e} X_{i,n,l} \right) \quad \text{for } i = 1, n < j, j = s_s - 1 \quad (3-20)$$

$$D_{i,n,j} = B_{i,n} \cdot X_{i,n,j} \quad \text{for } i = 1, n < j < s_s - 1 \quad (3-21)$$

$$D_{i,n,j} = B_{i,n} \cdot X_{i,n,j} \quad \text{for } i \geq 2, n < j \quad (3-22)$$

Policy Alternative – Policy 2

The vehicle trajectory evolution is relatively straightforward for policy 2, since no passenger leftover is involved.

The probability $P_{i,j}$ that stop j is bypassed by vehicle i is simply the joint probability that no passengers call for alighting vehicle i at stop j $\left(\prod_{l=1}^{j-1} \left(1 - \frac{D_{i,l,j}}{B_{i,l}} \right)^{B_{i,l}} \right)$ and that no passengers call for boarding vehicle i at stop j $\exp(-B_{i,j})$.

$$P_{i,j} = \left(\prod_{l=1}^{j-1} \left(1 - \frac{D_{i,l,j}}{B_{i,l}} \right)^{B_{i,l}} \right) \cdot \exp(-B_{i,j}) \quad \forall i \in O \quad (3-23)$$

The average number of passengers boarding vehicle i at stop j is the product of the passenger boarding rate at stop j and vehicle i 's headway.

$$B_{i,j} = \lambda_j \cdot (d_{i,j} - d_{i-1,j}) \quad \forall i \in O \quad (3-24)$$

A fraction $(X_{i,n,j})$ of the passengers boarding on vehicle i at stop j will alight at stop j .

$$D_{i,n,j} = B_{i,n} \cdot X_{i,n,j} \quad \forall i \in O, n < j \quad (3-25)$$

Segment 2 ($s_s \leq j \leq s_e$):

Basic Policy – Policy 1

The basic policy requires that the control vehicle skips all stops on the skipping segment.

$$P_{i,j} = 1 \quad \text{for } i = 1 \quad (3-26)$$

Therefore, the immediately following vehicle has to respond the boarding calls from those passengers left over by the control vehicle. So, for $i = 2$,

$$P_{i,j} = \left(\prod_{l=1}^{j-1} \left(1 - \frac{D_{i,l,j}}{B_{i,l}} \right)^{B_{i,l}} \right) \cdot \exp[-\lambda_j \cdot (a_{i,j} - d_{i-1,j}) - L_j] \quad (3-27)$$

Other vehicles ($i > 2$) keep regular probability estimation as of no control applied.

$$P_{i,j} = \left(\prod_{l=1}^{j-1} \left(1 - \frac{D_{i,l,j}}{B_{i,l}} \right)^{B_{i,l}} \right) \cdot \exp[-\lambda_j \cdot (a_{i,j} - d_{i-1,j})] \quad (3-28)$$

All passengers originating during the control vehicle's leading headway will be left over for the immediately following vehicle to pick up, which implies no passenger boarding on the control vehicle.

$$L_j = \lambda_j \cdot (d_{1,j} - d_{0,j}) \quad (3-29)$$

$$B_{i,j} = 0 \quad \text{for } i = 1 \quad (3-30)$$

$$B_{i,j} = \lambda_j \cdot (d_{i,j} - d_{i-1,j}) + L_j \quad \text{for } i = 2 \quad (3-31)$$

$$B_{i,j} = \lambda_j \cdot (d_{i,j} - d_{i-1,j}) \quad \text{for } i > 2 \quad (3-32)$$

Certainly, no passenger having a destination on the skipping segment can board the control vehicle.

$$D_{i,n,j} = 0 \quad \text{for } i = 1, \quad n \leq j \quad (3-33)$$

Those passengers having the destinations on the skipping segment will actually ride vehicle 2.

$$D_{i,n,j} = \sum_{m=1}^i \lambda_n \cdot (d_{m,n} - d_{m-1,n}) \cdot X_{m,n,j} \quad \text{for } i = 2, s_s \leq n < j \quad (3-34)$$

Stop $s_s - 1$ merits more attention since some passengers ($\sum_{l=1}^{n-1} B_{1,l} \cdot X_{1,l,j}$) are forced to alight there by the control vehicle. These passengers will also ride vehicle 2.

$$D_{i,n,j} = \sum_{m=1}^i \lambda_n \cdot (d_{m,n} - d_{m-1,n}) \cdot X_{m,n,j} + \sum_{l=1}^{n-1} B_{1,l} \cdot X_{1,l,j} \quad \text{for } i = 2, n = s_s - 1 \quad (3-35)$$

$$D_{i,n,j} = B_{i,n} \cdot X_{i,n,j} \quad \text{for } i = 2, n < s_s - 1 \quad (3-36)$$

$$D_{i,n,j} = B_{i,n} \cdot X_{i,n,j} \quad \text{for } i > 2, n \leq j \quad (3-37)$$

Policy Alternative – Policy 2

For policy 2, the probability that the control vehicle stops on the skipping segment is entirely determined by the calls to alight from the onboard passengers.

$$P_{i,j} = \left(\prod_{l=1}^{j-1} \left(1 - \frac{D_{i,l,j}}{B_{i,l}} \right)^{B_{i,l}} \right) \quad \text{for } i = 1 \quad (3-38)$$

Similarly, the passengers left over by the control vehicle, if any, will board the immediately following vehicle.

$$P_{i,j} = \left(\prod_{l=1}^{j-1} \left(1 - \frac{D_{i,l,j}}{B_{i,l}} \right)^{B_{i,l}} \right) \cdot \exp[-\lambda_j \cdot (a_{i,j} - d_{i-1,j}) - L_j] \quad \text{for } i = 2 \quad (3-39)$$

$$P_{i,j} = \left(\prod_{l=1}^{j-1} \left(1 - \frac{D_{i,l,j}}{B_{i,l}} \right)^{B_{i,l}} \right) \cdot \exp[-\lambda_j \cdot (a_{i,j} - d_{i-1,j})] \quad \text{for } i \geq 2 \quad (3-40)$$

The expected number of passengers (L_j) left over by the control vehicle will also be determined by the probability ($P_{i,j}$) that the control vehicle does not serves the stop, and the opposite for the expected number of passengers boarding on the control vehicle.

$$L_j = \lambda_j \cdot (d_{1,j} - d_{0,j}) \cdot P_{i,j} \quad \text{for } i = 1 \quad (3-41)$$

$$B_{i,j} = \lambda_j \cdot (d_{i,j} - d_{i-1,j}) \cdot Q_{i,j} \quad \text{for } i = 1 \quad (3-42)$$

Again, the passengers (L_j) left over by the control vehicle are all taken by vehicle 2.

$$B_{i,j} = \lambda_j \cdot (d_{i,j} - d_{i-1,j}) + L_j \quad \text{for } i = 2 \quad (3-43)$$

$$B_{i,j} = \lambda_j \cdot (d_{i,j} - d_{i-1,j}) \quad \text{for } i > 2 \quad (3-44)$$

If $i = 2$ and $s_s \leq k \leq j$, $D_{i,n,j}$ includes not only the passengers $[\lambda_j \cdot (d_{i,j} - d_{i-1,j}) \cdot X_{i,n,j}]$ originally waiting for vehicle 2 and having the destination at stop j , but also the leftover passengers ($L_j \cdot X_{1,n,j}$) originally waiting for the control vehicle and having the destination at stop j .

$$D_{i,n,j} = \lambda_j \cdot (d_{i,j} - d_{i-1,j}) \cdot X_{i,n,j} + L_j \cdot X_{1,n,j} \quad (3-45)$$

Otherwise, i.e. ($i = 2$ and $n < s_s$) or ($i \neq 2$):

$$D_{i,n,j} = B_{i,n} \cdot X_{i,n,j} \quad (3-46)$$

Segment 3 ($s_e + 1 \leq j \leq N$):

Basic Policy – Policy 1

$$P_{i,j} = \left(\prod_{l=1}^{j-1} \left(1 - \frac{D_{i,l,j}}{B_{i,l}} \right)^{B_{i,l}} \right) \cdot \exp[-\lambda_j \cdot (d_{i,j} - d_{i-1,j})] \quad \forall i \in O \quad (3-47)$$

$$B_{i,j} = \lambda_j \cdot (d_{i,j} - d_{i-1,j}) \quad \forall i \in O \quad (3-48)$$

For the control vehicle, stop-skipping control does not affect the passenger distribution with the origin stop upstream of stop $s_s - 1$ or on segment 3.

$$D_{i,n,j} = B_{i,n} \cdot X_{i,n,j} \quad \text{for } i = 1, \quad n < s_s - 1 \text{ or } s_e < n < j \quad (3-49)$$

However, for the passenger distribution with the origin stop $s_s - 1$, it is necessary to use the number of passengers originating during the control vehicle's leading headway, instead of the expected number of passengers, to proportionately estimate the number of O-D trips, because not all originating passengers have the chance to board the control vehicle due to the stop-skipping control.

$$D_{i,n,j} = \lambda_n \cdot (d_{i,n} - d_{i-1,n}) \cdot X_{i,n,j} \quad \text{for } i = 1, \quad n = s_s - 1 \quad (3-50)$$

Certainly, there are no passengers boarding the control vehicle on the skipping segment and then alighting on segment 3.

$$D_{i,n,j} = 0 \quad \text{for } i = 1, \quad s_s \leq n \leq s_e \quad (3-51)$$

Actually, these passengers are picked up by the immediately following vehicle ($i = 2$).

$$D_{i,n,j} = \sum_{m=1}^i \lambda_n \cdot (d_{m,n} - d_{m-1,n}) \cdot X_{m,n,j} \quad \text{for } i = 2, \quad s_s \leq n \leq s_e \quad (3-52)$$

The passengers boarding on vehicle 2 on segment 1 and having the destination on segment 3 can be estimated with the equation below.

$$D_{i,n,j} = \lambda_n \cdot (d_{i,n} - d_{i-1,n}) \cdot X_{i,n,j} \quad \text{for } i = 2, \quad n < s_s \text{ or } s_e < n < j \quad (3-53)$$

Because no stops on the skipping segment are involved, no leftover passengers are included in equation (3-53).

For other vehicles, $D_{i,n,j}$ is simply the product of the number of passengers ($B_{i,n}$) boarding at the origin stop n and the ratio $X_{i,n,j}$.

$$D_{i,n,j} = B_{i,n} \cdot X_{i,n,j} \quad \text{for } i > 2, n < j \quad (3-54)$$

Policy Alternative – Policy 2

For policy 2, same expressions can be applied to all vehicles, including the control vehicle and vehicle 2. However, to be noted that, for vehicle 2, $B_{i,n}$ in equation (3-57) may include the passenger left over by the control vehicle, if stop n was skipped by the control vehicle, as indicated by equation (3-43).

$$P_{i,j} = \left(\prod_{l=1}^{j-1} \left(1 - \frac{D_{i,l,j}}{B_{i,l}} \right)^{B_{i,l}} \right) \cdot \exp[-\lambda_j \cdot (a_{i,j} - d_{i-1,j})] \quad \forall i \in O \quad (3-55)$$

$$B_{i,j} = \lambda_j \cdot (d_{i,j} - d_{i-1,j}) \quad \forall i \in O \quad (3-56)$$

$$D_{i,n,j} = B_{i,n} \cdot X_{i,n,j} \quad \forall i \in O, n < j \quad (3-57)$$

With these equations from (3-3) through (3-57), all vehicles' trajectories can be predicted, and the number of passenger boardings and alightings at each stop for each vehicle can be estimated.

3.3. Simulation Study

3.3.1. Simulation Logic

In order to evaluate and compare the performance of stop-skipping control with the two different policies, a simulation study is conducted. In both policy 1 and policy 2, an initial decision about the start and end stops of the skipped segment (s_s and s_e) is

decided at the decision location s_1 but this decision could be revised at any time until the control vehicle arrives at the stop just before the skipping segment $s_s - 1$.

Basic Policy -- Policy 1

For policy 1, the first step for the vehicle operator to fulfill the control decision is to drop off the passengers who have their destinations on the skipping segment. This will occur at the stop just upstream of the skipping segment. As soon as the control vehicle has completed this step, the control vehicle will skip all stops on the skipped segment. There is only one skipped segment on the route for the control vehicle.

Policy Alternative -- Policy 2

In contrast, the simulation of policy 2 is significantly more complicated. Initially, the model is used to determine a start and end stop for the skipping segment. Moreover, the definition of the skipping segment can change at any point after the disruption. More specifically, the simulation model is continuously updated with information about the location of the control vehicle and the location of the upstream impacted vehicles, and the number of passengers boarding and alighting these vehicles. By updating this information as the control vehicle moves through the skipping segment, the decision of when to end the control (i.e., the choice of the end stop S_e) can change dynamically. In this way, in contrast to policy 1, policy 2 is an entirely adaptive decision-making process in the sense that, whenever new operating condition information is acquired, the previous stop-skipping control decision can be adjusted accordingly.

Furthermore, for the policy alternative, this simulation study deliberately allows the control vehicle to skip more skipping segments after it has completed one segment. In

this case, additional skipping can be employed to take the advantage of the policy alternative. That is to say, the adaptive decision using the policy alternative will not add any further inconvenience to passengers onboard the control vehicle.

3.3.2. Simulation Study Design

Besides simulating the basic dynamics of the decision-making process and the vehicle trajectory evolution, another significant task of the simulation study is to identify the conditions where the different policies may be preferred.

Similar to the study on the holding control problem, the original stop-skipping control problem is formulated within the context of a deterministic model, but including probabilities associated with the boarding and alighting processes. However, transit operations are exposed to many stochastic factors, particularly for the vehicle travel times between stops. Also of concern is the directional distribution of passengers along the route, which certainly impacts the number and distribution of passengers affected by the stop-skipping control. Therefore, it is also of particular interest to observe how the stop-skipping policies perform within the circumstances defined by different passenger distribution patterns, as well as the differences in vehicle travel time variation. Furthermore, the duration and the location of the vehicle disruption may also considerably affect the performance of stop-skipping control.

Passenger Distribution Pattern Design

Three passengers boarding profiles, with different locations where passenger boarding peaks, have been designed in the holding control problem study in Chapter 2. However, these passenger boarding profiles may not fit in the purpose of this simulation

study for the stop-skipping control problem, because they are designed without paying much attention on the passenger O/D distribution along the route. On the other hand, the passenger O/D distribution pattern along the route may have significant implication on the performance of the two stop-skipping policies. Therefore, another set of discrete passenger distribution patterns were designed particularly for this simulation study, from which other intermediate cases can be interpolated. In specific, three passenger distribution patterns are used:

- Pattern 1: Normal Pattern (symmetric pattern), with the highest passenger load in the middle of the route;
- Pattern 2: Downtown-oriented pattern, with the highest passenger load skewed to the downstream portion of the route; and,
- Pattern 3: The reverse of the downtown-oriented pattern, with the highest passenger load skewed to the upstream portion of the route.

These patterns are also shown in Fig. 3.3, in the special case that vehicle headways are perfectly even across the route. Note that pattern 3 is just the mirror case of pattern 2. Though the three patterns are arbitrarily designed, the loading profiles can be considered the representative of similar patterns in the real world.

Passenger boarding rate profiles for different passenger distribution patterns are shown in Fig. 3.4. Pattern 1 and pattern 2 share the identical passenger boarding rates at all stops. Within the simulation, given the average passenger arrival rate at each stop and the O/D distribution ratio for each pair of stops, Poisson arrivals can be randomly generated and assigned to a destination stop.

Obviously, passenger loading pattern 2 and pattern 3 present two extreme conditions for the stop-skipping policy 2, since its effectiveness relies on the possibility of no passenger calling for alighting at the stops on the skipped segment. Certainly, passenger distribution pattern 3 is expected to highlight difficulties with policy 2 for two reasons: the vehicle boarding peaks on the upstream segments of the route and passenger alighting dominates the remaining part of the route; and, very few passengers can benefit from the control at the downstream route segment once the vehicle passes the peak passenger boarding stops, since the boarding demand is very low. Patterns 2 and 3 are of particular interest also because they indeed exist in the real world on some peak-period routes, and were also analyzed in Jordan and Turnquist (1979) as well as in Furth (1986). Finally, in contrast, the conditions presented by patterns 1 and 2 are likely to be favorable for policy 2.

Travel Time Variation

Similar to the holding control problem in Chapter 2, the coefficient of variation (COV) of vehicle travel time between two adjacent stops is also important for the stop-skipping control problem, because one would expect an adaptive stop-skipping policy, in the case of policy 2, to be sensitive to the amount of travel time variation. Though the range of COV from 0.1 to 0.3 is employed in the holding control problem, to cover a wider variety of situations, the COV is designed to vary from 0.1 up to the maximum 0.4 in this simulation study for stop-skipping control problem. Again, the vehicle travel time is assumed to be normally distributed.

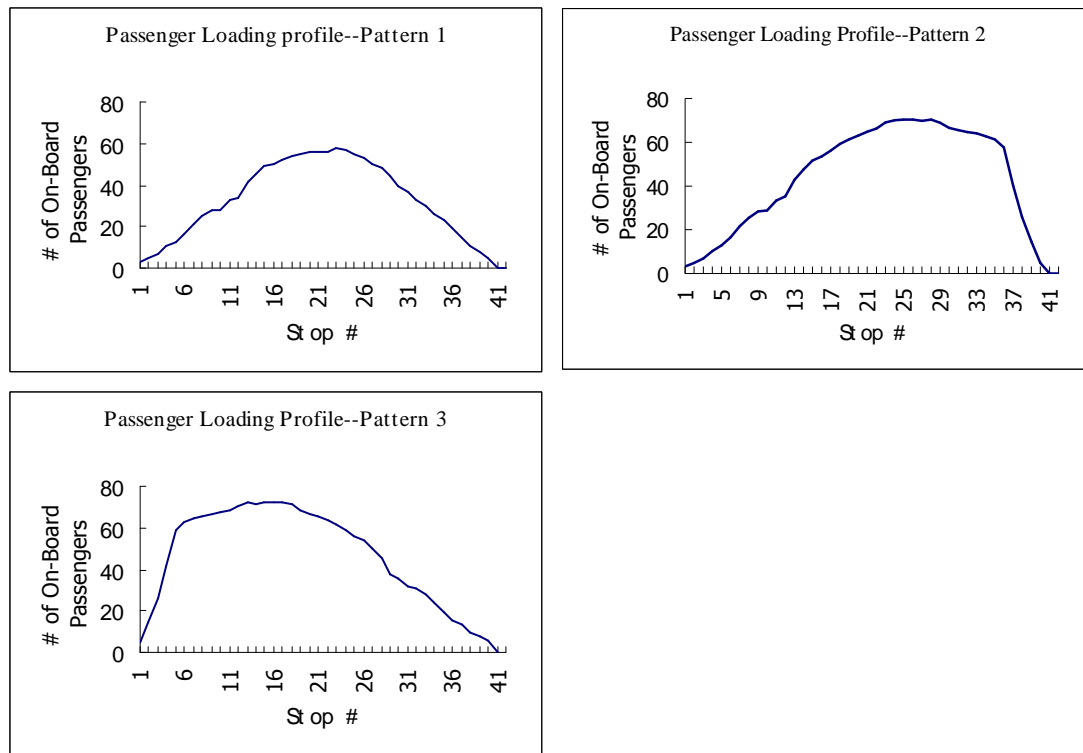


Fig. 3.3. Passenger Loading Profiles (Stop-Skipping Control)

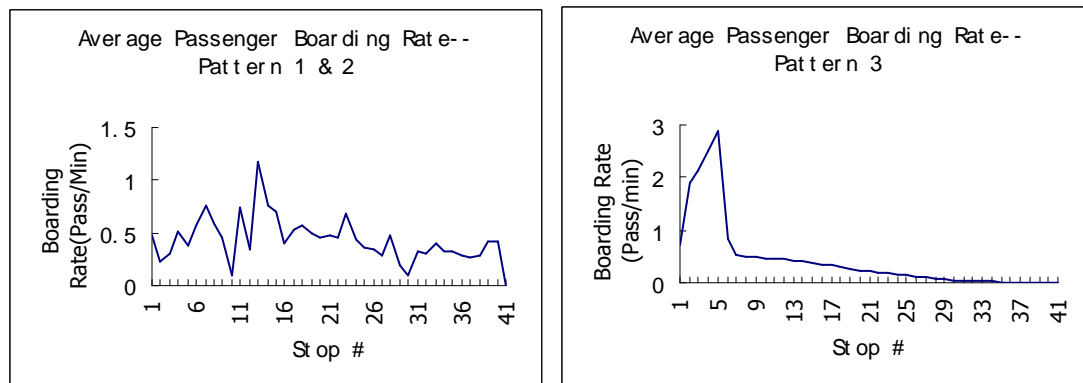


Fig. 3.4. Passenger Boarding Profiles (Stop-Skipping Control)

Other Operating and Simulation Parameters

Other operating and simulation parameters are given in Table 3.1.

Table 3.1. Other Operating & Simulation Parameters

Parameters	Values
One Way Trip Time (min)	Approx. 60 (depending on passenger boarding)
Average Dispatching Headway (min)	7.5
Total Number of Stops/Terminals	41
Disruption Length (min)	1 to 7 minutes, in 1 min increments
Disruption Location	Downstream of stop 1 to 18, in 1-stop increments
AVL Polling Rate (sec)	40
α, β_1, β_2 (sec)	15, 2 and 1, respectively
Number of Impacted Vehicles	4, excluding the boundary vehicle and control vehicle
w_1, w_2, w_3	1, 0.5, and 5, respectively

Some of the parameters in Table 3.1 have also been introduced in Chapter 2. For the test transit route, the one way trip time is set as approximately 60 minutes, including vehicle dwell times; the service headway is set as 7.5 minutes; and the total number of stops is set as 41. Furthermore, the vehicle disruption of different durations (varying from 1 to 7 minutes) at different locations (varying from the location just downstream of stops 1 through the location just downstream of stop 18) is simulated and tested in the simulation study. Also, the objective value is calculated from the control vehicle and 4 impacted vehicles.

3.3.3. Simulation Results

As introduced in last sub-section, with 2 policies, 3 passenger distribution patterns, 18 vehicle disruption locations, 7 disruption lengths, as well as 4 categories of travel time variation level (COV), a total of 3024 cases (or 1512 cases for each policy) are generated to represent all possible combinations of these four factors. For each specific case (or combination), 100 simulation runs are conducted, and the average passenger cost reduction for each case is computed accordingly.

The system under examination can be evaluated at different levels. At the highest level, the system is evaluated by comparing the impacts of the two different stop-skipping policies. Within this context, the simulation is based entirely on CRN (Common Random Number) simulation in the sense that, for each simulation run of each specific case, both policies share commonly generated random factors. The random factors include the travel time between adjacent stops, passenger arrivals at each stop, and the associated passenger destination stops. In other words, for a given passenger distribution pattern, disruption location, disruption length, and travel time COV, 100 identical simulation runs are used for both policies. Similar to that in the holding control problem study, with such a CRN-based simulation study, one may easily conclude that, given that the other factors are the same, the difference in system performance results solely from the two different stop-skipping policies.

At a lower level, within each policy, the simulation also models how the passenger distribution pattern, vehicle disruption location and length, as well as vehicle travel time variability, together affect the performance of the policies. To examine the impact of the system performance resulting from these four factors, a similar CRN-based simulation study is also desirable. However, the very nature of a simulation study also needs to consider the diversity of the realizations of all random factors that may emerge in the real world. As argued earlier, for each case (or combination of the policy, passenger distribution patterns, vehicle disruption location, disruption lengths and travel time variation level), only 100 simulation runs are conducted, and in each simulation run, the random factors, namely the travel time between each pair of adjacent stops, passenger

arrivals at each stop, and the associated passenger destination stops, are randomly generated. Obviously, 100 simulation runs can by no means cover exhaustively the realizations that may result from these random factors, and hence a CRN-based simulation with 100 realizations may be very limited to represent the real world. Therefore, the simulation in this particular study is not CRN-based at the lower level, so that it can cover more realizations of the random factors. This means that as the passenger distribution, the disruption location, the disruption length, and the travel time COV vary, a new set of 100 simulation runs is generated. However, this also makes it difficult to compare the system performance across these levels to analyze how the vehicle disruption location and length, as well as of vehicle travel time variability, affect the policy performance.

The following simulation results are presented by different passenger distribution patterns, since the passenger distribution patterns are believed to have the most significant impact on the performance of different policies among all the major factors.

Normal and Downtown-Oriented Passenger Distribution Patterns

Preliminary analysis has indicated that the disruption location, disruption length and vehicle travel time variation all affect the performance of each of the two different policies. Merely for the convenience of analyzing simulation results, linear regression models are proposed to depict how these three factors determine the policy performance under different passenger distribution patterns as follows (below each regression function are the t-statistics of the coefficients).

Normal Passenger Distribution Pattern and Basic Policy (Policy 1)

$$y = 172 - 32 \cdot x_1 + 122 \cdot x_2 - 112 \cdot x_3 \quad R^2 = 0.88 \quad (3-58)$$

(9.72) (-33.78) (48.95) (-2.53)

Normal Passenger Distribution Pattern and Policy Alternative (Policy 2)

$$y = 354 - 45 \cdot x_1 + 91 \cdot x_2 \quad R^2 = 0.78 \quad (3-59)$$

(17.38) (-32.92) (25.91)

Downtown Oriented Distribution Pattern and Basic Policy (Policy 1)

$$y = 177 - 34 \cdot x_1 + 115 \cdot x_2 - 114 \cdot x_3 \quad R^2 = 0.85 \quad (3-60)$$

(8.91) (-32.08) (43.14) (-2.33)

Downtown Oriented Distribution Pattern and Policy Alternative (Policy 2)

$$y = 275 - 42 \cdot x_1 + 116 \cdot x_2 + 116 \cdot x_3 \quad R^2 = 0.84 \quad (3-61)$$

(12.12) (-34.80) (37.29) (2.10)

Wherein,

y : Mean of the overall passenger cost reduction, namely the mean of all realizations of the objective value “ Z ”. Note the eventual realization “ y ” of the objective value may deviate significantly from the expected value “ Z ” due to the vehicle operation stochasticity;

x_1 : Vehicle disruption location in terms of stop number, which is essentially seen as a continuous variable;

x_2 : Vehicle disruption length in minutes; and,

x_3 : Vehicle travel time COV, ranging between 0.1 and 0.4.

These results indicate first that policy 2 seems to outperform policy 1, all other things being equal. This is seen in the value of the constant term in each equation in (3-58) through (3-61): the average cost reduction is over twice as large for the normal passenger distribution, and about 55% higher for the downtown-oriented passenger distribution.

Also, except for the case of the normal passenger distribution pattern and policy 2 in (3-59), the vehicle disruption location, the disruption length and the vehicle travel time variation all show statistical evidence that they do affect the policy performance significantly in a linear manner. For both policies, the average passenger cost reduction consistently decreases as the disruption occurs further downstream, and increases as the vehicle experiences a longer disruption. These results are fairly intuitive: one might expect the benefits of stop-skipping to decrease as the vehicle moves further downstream; however, a longer disruption results in higher benefits from stop-skipping.

However, the travel time variation shows different effects in the two policies. For the basic policy in (3-58) and (3-60), the total passenger cost reduction decreases as the vehicle travel time is subject to higher variability: this is obvious from the negative coefficient on x_3 . In contrast, the opposite holds for the policy alternative when the passenger distribution pattern is downtown-oriented in (3-61), or the vehicle travel time variability shows negligible effects on the policy alternative with the case of normal passenger distribution pattern in (3-59). With a higher COV, the eventual vehicle operations can evolve in a quite different way than what was expected when the stop-skipping control decision was made. This underlying fact may greatly discount the benefits of the stop-skipping control under policy 1. On the contrary, the policy alternative can adapt the control decision to the continuously-varying operating conditions. This, in turn, can augment the advantages of the stop-skipping control, especially when the vehicle travel time variation is considerable. Yet, based on equation

(3-59), for the normal passenger distribution pattern, the travel time variation is not included because it is not found to be statistically significant.

Also, the absolute values of the coefficients on x_3 are somewhat modest when compared to the intercept terms and the contributions from the other two factors (disruption location and duration) in the regression equations, for both policies. Accordingly, it is not unreasonable to conclude that the performance of both policies individually is nonetheless stable relative to the vehicle travel time variation, from the perspective of the average overall passenger cost reduction. However, the vehicle travel time variation contributes in opposite directions for the different policies, so that the relative performance of the basic policy and the policy alternative is fairly sensitive to the vehicle travel time variability, especially for the downtown-oriented passenger distribution pattern.

A further illustration of how the overall passenger cost reductions based on both policies vary with the vehicle disruption location and disruption length is shown in Fig. 3.5, given that the vehicle travel time COV is 0.2. In these three-dimensional graphs, the vertical axis is the passenger cost reduction y , and the horizontal axes are the disruption location x_1 and disruption length x_2 (in minutes).

Given a certain vehicle travel time variation, policy 2 can always outperform policy 1 when the vehicle disruption occurs further upstream and does not last too long. However, also for policy 2, the passenger cost reduction varies more rapidly with the disruption location than for policy 1. For instance, at a travel time variability level with COV of 0.2, with the normal passenger distribution pattern, for policy 2, as the vehicle

disruption lasts as long as 5 minutes, the total passenger cost reduction falls below 100 passenger minutes when the vehicle disruption occurs downstream of stop 12. On the contrary, even when the vehicle disruption occurs downstream of stop 17, policy 1 can still achieve a passenger cost reduction as much as 108 passenger minutes. On the other hand, the passenger cost reduction for policy 1 increases more rapidly with the disruption length, when compared with policy 2. This makes sense because for policy 2, as the vehicle disruption grows longer, as it becomes very difficult for the control vehicle to skip enough stops to gain the lost time, and to find enough downstream stops to justify the extra cost resulting from skipping stops. Fig. 3.6 illustrates a comparison of how many stops are required to be skipped to restore the service from a disruption of moderate length (4 minutes) at different disruption locations, based on different policies.

Fig. 3.5 basically shows that, as the vehicle disruption location moves downstream, the required number of stops to be skipped based on policy 2 increases sharply. This implies that policy 2 struggles to skip more stops to restore the system, when there are still a sufficient number of downstream stops to justify the stop-skipping control. Also for policy 2, as the vehicle disruption location moves further downstream beyond stop 13, the tradeoff of passenger cost and benefit ‘forces’ the control vehicle to skip fewer stops, since there are not sufficient stops to skip, and there are not enough downstream stops to outweigh the cost incurred by the stop-skipping. It also makes sense that the required number of skipped stops from policy 2 keeps decreasing, since the number of downstream stops that can benefit from the stop-skipping control as the disruption location moves further downstream will also decrease.

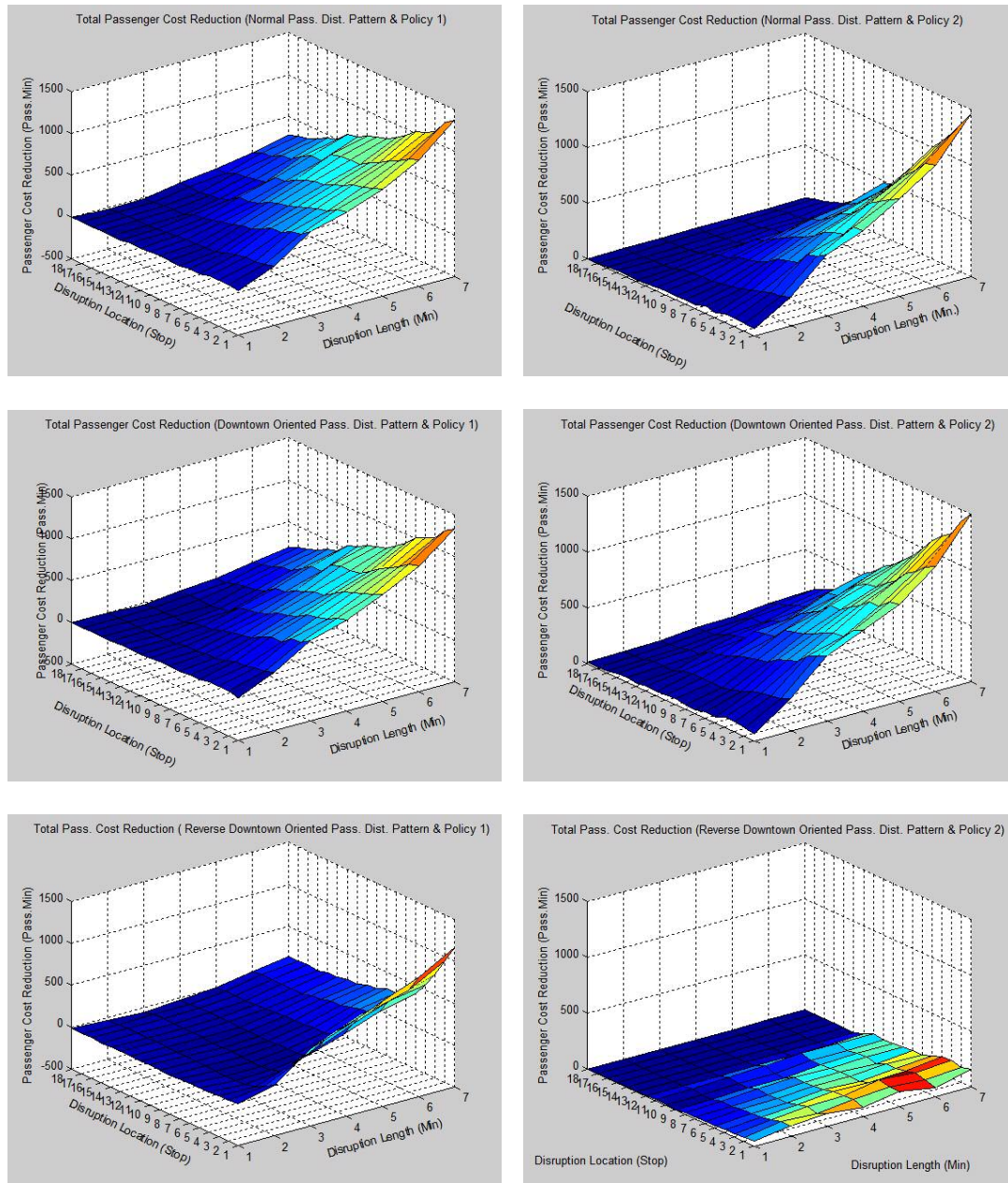


Fig. 3.5. Overall Passenger Waiting Time Reductions (Stop-Skipping Control)

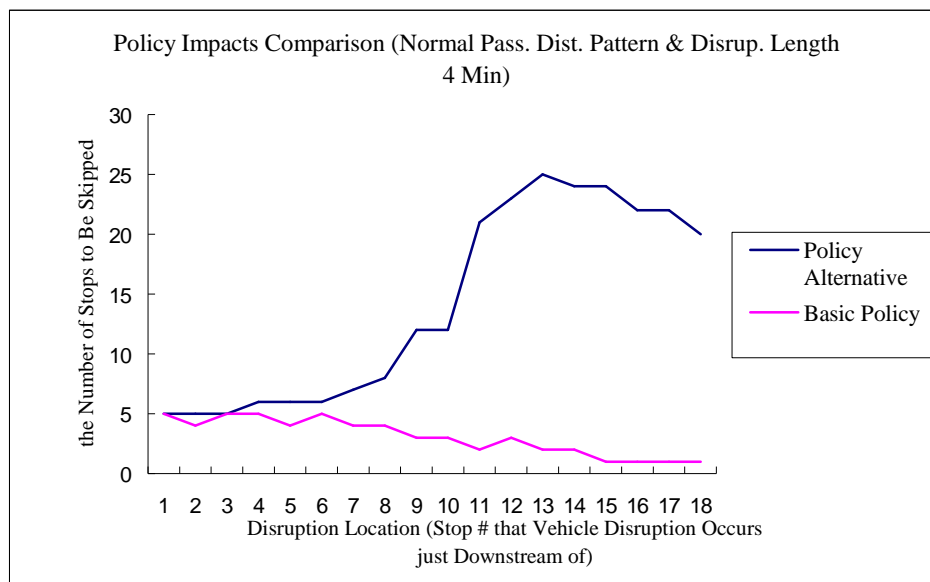


Fig. 3.6. Comparison of Skipping Segment Length and Location

Furthermore, equations (3-60) and (3-61) also show that, for the downtown-oriented passenger distribution pattern, the relative performance of the basic policy and the policy alternative is almost insensitive to the vehicle disruption length. Also, under almost all circumstances, the policy alternative can approximate or outperform the basic policy in terms of overall passenger waiting time reduction. In view of this, it may be concluded that the policy alternative is more desirable for the downtown-oriented passenger distribution pattern than the normal passenger distribution pattern. This is consistent with the underlying mechanics behind the policy alternative, which drives the vehicle to skip the prescribed stops relying on the probability that no onboard passenger calls for alighting at these stops. For a downtown-oriented passenger distribution pattern, passenger alightings occur primarily on the downstream route segment, which creates greater probability for the control vehicle to skip the stops, and the policy alternative performs more efficiently than within a normal passenger distribution pattern.

Reverse of Downtown-Oriented Passenger Distribution Pattern

With the reverse of downtown-oriented passenger distribution pattern, the total passenger cost reduction for the basic policy is still considerable (above 5 passenger hours) when compared with that for policy 2. This occurs when the vehicle disruption does not occur too far downstream to be beyond the peak passenger boarding stops (stops 1 through 5), provided that the disruption is sufficiently long (e.g. 4-7 minutes). For instance, for a vehicle travel time COV 0.2, as the vehicle disruption lasts 7 minutes and the disruption location is just downstream of stop 5, the overall passenger cost reduction is 327 passenger minutes for policy 1. However, for policy 2, in almost all situations, the passenger cost reduction is much smaller (below 150 passenger minutes). This is primarily because the control vehicle has very little chance to skip any stops when the vehicle is heavily loaded and when passenger alighting dominates passenger boarding at the downstream stops. Furthermore, when compared to the normal and downtown-oriented passenger distribution pattern, with a reverse downtown-oriented passenger distribution pattern, the performance of stop-skipping is even less sensitive to the vehicle travel time variability, since the passenger demand on the downstream route segment which is supposed to be affected most by the travel time variability is low.

Overall

Though not shown in Fig. 3.5, by taking advantage of the adaptive decision-making, it seldom occurs in policy 2 that there is a net increase in the total passenger waiting time. On the contrary, the basic policy is not adaptive; hence it may, though unintentionally,

result in an increase in passenger waiting time, especially when the vehicle disruption is not fairly long (1 or 2 minutes) and travel time is subject to large variation (COV 0.4).

The difference in the total passenger time reduction between the two policies can be easily seen with the empirical regression equations (3-58) through (3-61). Fig. 3.7 also gives a comparison of the policy performance for a vehicle travel time COV 0.2 for further illustration. The values in these figures are the difference of passenger cost reduction for the two policies (policy 2 minus policy 1). A positive value on the vertical axis indicates higher passenger costs for policy 1, when compared with policy 2 (i.e., policy 2 is performing better, on average).

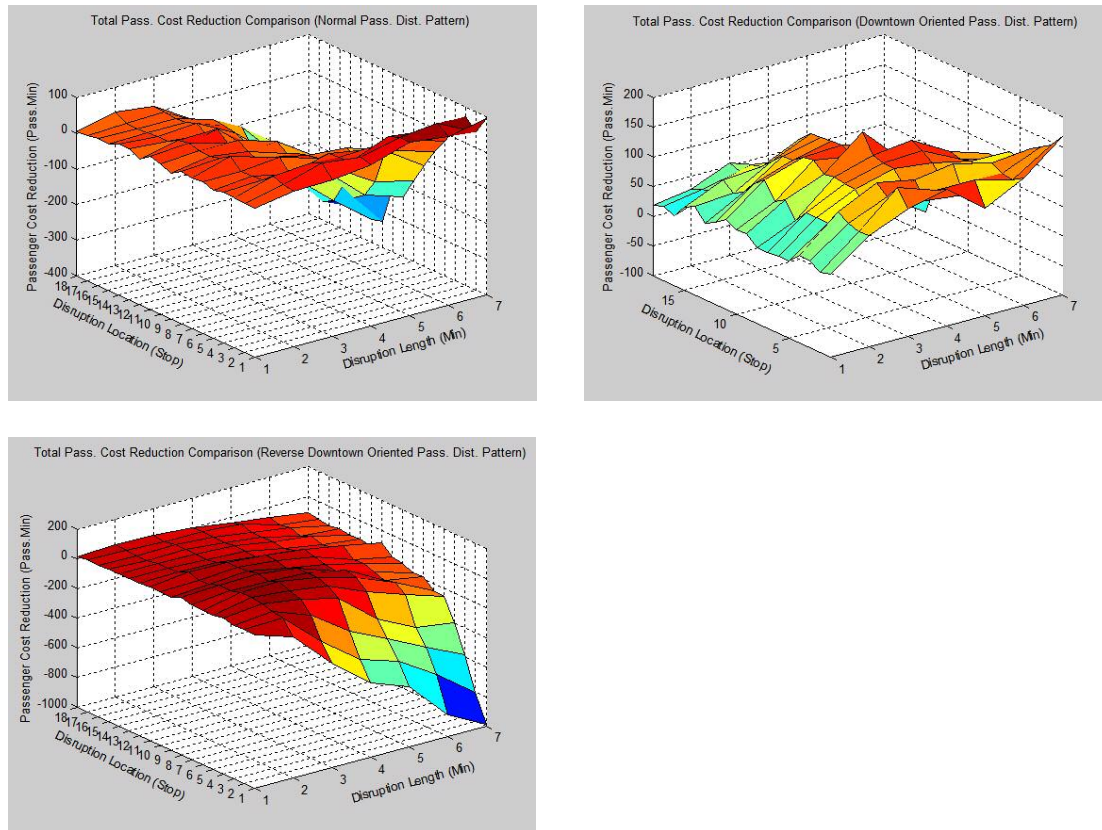


Fig. 3.7. Comparison of Passenger Cost Reduction between Policy 1 and Policy 2 (Stop-Skipping Control)

From Fig. 3.7, for both the normal and downtown-oriented passenger distribution patterns, one may see that, under a majority of circumstances, the performance of the two policies is comparable with an absolute difference in passenger cost reduction of no more than 100 passenger minutes. When considering all levels of travel time variability (COV varies from 0.1 to 0.4, not shown in Fig. 3.7), for the normal passenger distribution pattern, policy 1 seldom shows significantly better performance than policy 2, and this happens only when there are long vehicle disruptions (larger than or equal to 6 minutes), the disruption occurs further downstream (downstream of stop 8), and the vehicle travel time variability is low (COV is below 0.2). Also, the relatively large passenger cost savings from policy 2, compared to policy 1, happens only when the vehicle disruption occurs further upstream (upstream of stop 5) and the vehicle travel time variability is high (COV is above 0.3).

For the downtown-oriented passenger distribution pattern, the performance of the two policies is even closer. In this case, the difference in the two stop-skipping policies is very modest. Policy 2 outperforms policy 1 in an overwhelming majority of situations, because the downtown oriented passenger distribution pattern provides more desirable conditions for policy 2. Nonetheless, the relative passenger cost difference is small. Similar conditions (long disruption lengths, downstream disruption locations, low COV) are needed for policy 1 to outperform policy 2 significantly. Also, an upstream disruption location and large vehicle travel time variability is necessary for policy 2 to achieve greater passenger cost savings over policy 1.

In addition, the required higher travel time COV for policy 2 to outperform policy 1 under both passenger distribution patterns further demonstrates the previous argument that the relative performance of the policies is fairly sensitive to the vehicle travel time variability.

One may also compare the difference in total cost reductions between policy 1 and policy 2, across all cases. Table 3.2 shows the relative difference in the two policies, and the percentage of cases with that relative difference.

Table 3.2. Relative Performance Comparison of Policy 1 and Policy 2

Relative Performance (Policy 2 – Policy 1) (Passenger Minutes)	Percentage of Cases (Combinations)		
	Normal Pass. Dist. Pattern	Downtown-Oriented Pass. Dist. Pattern	Reverse Downtown-Oriented Pass. Dist. Pattern
-300 and less	2.8	0.0	10.1
(-300, -200]	9.1	0.0	3.8
(-200, -100]	10.5	1.0	12.7
(-100, 0)	20.9	8.5	24.4
[0, 100]	49.2	52.4	48.6
(100, 200]	6.5	32.3	0.4
200 and larger	1.0	5.8	0.0
Total	100.0	100.0	100.0

Table 3.2 shows that, under both normal and downtown-oriented passenger distribution patterns, the basic policy and the policy alternative perform similarly in terms of total passenger cost reduction for a majority of cases. For normal and downtown-oriented passenger distribution patterns, there are 70.1 percent and 60.9 percent of cases, respectively, for which the relative performance difference is within ± 100 passenger minutes. However, there are still a significant percentage of cases for which the two policies perform quite differently. For instance, for the normal passenger distribution pattern, for as high as 22.4 percent of the cases, the basic policy can outperform the policy alternative by more than 100 passenger minutes. Also, for the downstream

passenger distribution pattern, for more than 38 percent of the cases, the policy alternative may achieve at least 100 passenger minutes more passenger cost reduction than the basic policy.

For the reverse downtown-oriented passenger distribution pattern, policy 1 consistently approximates or outperforms policy 2. There are 73.0 percent of cases for which the relative performance difference is within 100 passenger minutes, and for almost all remaining cases the policy 1 can achieve at least 200 passenger minutes more extra passenger cost reduction than policy 2. Furthermore, the total passenger cost difference between the two policies can be appreciably large when the vehicle disruption occurs at the beginning of the route and lasts a relatively long time. As an example, when vehicle travel time COV equals 0.2 and the vehicle disruption occurs just upstream of the second stop and lasts for 7 minutes, the passenger waiting time reduction achieved with policy 1 is 1146 passenger minutes and only 157 passenger minutes with policy 2. The difference is 989 passenger minutes, which is quite considerable.

In summary, with the normal passenger distribution pattern, the two policies show selective superiority in terms of the passenger cost reduction they can achieve, depending on the operation conditions and the characteristics of the vehicle operation disruption. With the downtown-oriented passenger distribution pattern, the policy alternative shows some superiority to the basic policy in that the policy alternative can result in larger passenger cost reduction than the basic policy for an overwhelming majority of cases. However, the basic policy dominates the policy alternative for more than fifty percent of cases, and can approximate the policy alternative for another fifty percent of cases.

3.3.4. Capacity Constraint Analysis

Although it is assumed in the problem formulation section that vehicle passenger capacity is not a serious issue, in some cities, the vehicle passenger capacity could become a significant issue during peak hours. Therefore, it is important to assess the sensitivity of this model, which does not include a capacity constraint, to operating conditions where capacity is an important issue. To do this, the simulation is applied in a context where the decision-making methodology remains the same, but the capacity constraint is explicitly considered in the simulation to mimic more realistic vehicle operations. The simulation scenarios are applied to the downtown-oriented passenger distribution pattern case, and the simulation results are compared with those presented previously.

This comparison is summarized in Table 3.3. In this table, the difference in passenger cost is compared for policy 1 and policy 2. In this case, the cost is broken out into two parts. The out-of-vehicle passenger cost is computed as the realized values of the first and third terms in equation (3-1) for policy 1, or the first term in equation (3-2) for policy 2, respectively. The in-vehicle passenger cost is calculated using the realized values in the second term of equations (3-1) and (3-2), for policy 1 and 2, respectively. The average passenger cost and the standard deviation for the capacity-constrained scenarios and the unconstrained scenarios were calculated separately, and the difference is directly presented in Table 3.3.

Table 3.3. Comparison of Simulation Results with/without a Capacity Constraint

		Total Passenger Cost Difference (Passenger-Minutes) (Capacity Constrained ¹ – Capacity Unconstrained ¹)	
		Out-of-Vehicle Passenger Cost	In-Vehicle Passenger Cost
Control Under Policy 1	Average	–3.27 (2156.04 – 2159.31) ²	–1.8 (5281.8 – 5283.6) ²
	SD	513.6 Constrained 513.1 Unconstrained	359.0 Constrained 358.7 Unconstrained
Control Under Policy 2	Average	33.1 (2146.1 – 2113.0) ²	–14.9 (5231.8 – 5246.7) ²
	SD	495.5 Constrained 515.9 Unconstrained	355.6 Constrained 364.4 Unconstrained

¹Capacity is assumed to be 75 passengers/vehicle, including standees.

²Included in the parenthesis is how the number (e.g. –3.27) is calculated.

In Table 3.3, one notes that the difference in overall out-of-vehicle cost is basically negligible for policy 1. In contrast, for policy 2, the overall out-of-vehicle cost increases. This might be expected, since the capacity constraint results in more passengers waiting for an available space on a bus, and the policy 2 is subject to more influence from the capacity constraint because it basically relies on the probability that a stop does not have a passenger alighting. Capacity constraints surely affect these probabilities. The negative value of the overall out-of-vehicle passenger cost for policy 1 may simply result from the fact that increased waiting cost for the passengers left over by the vehicle might also reduce the waiting cost for the passengers at the downstream stops at the same time. Nevertheless, the results for both policies are not statistically different from zero, based on the value of the standard deviation. In terms of the in-vehicle cost, the capacity constraint results in slight reductions in cost relative to the unconstrained case, as the reduction in on-board passengers results in lower dwell times at stops for the on-board passengers.

Overall, the statistics in Table 3.3 suggest that explicit consideration of a vehicle capacity constraint may not alter the overall conclusions substantially. As a result, the performance of a stop-skipping control decision made based on the methodology proposed in this paper could be fairly stable, if vehicle passenger capacity is sufficiently large to accommodate passenger demand in normal situations. Of course more analysis of this capacity-constrained case is warranted, and is left for future study.

3.4. Conclusions

The study of the real-time stop-skipping control problem in this chapter moves a step further to investigate the possibility of implementing stop-skipping control in a real-time manner through proposing a policy alternative for better guiding the control vehicle to skip stops. Theoretically, the model formulated in this study can respond to transit service disruptions more rapidly than those developed in the past studies, where the stop-skipping control is formulated as a vehicle dispatching problem at the terminal.

Based on a simple analysis of the basic policy that is extensively used in practice and a majority of previous studies for stop-skipping control, the policy alternative is developed and assumed to be more appropriate for a real-time application. Also, the policy alternative can avoid the inconvenience that the basic policy may bring to some onboard passengers who have the destinations on the skipping segment. Specifically, this policy alternative allows the control vehicle to drop off the onboard passengers at the same stops as they planned to alight, even when their destination stops are already prescribed to be skipped by the control vehicle. In contrast, the basic policy dictates that

the control vehicle must skip all prescribed stops, and it must drop off the on-board passengers that have the destinations among the prescribed skipped stops.

This chapter has provided a deterministic modeling framework for deriving the start and end stops for a skipping segment, under both policies. In addition, the study in this chapter has shown that such a model can be solved to optimality in real time, using an explicit enumeration method. With this problem formulation, a simulation study is conducted to examine how the performance of the two stop-skipping policies varies with the passenger distribution pattern, the vehicle disruption location, the vehicle disruption length, and the vehicle travel time variability. These simulation results suggest:

- For both normal and downtown-oriented passenger distribution patterns, the policy alternative can approximate the basic policy in a majority of cases, in the sense that they both perform similarly in terms of the total passenger cost reduction. However, considerable performance difference of the two policies can still be seen in a significant percentage of cases. For the normal passenger distribution pattern, as the vehicle disruption occurs near the start of the route and does not last very long, the policy alternative can outperform the basic policy. Otherwise, the basic policy can achieve more passenger cost reduction than the policy alternative. For the downtown-oriented passenger distribution pattern, a similar conclusion still holds; furthermore, the policy alternative can outperform the basic policy in an overwhelming majority of situations. In this sense, the downtown-oriented passenger distribution pattern forms the most desirable condition for the policy alternative.

- For the reverse downtown-oriented passenger distribution pattern, the policy alternative seldom has a chance to outperform the basic policy, since the strongly-skewed passenger boarding profile provides little possibility and no incentive for the control vehicle to skip any stops.
- From the perspective of the overall passenger waiting time reduction, the performance of the two policies varies with the vehicle travel time variation in opposite ways. Vehicle travel time variation discounts the benefits that can be made under the basic policy, but can augment the total passenger cost reduction resulting from the policy alternative. In terms of their contributions to the total passenger cost reduction from stop-skipping control, both policies are not as sensitive to the vehicle travel time variability as to the vehicle disruption location and the disruption length. However, the relative performance of the two policies is fairly sensitive to the vehicle travel time variability. The vehicle travel time variability could contribute significantly to the performance difference of the two policies.
- The policy alternative, which is a truly adaptive real-time control strategy, can avoid increases in passenger cost, since it has the advantage to adapt the stop-skipping control to the continuously varying operation condition. The statistical evidence shows that higher vehicle travel time variability can create more desirable conditions for the policy alternative. However, the basic policy could result in an increase in passenger cost, especially when the vehicle disruption is not very long and travel time is subject to large variation.

Finally, the policy alternative may be more preferable from the perspective of the transit agencies in the sense that it does not need to force any passengers to alight off the vehicle. This may have significant implication to the transit agencies, especially if they consider implementing stop-skipping control regularly and in a real-time manner.

CHAPTER 4 VEHICLE DISPATCHING WITH SWAPPING AT A TRANSFER TERMINAL

This chapter takes a significant departure from the study on the single route based operations control problems in the previous two chapters to deal with the problem of vehicle dispatching with swapping at a transfer terminal, where multiple transit routes intersect. Furthermore, differentiating this with the other two operation control problems in this dissertation, the vehicle dispatching problem in this chapter addresses infrequent transit service. Many previous studies have suggested that operations control may help to improve the transfer synchronization and reduce overall passenger cost in terms of their waiting time at either the transfer terminal or the downstream stops, primarily for infrequent service.

4.1. Introduction

Passenger transfers have traditionally been a major concern of public transit agencies. A public transit transfer terminal synchronizes the arrival of incoming vehicles with the departure of outgoing vehicles so as to reduce the possibility of missing connections and the transfer delay for passengers. Transfer synchronization at the terminal is primarily achieved by optimally designing the timetable for each route so that the vehicle arrival and departure for each route can be properly matched. US UMTA (1983) presented evaluations for some transfer systems under operation, and Vuchic et al. (1981) provided fairly comprehensive procedures for design of a timed transfer system. Hall (1985) developed and evaluated a model for scheduling vehicle arrivals at transportation terminals where vehicles are randomly delayed en route. He developed

optimal "slack" time between the scheduled arrival for a feeder line and the scheduled departure for a transfer line, and concludes that coordinating arrivals with departures is the most important when the headway is large relative to the average vehicle delay. Lee and Schonfeld (1991) also investigated the slack time optimization at a transit terminal so as to reduce the transfer passenger cost for simple systems with transfers between one bus route and one rail line. Bookbinder and Desilets (1992) combined a simulation procedure with an optimization model to determine timetables that minimize the overall inconvenience to passengers, represented by a disutility function, on the basis of a given transit network and stochastic bus arrival times. Knoppers and Muller (1995) investigated the possibilities and limitations of coordinated transfers between feeder lines and the connecting lines in public transit. Their results show that the coordination of timetables intends to achieve a tradeoff, between the transfer cost and the possible cost incurred to the downstream segments on the connecting line, by optimizing the synchronization control margins. Also, they showed that the coordination of timetables is only worthwhile when the schedule deviation on the feeder lines at the transfer station is less than 40 percent of the headway on the connecting line.

As traffic congestion has become increasingly common in central urban areas, transit agencies may find it more difficult to maintain the vehicle arrival/departure time synchronization at a transfer terminal. Accordingly, much research has also been conducted to apply operations control at a terminal to enhance the synchronization of bus departure. Among this research, holding control strategies, being thoroughly discussed in Chapter 2, again are extensively investigated and used for transfer synchronization

purposes. Abkowitz et al. (1987) developed a computer program to simulate four different transfer strategies, including unscheduled transfers, scheduled transfers without vehicle waiting, scheduled transfers where the lower frequency bus is held until the higher frequency vehicle arrives, and scheduled transfers when both buses are held until a transfer event occurs. This study found that the double holding strategy is advantageous whenever the headways on intersecting routes are compatible; otherwise, the no holding strategy is preferred. Dessouky et al. (1999) developed bus delay and lateness forecasting models and applied them for evaluating a variety of holding-based operation control strategies at transfer stations through simulation. Their results, based on empirical data collected in Los Angeles, show that segment delay of the infrequent bus lines is negatively correlated with the lateness at the segment start point, which indicates that buses can catch up when they fall behind schedule. Also, the simulation analysis shows that the most significant benefit of bus tracking technology is achieved when the buses experience major delay, especially when only a small number of connecting buses exist. Hall et al. (2001) examined three vehicle dispatch policies: (1) the vehicle is dispatched immediately; (2) the vehicle is held until a predetermined time in anticipation of the connecting buses; and, (3) the vehicle is held until a predetermined time, but allowing buses to depart as soon as all connecting buses have arrived. The results show that the objective function, the passenger waiting time, can show four different behaviors with regard to the minima for identically and normally distributed vehicle arrival lateness. Further, the results also show that more flexible policies, e.g. the third policy, offer some potential for improvement in the passenger waiting time, but the practical difference

between the policies is small. More recently, Dessouky et al. (2003) compared the terminal holding control strategies that depend on technologies for communication, tracking and passenger counting, to those that depend solely on local information. Results demonstrate that technology can help in determining optimal holding times that minimize the average passenger trip time, by balancing the time saved for late-arriving transfer passengers against the delay for passengers who are either already on-board or will board at subsequent stops. The benefits of the technology are improved especially for the case of a large number of connecting buses, little schedule slack time, and relatively large headway.

In addition to holding strategies at the terminal, the operations control study for transfer synchronization in this dissertation also considers the real-time strategy of dispatching a vehicle from one route onto another at the route terminal. In the context of service planning, this is called interlining. Interlining allows the use of the same revenue vehicle and/or operator on more than one route without going back to the garage (Sacramento Regional Transit, 2003). Interlining is often considered as a means to minimize vehicle requirements as well as to provide transfer enhancement for passengers. For interlining to be feasible, two (or more) routes must share a common terminus or be reasonably proximate to each other to minimize deadheading (Sacramento Regional Transit, 2003). From this general definition of interlining, there are two points deserving attention. First, interlining is able to minimize the number of vehicles required. For this purpose, interlining has to be built in the timetable as a routine behavior for the involved routes. However, the use of interlining can make the vehicle schedule fairly complex. On

the other side, interlining also has the ability of enhancing the passenger trip connection at terminal, when many transfers are made from one route to the other.

As an extension of this concept, there may be some potential of applying interlining as an operation control tool on a real-time basis. However, there is little research that reveals any explicit consideration of this strategy. Furth and Nash (1985) have conducted a study of “pooling” vehicles at a common dispatching terminal to dispatch vehicles among several routes on a real-time basis. The intent of vehicle pooling is to improve the reliability of service. The real-time dispatching, with the vehicle pools is essentially supported by swapping vehicles from different routes. The study result shows that vehicle “pooling” can improve the transit schedule adherence and can facilitate interlining to reduce the need for slack time at the terminal. In addition, some swapping research work can be found in the air transportation area. Jarrah et al. (1993) created a decision support framework for airline flight cancellations and delays. Swapping aircraft among scheduled flights, as a basic tool in the framework, is considered in their study to respond to the aircraft shortage that occasionally occurs during day-to-day airline operation. Talluri (1996) investigates the swapping application in daily airline fleet management. In this study, a simple algorithm for making the swapping (changing the assignment of a specified flight leg to different equipment) is developed while the composition of equipment overnighing at the various stations is not affected.

For the purposes of both being consistent with Furth and Nash’s (1985) study and not conflicting with the typical terminology used in timetable design, the term “vehicle swapping” is used throughout this study to describe real-time “interlining”.

From the literature review above, some general conclusion can be drawn.

- The passenger transfer problem can be addressed in part by the timetable design; but,
- As transit operation becomes more variable, holding, as a real-time operation control strategy, can help stabilize and enhance the timed transfer at a transfer terminal;
- A holding strategy can be advantageous for the transferring between infrequent transit services, although there is no significant benefit which can be achieved by applying holding for the transferring between high-frequency transit service; and,
- Swapping vehicles among different routes may be advantageous to reduce transfer inconvenience through more flexible vehicle dispatching.

These conclusions naturally raise two underlying questions: can the typical holding-based control strategy at the transfer terminal be augmented by a “swapping” control strategy; and, is there any other beneficial by-product from such a strategy combination? Accordingly, this portion of the dissertation examines the potential of applying vehicle swapping, along with holding, on a real-time basis, and investigates the potential improvements from the strategy combination in terms of passenger waiting time at both the transfer terminal and the downstream stops/stations. This is different than Furth and Nash’s (1985) study, in that schedule adherence is the major measure of performance in their study. Instead, in this study, the transit user perspective is represented by the passenger waiting cost, which may make the real-time vehicle swapping more applicable.

Moreover, Furth and Nash's (1985) study evaluates the benefit of vehicle pooling in an analytical way, while this study primarily uses simulation to investigate the potential of vehicle swapping on a real-time basis. Finally, this research examines the performance of real-time vehicle swapping combined with a holding strategy, rather than as a unique control strategy.

The remaining part of this chapter will investigate the potential benefits of incorporating vehicle swapping into the strategy of vehicle dispatching at the transfer terminal, and formulate the problem of vehicle dispatching with swapping as a deterministic optimization problem to optimize the vehicle holding times and vehicle swapping strategy at the transfer terminal, so that the total cost of the passengers originating at the transfer terminal and the downstream stops and the passengers transferring at the terminal can be minimized. For a limited number of transit routes intersecting at the transfer terminal, vehicle swapping scenarios are enumerable. Furthermore, given a vehicle swapping scenario, the problem of vehicle dispatching with swapping at the terminal degrades to a holding problem for transfer purpose, as investigated in many previous studies, e.g. Dessouky et al. (1999). For such holding problem, the early arriving vehicles can only be held to a limited number of time points when the late vehicles arrive. Therefore, the vehicle holding times are also enumerable. Enumerable vehicle swapping scenarios and enumerable vehicle holding times make the exhaustive search method again appropriate to solve the problem of vehicle dispatching with swapping to optimality. As in the last two chapters, a simulation study is again conducted to compare the performance of two vehicle dispatching strategies, namely the

holding-only strategy and the strategy of both holding and swapping. These two strategies will be introduced in detail in the following sections of this chapter. The main flow in this chapter can also be depicted as in Fig. 4.1.

Again, the simulation study conducted in this chapter is very similar to the simulation studies conducted in previous two chapters, starting with impacting factor identification, followed by scenario design, then simulating each scenario for a specified number of times for both vehicle dispatching strategies, namely holding only strategy and the strategy of holding and vehicle swapping, finally comparing the total system cost reduction from both strategies in terms of the waiting time of both originating passengers and transfer passengers as well as the operator cost. Furthermore, the simulation study is the most important part in this chapter, and much more detail of the simulation study will be given in the simulation section. Therefore, a diagram similar to Fig. 2.2 for depicting the flow of simulation is not presented here to avoid redundancy.

Specifically, the remainder of the chapter is organized into four sections. Section 4.2 describes vehicle dispatching strategies at a transit transfer terminal and their potential effects on transit service. Section 4.3 provides the mathematical formulation for the problem of optimizing the vehicle dispatching strategy by integrating real-time swapping with vehicle holding. Section 4.4 discusses the simulation experiments and the sensitivity analysis. Finally, the findings and conclusions are summarized in the final section, and additional research to follow this study is suggested.

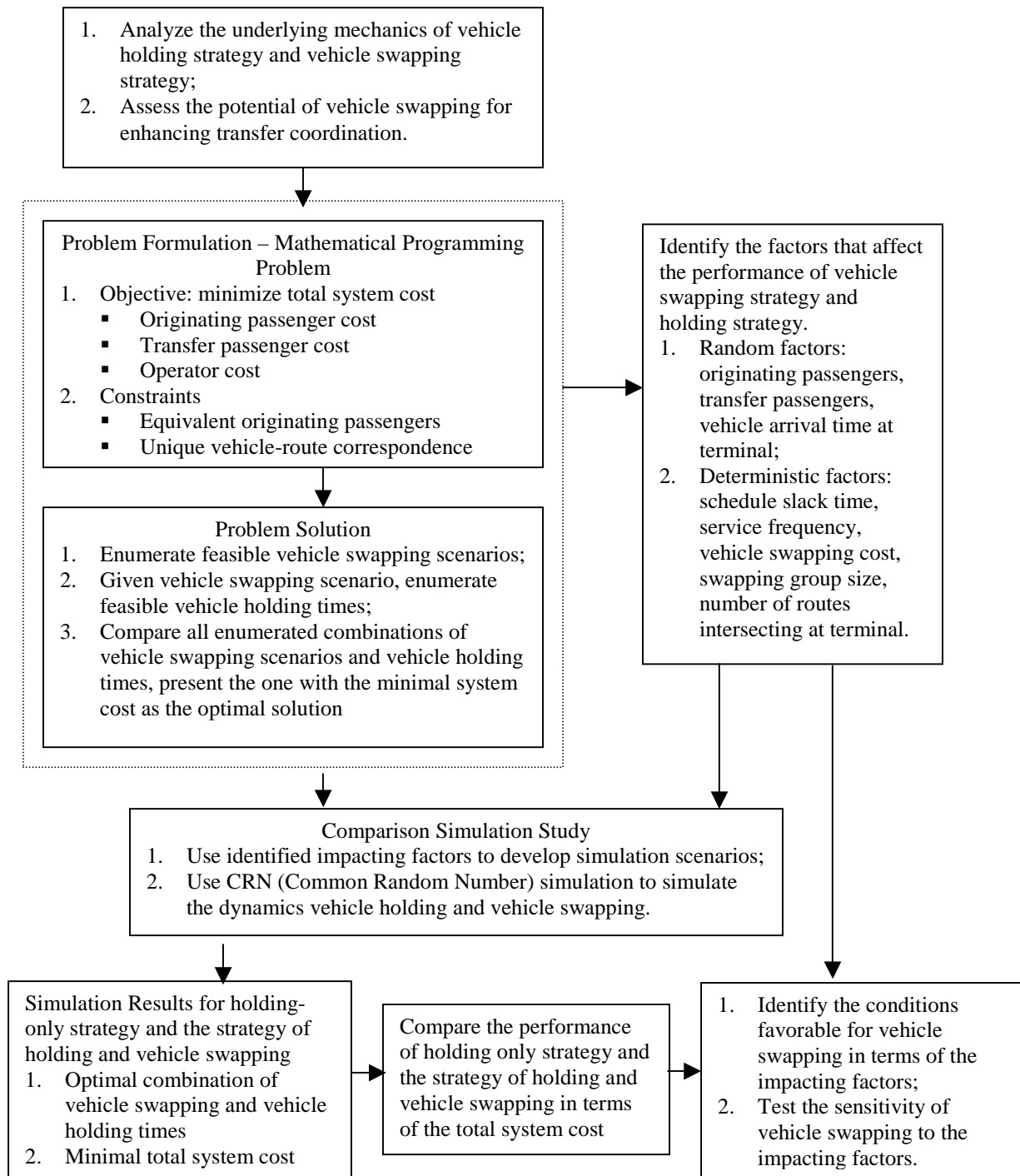


Fig. 4.1. Chapter Flowchart (Vehicle Dispatching with Swapping)

4.2. Vehicle Dispatching Control Strategies

For the case of multiple synchronized transit routes terminating at a common transfer terminal, a late vehicle arrival would incur two types of extra cost to passengers. The passengers originally waiting at the terminal (or downstream) for this vehicle would experience extra waiting time; and, passengers onboard the late vehicle making a transfer may miss their connection, which implies that they would have to wait for the next arriving vehicle on a transfer route. The following two sub-sections briefly describe the dynamics of the holding and swapping strategies as operation control tools to reduce these passenger costs.

4.2.1. Holding

The overall passenger boarding can be divided into two categories: originating passengers, which represent the passengers originating at the terminal or at downstream stops on one particular route; and, transfer passengers from other routes at the terminal. It is universally recognized that a holding strategy at the transfer terminal can be applied to balance the extra waiting time cost experienced by the originating passengers and the transfer passengers. This is done by holding the on-time or earlier-arriving vehicle so as to avoid the transfer passengers missing the connection, at a moderate cost for the originating passengers who experience extra waiting time (the holding time). Accordingly, three points merit some attention:

- Once holding is applied, the originating passengers experience extra waiting time.

- Originating passengers are not an active factor triggering holding control at a transfer terminal, but a passive factor considered for tradeoff with the benefits to transferring passengers.
- Empirically, the number of transfer passengers is often less than the originating passengers at the terminal and the downstream stops. Therefore, infrequent transit service may justify the holding strategy more than frequent service.

4.2.2. Vehicle Swapping

In contrast to the holding strategy, the vehicle swapping strategy follows a different logic in which both the originating passengers and the transfer passengers may actively trigger the control, to achieve an overall minimum passenger cost. This logic works in a more complicated way to reduce the extra waiting time cost for both originating and transfer passengers. A simple example below may help clarify the swapping mechanism and the differences with the holding strategy.

Assume that two transit routes A and B terminate at a common transfer terminal, and are scheduled to arrive at the terminal at the same time. The actual vehicle arrival times are T_A and T_B with *Scheduled Arrival Time* $< T_A < T_B$. The next vehicle arrival times are τ_A and τ_B ; the originating passengers boarding at the terminal and downstream stops are NP_A and NP_B ; the transfers passengers attracted by route A and route B are TP_{BA} and TP_{AB} respectively; and, no slack time is assumed to be applied at the terminal, meaning that the vehicles will depart immediately after passenger alighting and boarding are completed if they did not arrive before the scheduled time. For this case, holding of a vehicle on route A until the vehicle B arrives can be justified only when

$$(\tau_A - T_A) \cdot TP_{BA} > (T_B - T_A) \cdot NP_A \quad (4-1)$$

Inequality (4-1) states that the additional wait incurred by transfer passengers who miss the connection exceeds the additional wait to the originating passengers by holding until T_B . However, if swapping is applied, a vehicle from route A is dispatched onto route B before T_B , routes A and B will benefit in three ways:

- The originating passengers NP_B experience less wait time;
- There is no extra delay incurred to TP_{BA} ; and,
- The waiting time to TP_{AB} is reduced.

The only additional cost is for the originating passengers on route A , NP_A , who must wait until T_B to board. Therefore, swapping is justified when

$$(\tau_A - T_A) \cdot TP_{BA} > (T_B - T_A) \cdot NP_A - (T_B - T_A) \cdot (TP_{AB} + NP_B) \quad (4-2)$$

When compared with inequality (4-1), inequality (4-2) includes a second term on the right-hand side representing a reduction in waiting time for passengers on route B when swapping occurs. Essentially, in the case of only two routes, all of the improvements achieved by holding can be achieved by applying swapping instead; i.e. the swapping strategy always outperforms the holding strategy for this case.

4.2.3. Holding and Swapping

However, as the number of synchronized routes exceeds two, more passenger groups will be involved, and the conclusion above that swapping performs better than holding is not as clear. With the same example by adding another route C in the synchronized route group, and assuming that the arrival time for route C is T_C

(*Scheduled Arrival Time* $< T_A < T_C < T_B$), and the passenger transferring from route C to route A and B are TP_{CA} and TP_{CB} respectively, the implementation of strategies would get much more complicated. Route A can be held until either T_C or T_B , and swapping can be organized into four ways, AB , AC , BC and ABC . Moreover, a new strategy, which is essentially the combination of holding and swapping, could come out as follows:

A vehicle from route A is swapped onto Route B , but this vehicle can also be held to depart until T_C , so as to benefit TP_{CB} , which would miss the connection without any holding applied to this vehicle.

With even this simple example, it is clear that swapping can be extended to the case of more than two routes, and, moreover, swapping may not completely achieve the same passenger cost reduction as holding, simply by replacing holding with swapping without any support from holding. In the example, swapping and holding can work together, but this is not easily seen in the case of only two routes. Furthermore, for a situation with a relatively large number of synchronized transit routes, the vehicle dispatching decision could become extremely complicated due to the potential combination of both swapping and holding. However, one can also see that the role played by either holding or swapping is consistent: holding tends to reduce waiting cost only for transfer passengers, and swapping does so for both originating and transfer passengers.

4.3. Vehicle Dispatching Process & Model Formulation

This section provides more clarification for the application of these control strategies, through a more formal model formulation.

4.3.1. Assumptions and Vehicle Dispatching Process

This vehicle dispatching study investigates real-time vehicle dispatching strategies based on the real-time transit vehicle locations and relevant travel time forecasting models with AVL technology available. Furthermore, three more assumptions are made:

- The route that each vehicle is designated to serve and the vehicle holding time on each route represent the vehicle dispatching decisions made at a transit terminal. These decisions are assumed to be made iteratively on a frequent basis to be adaptive to the traffic condition change and other major operations factors, e.g. every time interval consistent with the AVL data polling process.
- Due to the frequency of this decision-making process, travel time variability does not need to be considered explicitly when the vehicle arrival time at the terminal is predicted. This assumption is somewhat argued in previous chapters.
- At the terminal, each route has an individual terminal area (a bus bay, for example), and the travel times for a vehicle to travel between areas within the terminal are negligible, so that vehicles moving from one route to another can do so within the decision cycle.

With the assumptions above, the vehicle dispatching process at the transfer terminal is described as follows:

Step 1: Put all vehicles serving the routes that are synchronized at the terminal into a decision set O .

Step 2: Check the number (m) of vehicles in the decision set O , if $m = 0$, go to Step 7.

Step 3: Set the earliest scheduled arrival time t of the vehicles within O to zero.

Step 4: Set all arrival times of the vehicles within O which arrived earlier than t to zero also. Predict the arrival times for other vehicles which have not arrived at the terminal by t (T_A , T_B , etc). Predict the next arrival times for all vehicles following those within O (τ_A , τ_B , etc).

Step 5: Make a vehicle dispatching decision for each vehicle within O based on the arrival times which are set or predicted in Step 4.

Step 6: Check each vehicle within O . If the larger of its actual arrival time and the scheduled departure time on the route it is to serve has passed, and the designated holding time on that route from the last decision-making is less than the cycle time for decision-making, hold the vehicle for the designated holding time at the area of the route it originally serves, then dispatch it from the area of the route it is to serve, and eliminate it from O . Go back to Step 2.

Step 7: Wait until the time of making next control strategy decision, then return to Step 3.

4.3.2. Model Formulation

In the vehicle dispatching process, making the vehicle dispatching decision for each vehicle in Step 5 is the key. Implementing control strategies at a transfer terminal aims to reduce overall passenger cost, which consists of the costs to both originating passengers at the terminal and at impacted downstream stops, and passengers transferring between routes. These two types of costs constitute the primary elements to be minimized. However, swapping may also incur additional implementation cost, because swapping

requires each vehicle operator to be familiar with the service on more than one route, which adds more burden to the operator training program and/or more advanced technologies to assist the vehicle operators to drive on other routes. Moreover, swapping may incur a violation of operators' work schedule, vehicle maintenance activities and the like. Therefore, a threshold cost for each swap must also be taken into account in the problem formulation.

In light of the analysis in previous sections, the overall passenger cost is primarily determined by the vehicle arrival times and the number of transfer and originating passengers. For the purpose of model formulation, a complete list of variables representing the significant passenger costs and also the swapping cost for vehicle dispatching strategies are defined as follows. Since the study in this chapter focuses on the synchronization and passenger transfer among multiple routes, rather than the vehicle operations on only one single route, a separate variable index is preferred here that may differ slightly than that employed in the previous two chapters.

N : Total number of synchronized transit routes intersecting at the transit transfer terminal.

NP_i : The equivalent originating passenger boardings at the terminal and downstream stops/stations on route i . Herein, the 'equivalent' is used to account for the different waiting times experienced by the originating passengers on different stops/stations due to slack time built in the timetable.

NTP_i : The originating passenger boardings at the terminal on route i .

$NSP_{i,k}$: The originating passenger boardings at downstream stop k on route i .

$TP_{i,j}$: Passengers transferring from route i to route j ;

S_i : The number of downstream stops for route i .

$Slack_{i,k}$: The average slack time built in the timetable for stop k on route i .

$\psi_{i,k}$: The equivalent “weight” for the originating passengers at the downstream stop k on route i , with the assumption that the weight for originating passengers at the terminal is 1.

SA_i : The scheduled vehicle arrival time for route i .

A_i : The actual vehicle arrival time for route i .

$TSlack_i$: The time difference between the scheduled arrival time and scheduled departure time at terminal for route i (i.e. terminal slack time).

D_i : The scheduled vehicle departure time for route i , which is the summation of SA_i and $TSlack_i$.

$WT_{i,j}$: Waiting time for $TP_{i,j}$.

WT_i : Waiting time for NP_i .

H_i : Holding time on route i . To be consistent with the previous studies on transfer synchronization, the holding time in this chapter is defined as the time that the vehicle is held at the terminal after the scheduled departure time if the vehicle arrives earlier than the scheduled departure time, or the time after the vehicle arrival time otherwise;

$\delta_{i,j}$: Binary variable describing a vehicle swap. If the vehicle originally serving route i is assigned to serve route j , then it equals 1, otherwise 0.

τ_i : The predicted departure time for the next coming vehicle on route i with or without swapping and holding.

$DY_{i,k} | WN_i$: The average vehicle departure lateness at stop k on route i conditional on the departure lateness at the dispatching terminal. This value can be concluded from the historical AVL data. For the sake of simplicity, in this particular study, the relationship between this lateness value and the actual departure lateness occurring at the dispatching terminal is assumed to vary following a piecewise function, which is reduced only by the slack time built in the timetable at each stop/station.

IC : The threshold cost incurred by each swap. Herein, a swap occurs when a transit route is served by a vehicle originally scheduled for another route. It is called a threshold cost in that, if the decision finds that the passenger cost savings from swapping is less than this swapping cost, the swapping will not occur.

With these variable definitions, the vehicle dispatching problem at a transfer terminal can be formulated into the model as follows. The vehicle holding time on the i^{th} route H_i and the binary variable $\delta_{i,j}$ representing any swap are the two sets of decision variables.

The objective function is:

$$\text{Minimize } Z = \left\{ \sum_{i=1}^N NP_i \cdot WN_i + \sum_{i=1}^N \sum_{j=1}^N TP_{i,j} \cdot WT_{i,j} + IC \cdot \sum_{j=1}^N \sum_{i=1 \& i \neq j}^N \delta_{i,j} \right\} \quad (4-3)$$

In this objective function, the first component represents the originating passenger cost; the second term defines the transfer passenger cost; and the last term determines any cost for swapping.

The scheduled vehicle departure time for each route is simply the sum of the scheduled vehicle arrival time and the built-in slack time at the terminal.

$$D_i = SA_i + TSlack_i, \quad \forall \text{ route } i \quad (4-4)$$

The expected extra waiting time for the equivalent originating passengers at the terminal and downstream is the difference between the actual vehicle departure time and the scheduled vehicle departure time (i.e., the departure time lateness). The actual vehicle departure time would be the summation of the holding time and the larger of scheduled vehicle departure time and the vehicle arrival time.

$$WN_i = \max\left(\sum_{j=1}^N \delta_{j,i} \cdot A_j, D_i\right) + H_i - D_i, \quad \forall \text{ route } i \quad (4-5)$$

However, because the originating passengers at each downstream stop/station may experience different waiting time due to the slack time built in the timetable, it is not reasonable to add identical extra waiting time as in equation (4-4) for each originating passenger. With the stepwise linear relationship assumed in previous sections, the departure time lateness for downstream stops would be conditional on the lateness at the transfer terminal.

$$DY_{i,k}|WN_i = \max\left(0, WN_i - \sum_{j=1}^k Slack_{i,j}\right) \quad \forall \text{ route } i, \text{ stop } k \quad (4-6)$$

Accordingly, the equivalent weight for the originating passengers at downstream stops/stations would be the ratio of the actually experienced vehicle departure lateness at downstream stops/stations to the lateness at the dispatching terminal.

$$\psi_{i,k} = (DY_{i,k}|WN_i)/WN_i \text{ if } WN_i \neq 0 \text{ otherwise } \psi_{i,k} = 1 \quad \forall \text{ route } i, \text{ stop } k \quad (4-7)$$

This weight is used to calculate the equivalent “passengers” experiencing delay in the following way. The overall equivalent originating passengers would be the total of the originating passengers at the terminal and the equivalent passengers at downstream stops/stations in proportion to the corresponding equivalent weight for each individual downstream stop/station.

$$NP_i = NTP_i + \sum_{k=1}^{S_i} NSP_{i,k} \cdot \psi_{i,k} \quad \forall \text{ route } i \quad (4-8)$$

Each arriving vehicle has to be matched with a route on which to depart, meaning:

$$\sum_{j=1}^N \delta_{i,j} = 1 \quad \forall \text{ route } i \quad \text{and} \quad (4-9)$$

$$\sum_{i=1}^N \delta_{i,j} = 1 \quad \forall \text{ route } j \quad (4-10)$$

If the transfer passengers arrive at the terminal earlier than the vehicle departure time on the route to which they are supposed to transfer, the extra waiting time to make the transfer would simply be the difference between the vehicle departure time and the passenger arrival time, otherwise, the passenger must wait until the next departing vehicle on the transfer route.

$$WT_{i,j} = \max\left(\sum_{i=1}^N \delta_{i,j} \cdot A_i, D_j\right) + H_j - A_i \quad \text{if } \max\left(\sum_{i=1}^N \delta_{i,j} \cdot A_i, D_j\right) + H_j \geq A_i \quad \forall \text{ route } i,j \quad (4-11)$$

$$WT_{i,j} = \tau_j - A_i \quad \forall \text{ route } i,j \quad \text{otherwise} \quad (4-12)$$

The algorithm of predicting τ_j will be introduced in the following section.

$$\delta_{i,j} = \{0,1\} \quad \forall \text{ route } i,j \quad \text{and} \quad H_i \geq 0 \quad \forall \text{ route } i \quad (4-13)$$

4.3.3. Solution Method

With the objective function and constraints in the previous section, this problem is formulated as a mixed integer optimization problem, which makes it fairly difficult to address the problem in an analytical way. However, other approaches are possible.

For the vehicle dispatching problem with holding, Dessouky et al. (1999) proves that the earlier-arriving vehicle should only be held until the specific time instants at which a later vehicle arrives. As described in the model formulation section, the holding time on the i^{th} route H_i and binary variables $\delta_{i,j}$ representing swapping are the only decision variables in the objective function. Once all $\delta_{i,j}$ are set, the problem reduces to a pure holding problem as in Dessouky et al. (1999). In this case, the objective function can be optimized by enumerating each predicted vehicle arrival time, so that the optimal holding time for each route can be easily obtained without much computational effort. On the other hand, for a transfer terminal with no more than ten routes synchronized for the timed transfer, the full enumeration of the possible swapping scenarios would not be computationally burdensome for a personal computer for the purpose of real-time vehicle dispatching decisions. Therefore, the problem can be solved through explicit enumeration with reasonable computational effort.

Deeper insights can be gained from a closer look at the objective function. One can see that the swapping and holding strategies can only function to change the actual vehicle departure time on each route; there is nothing they can do with the actual arrival time A_i . In addition, a general conclusion can also be drawn that the vehicle dispatching

strategy is determined by the estimated vehicle arrival times, the scheduled departure times, and the passenger loads.

Finally, it is more operationally practical for swapping to be applied as a real-time control strategy if the routes intersecting at the terminal can be divided into several sub-groups (i.e. a swapping group). That is, there may be only a subset of routes included in a swapping decision, instead of the full set of routes: the full set may create unreasonable requirements on the vehicle operators to be familiar with all the routes in the group. Accordingly, if the total number of routes at the terminal can be divided into several swapping groups, the vehicle dispatching decision within a swapping group would be independent of such decisions within other groups (including routes not in a swapping group). Also, for routes not in a swapping group at all, the only decision for vehicle dispatching is the holding time, and this holding time decision is completely independent of the holding time or swapping decision on other routes. Therefore, the decision for these independent routes will be called single route-based. In contrast, for those routes within a swapping group, the vehicle dispatching strategy decision would be group-based, which means the holding and swapping strategies will be interacting within the group.

Based on this analysis, the study design in the following sections focuses on the case with one single swapping group whose size may vary. The vehicle dispatching strategy for the case with more than one swapping group can be made for each swapping group separately, and the overall passenger cost reduction is the simple sum of that resulting from the control strategies applied over each swapping group.

Moreover, according to the model constraints, the optimal strategy will be determined by not only the operation status of the vehicles from the coordinated routes in the current decision-making cycle, but also the operation status of the next set of arriving vehicles. Basically, the prediction of the departure times for the next set of vehicles is another optimization problem, but with several assumption made for the sake of problem simplicity:

- The third set of dispatched vehicles at the upstream terminals will arrive and depart the terminal on time. This may be justified in part because it is unlikely that there is any new location information for these vehicles during the current cycle of decision-making. In other words, no holding or swapping would be expected to be applied to the third set of vehicles.
- Those transfer passengers who will miss the current connection would not add significant passenger burden to the next set of arriving vehicles. This assumption allows one to say that the two optimization processes (i.e. for the current set of vehicles and the next set of vehicles respectively) are basically independent.

4.4. Simulation and Sensitivity Analysis

4.4.1. Simulation Logic

As introduced at the beginning of this chapter, to examine the potential of applying real-time swapping into vehicle dispatching strategies at a transfer terminal, simulation is used again to compare the vehicle dispatching strategy with holding only and the strategy

with both holding and vehicle swapping. The basic tasks in this particular simulation study employ the simulation to:

- Design a wide variety of reasonable transit operating scenarios capturing the relevant transit operating factors and their distributions;
- Conclude what operating environment may be favored by the real-time swapping strategy; and,
- Conduct a sensitivity analysis to observe how swapping performs under a variety of transit operating factors.

As argued in the last two chapters, a CRN-based simulation study is preferential for comparison purposes. Therefore, Common Random Number simulations are conducted separately for the situations with and without real-time swapping for vehicle dispatching strategies at the transfer terminal. For clarity, the advantages of the CRN based simulation study are repeated here. For both sets of simulations, common random number series are applied as inputs to generate a very large amount of scenarios, which may be considered to be sufficient to cover many reasonable situations in the real world. With this assumption, the simulation process can ensure several points:

- Rare cases have quite a few of instances in the simulation, so the conclusion drawn from the simulations can be quite general; and,
- Due to identical inputs to both simulation processes, the differences observed from the simulation output can be attributed to the system control variables, which are the vehicle dispatching strategies in this particular study.

Based on this, a flow diagram (Fig. 4.2) can help explain the basic logic in the simulation process.

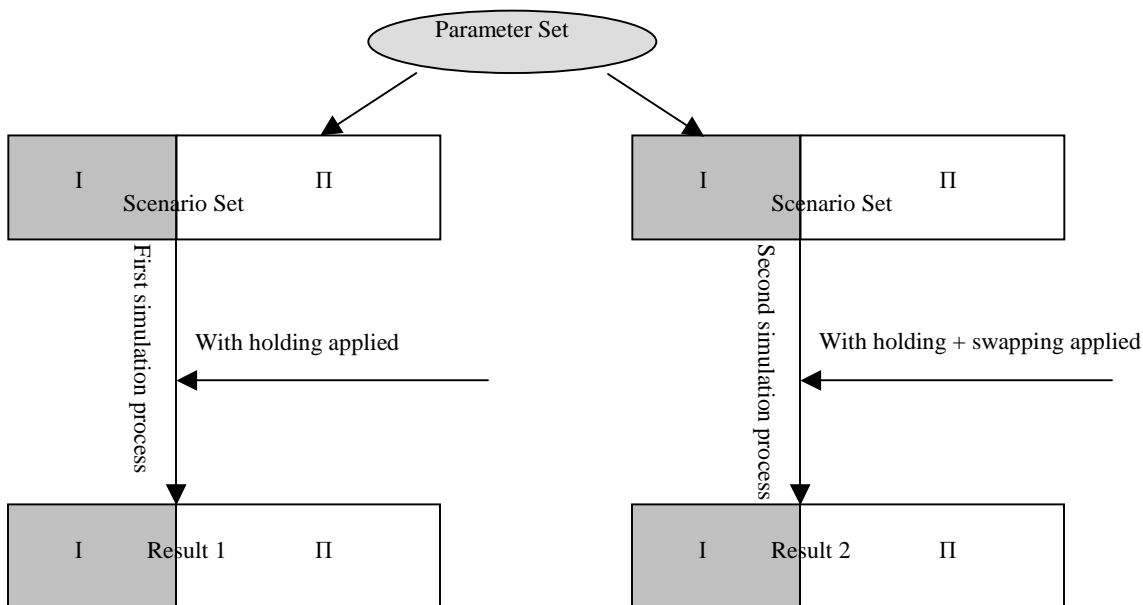


Fig. 4.2. Logic Diagram of Simulation Study for Terminal Swapping

In this diagram, there are two simulation processes, with and without swapping available. A common set of parameters is input to two simulation processes to generate two sets of identical transit operation scenarios on which different vehicle dispatching strategies are applied. In the second simulation process (i.e. with the swapping strategy available), only a portion of the input scenarios (the gray area in the scenario set) favors vehicle swapping. This portion of the scenarios is called “swapping-prone scenarios”. The swapping-prone input scenarios will lead to portion I (the gray area in the second simulation process) in the second result set (i.e. Result 2). As a contrast, another portion of the scenario set (the white area in the scenario set), the non-swapping-prone scenarios, will lead to the portion II (the white area in the second simulation process) in the second result set. Similarly, the scenario set and the result set in the first simulation process can

be also divided into two portions corresponding to those in the second simulation process. It is clear that the portion II in the first result set would be identical with the portion II in the second result set. Therefore, the comparisons between the overall results from the two simulation processes are essentially the comparisons between portion I in the holding only results and portion I in the holding and swapping results. This is the primary comparison of terminal dispatching strategies described in this chapter.

Also, the statistical difference between the two portions (portion I and portion II) in the input scenario set will help in identifying the transit operation environments in which the vehicle swapping strategy is generally favored. Moreover, the input parameter set defines the transit operation scenarios generated, and certainly the performance outcomes from the two simulation processes. A sensitivity analysis of the strategy performance to the variations of the input parameters can be helpful to further identify the favorable transit operation environments for the application of a vehicle swapping strategy.

Accordingly, the basic tasks for the simulation are to:

- Analyze the sensitivity of the strategy performance to the key transit operation factors; and,
- Identify the swapping-prone scenario set and conclude its statistical characteristics in comparison with the non-swapping-prone scenario set.

In more detail, the simulation study for comparing the vehicle dispatching strategies with and without vehicle swapping is depicted in Fig. 4.3.

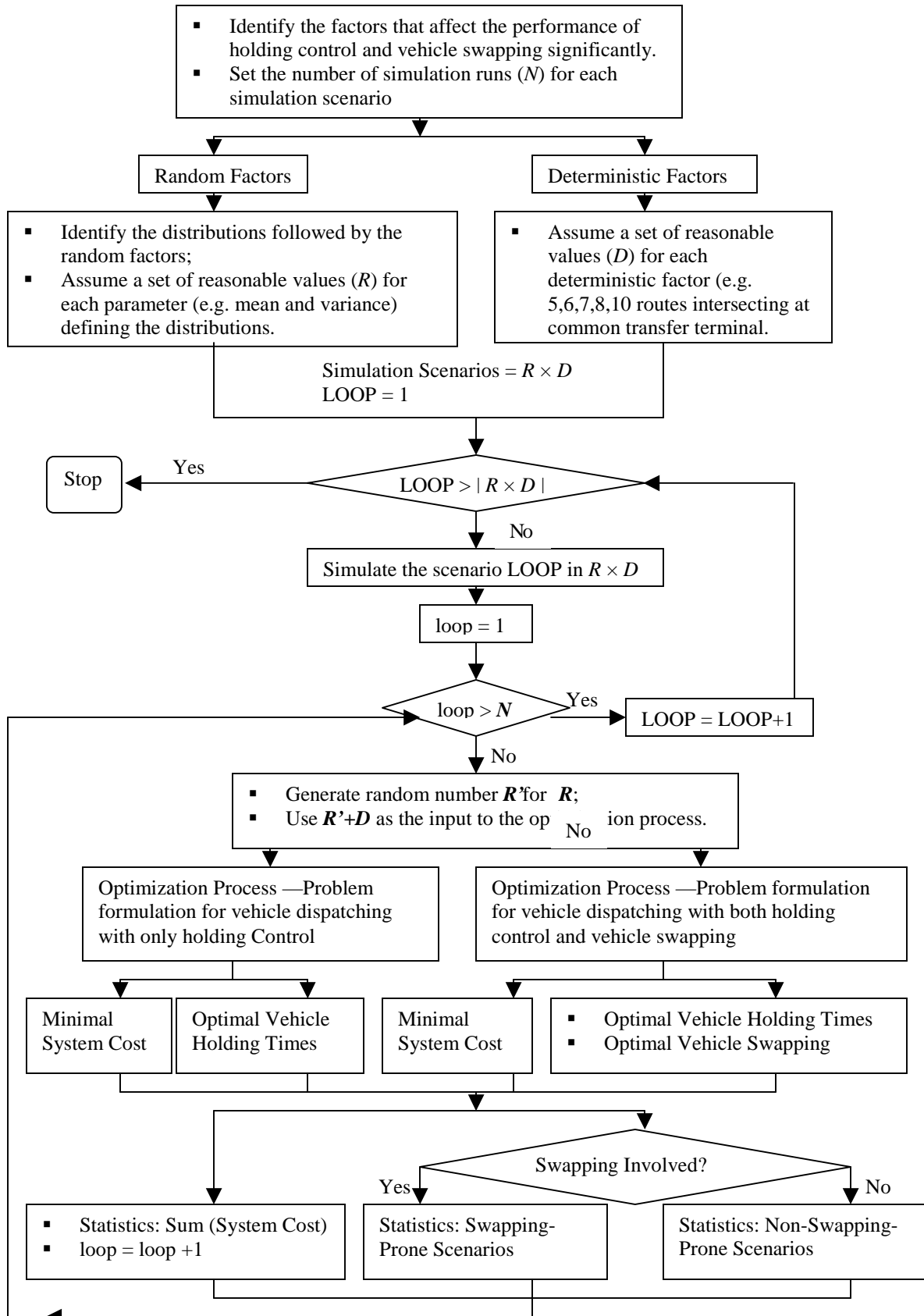


Fig 4.3. Simulation Flowchart – Vehicle Dispatching with Swapping Problem

As shown in Fig. 4.3, this simulation study for the problem of vehicle dispatching with swapping starts by identifying the key factors which may affect significantly the performance of both the holding control and the vehicle swapping strategy for the purpose of enhancing passenger transferring. The key factors (both random and deterministic) identified are further employed to construct simulation scenarios ($R \times D$). For each of these scenarios, a certain number (N) of common random factor sets are randomly generated to simulate the dynamics of both the holding control and the vehicle swapping at a transfer terminal. The problem formulations and solution method are used to solve for the optimal vehicle holding times and, under consideration of swapping, the optimal vehicle swapping combination. The simulation results (in terms of minimum system cost) are collected to compare the performance of the holding only strategy and the strategy of holding and swapping, and for further identifying the favorable conditions for the vehicle swapping strategy by comparing the statistics of the swapping-prone scenarios and the non-swapping-prone scenarios.

4.4.2. Assumptions

Several assumptions on which the simulations are based are listed below.

- 1) Passenger boarding is moderate and would not exceed the transit vehicle capacity.
- 2) A schedule-based holding strategy and any built-in slack time are applied only at downstream time points on each route. Also, the vehicle departure lateness (if any) at the dispatching terminal can be diminished at the downstream time

points only by the slack time built in the timetable; in other words, there is very little freedom for the vehicle operator to adjust vehicle speed;

- 3) Passenger boarding and alighting occur only at the terminal and downstream time points;
- 4) No transit vehicle overtaking is allowed. Actually, vehicle overtaking rarely occurs for infrequent transit service;
- 5) No significant transit service disruption occurs in the process of simulation;
- 6) In each simulation run, the number of transfer passengers between routes is accurately available for all incoming vehicles in the current decision-making cycle, but the distribution for those vehicles coming later are derived from historical data;
- 7) The overall magnitude of transfer passengers for all routes intersecting at the terminal is certain, but the distribution among routes varies from run to run in the simulation process. The process of generating transfer passengers will be introduced later in the chapter;
- 8) The terminal and downstream boarding distribution for originating passengers are derived from historical data for each route;
- 9) The number of originating passenger boardings on each particular route follows a normal distribution, which is derived from historical data. However, only the mean of the normal distribution is used in each simulation run for selecting a dispatching strategy and for strategy performance evaluation (i.e. no standard deviation is considered in the simulation). This mean value is

randomly generated following a discrete uniform distribution between 40 and 80 passengers, identical for each route. Therefore, the mean for each route can be different from other routes in each particular simulation run. Also, passenger boarding is assumed to occur with identical values at each time point on a particular route;

- 10) The set of routes in each swapping group is pre-defined;
- 11) Each vehicle is dispatched from upstream terminal right on schedule;
- 12) At any point of decision, the transit vehicle is assumed to be able to traverse the remaining distance to the terminal according to the scheduled running time; and,
- 13) The actual time spent traveling the distance which is supposed to be traveled in one AVL polling cycle follows a normal distribution, and no correlation exists between the successive travel time distributions on the same route.

In the assumptions above, (2) and (3) imply that only time points are considered in the simulation. Assumption (13) basically implies that the transit vehicle cannot catch up once it is late, and travel time stochasticity is neglected, which seems unreasonable. However, as stated in (5), no significant transit service disruption occurs in the process of simulation, meaning that no vehicle is too late at the decision-making point. In practice, the vehicle operator's ability to catch up is very limited in a short time period. Moreover, in this particular study, a vehicle's location and the predicted arrival time at the terminal is updated at a frequency of every polling cycle (40 seconds, for example). This may help capture a large portion of the stochasticity in the travel time prediction.

4.4.3 Test Scenarios and Basic Parameters

Several key concepts and parameters in the simulation and sensitivity analysis are defined below.

Number of Swappings: the number of swappings is counted as the number of routes that are served by transit vehicles from another route in the swapping group. For example, there are three vehicles, vehicle 1, vehicle 2 and vehicle 3, supposed to serve route 1, route 2 and route 3 respectively. However, the control strategy may dictate that the vehicle 1 serves route 2, vehicle 2 serves route 3, and vehicle 3 serves route 1. In this case, three swappings occur.

Group Size: the number of routes that are deliberately assigned to a swapping group.

Originating Passengers: the equivalent passengers (NP_i) originally boarding at the dispatching terminal and the downstream stops along each route.

Transfer Passengers: other passengers ($TP_{i,j}$) excluding the originating passengers above.

Swapping Threshold Cost: the incremental cost incurred by each swap (IC).

Basically, ten transit routes with common structures are assumed to be synchronized at a hypothetical transit transfer terminal. The “common structure” refers to the assumption that each route has the same number of time points, identical distance between time points, common operating schedule and all other static characteristics. Also, the common structure means that each route has same originating passenger distribution, transfer passenger distribution, and travel time distribution. Although the route structure is identical for each route, the simulated values for the originating passengers, transfer

passengers, travel times and so on in each simulation run may be different from route to route. This may lead to quite different passenger boarding profiles and vehicle lateness evolutions among routes for each simulation run.

Within this particular study design, the basic parameters applied in the simulation are given in Table 4.1. These basic parameters together construct a baseline case in the simulation study, and simulation is conducted to test the sensitivity of the swapping performance to the following parameters: number of routes, slack time at the transfer terminal, the standard deviation of the travel time distribution, the headway, the average originating and transfer passengers, and the incremental swapping cost.

Table 4.1. Transit Operation Parameters in Simulation (Terminal Swapping)

Parameters	Values
Total Number of Routes	10
Total Number of Intermediate Time Points on Each Route	6
Slack Time at Time Point (min)	0.4
Slack Time at Transfer Terminal (min)	1.0
AVL Data Polling Time Interval (sec)	40
Standard Deviation of Travel Time Distribution (average = 40 sec)	20
Average Route Headway (min)	30
Average Originating Passengers Boarding at Terminal and Downstream Stops	60
Average Transfer Passengers to/from Each Route	17
Swapping Cost (Passenger.Sec)	3600
Average One-way Trip Time per Route (Hr)	1

According to Table 4.1, for the baseline case, 10 routes in total terminate at the transfer terminal, operating at a headway of 30 minutes; each route has 6 time points, 1 hour one-way trip time, 0.4 minute built-in slack time at each time point, and 1.0 minute built-in slack time at the transfer terminal; the real vehicle travel time for traversing the distance that is supposed to be traveled in 40 sec with the given average running speed is assumed normally distributed with a standard deviation of 20 seconds; each route produces 17 transfer passengers, and attracts 60 originating passengers and 17 transfer

passengers; and, each vehicle swapping may incur an operator cost equivalent to 3600 passenger sec.

Furthermore, with the built-in slack time (0.4 minutes each in Table 4.1) at each time point for each route, along with the schedule-based holding strategy assumed in the previous section, the vehicle arrival time lateness at the transfer terminal will approximately follow a shifted lognormal distribution with an intercept of 166 sec and a lognormal distribution with mean of 246 sec and standard deviation of 116 sec, which is consistent with some previous studies, e.g. Abkowitz et al. (1987).

The average number of transfer passengers is assumed to be known as shown in Table 4.1. However, in each simulation run, the number of transfer passengers between each pair of transit routes is randomly generated by the following algorithm.

- Step 1: Randomly generate the total number of transfer passengers (capacity) that one transit route may attract or produce, using an uniform distribution ranging from 0 to 40 passengers;
- Step 2: Randomly generate a route number which would be used as the producing route;
- Step 3: Randomly generate another route number which would be used as the attracting route;
- Step 4: Randomly generate the transfer passengers between the producing route and the attracting route, using uniform distribution ranging from 0 to 20 passengers; compare this generated number with the minimum of: (1) the remaining capacity for the attracting route; (2) the remaining capacity of the

producing route; and, (3) a predefined maximum of 20 transfer passengers between any two routes; if the generated number is smaller than the minimum of the three, use the generated number as the final transfer passengers between the producing route and the attracting route. Otherwise, use the minimum instead.

Step 5: Update the remaining capacities for both the producing route and the attracting route.

Step 6: Check if all routes have been paired with the producing route generated at Step 2. If not, go to Step 3 to randomly generate another attracting route.

Step 7: Check if all routes have been used once as the producing route. If not, go to Step 2 again. If so, stop.

4.5. Simulation Results and Sensitivity Analysis

4.5.1. Sensitivity to Operating Factors

With the basic parameter set in Table 4.1, a sensitivity analysis is conducted through observing how the overall passenger cost reductions vary with each factor, while keeping others the same. The results are shown in Tables 4.2 through 4.8 below. The cells in italic font represent the values for the baseline case in each table. Herein, the performance (represented by the values in the tables) for either control strategy is essentially the average over the outputs of all synchronized routes based on 20,000 simulation runs with randomly generated inputs according to the basic parameter settings and assumptions outlined previously.

Table 4.2. Sensitivity to Swapping Group Size

Group Size (1)	Average Cost Reduction by Adding Swapping				Average Holding Time Reduction (sec) (6)	Percentage of Runs with a Swap (%) (7)	Average # of Swaps Made When a Swap Occurs (8)	Average # of Swaps per Simulation Run (9)
	Overall		Originating Passengers (Pass-Sec) (4)	Transfer Passengers (Pass-Sec) (5)				
	(Pass-Sec) (2)	(%) (3)						
2	327	3.2	128	201	2	2.3	2.0	0.05
3	959	9.4	361	599	6	5.9	2.0	0.12
4	1766	17.3	686	1081	12	10.6	2.0	0.22
5	2741	26.9	1048	1694	19	15.6	2.1	0.33
6	3800	37.3	1441	2360	27	20.7	2.2	0.44

Table 4.3. Sensitivity to Swapping Threshold Cost

Threshold Cost (Pass-Sec) (1)	Average Cost Reduction by Adding Swapping				Average Holding Time Reduction (sec) (6)	Percentage of Runs with a Swap (%) (7)	Average # of Swaps Made When a Swap Occurs (8)	Average # of Swaps per Simulation Run (9)
	Overall		Originating Passengers (Pass-Sec) (4)	Transfer Passengers (Pass-Sec) (5)				
	(Pass-Sec) (2)	(%) (3)						
0	7216	70.9	2005	5211	49	73.9	2.8	2.10
1800	3469	34.1	1294	2176	26	27.1	2.1	0.58
3600	1766	17.3	686	1081	12	10.6	2.0	0.22
5400	871	8.6	353	519	6	4.3	2.0	0.09
7200	376	3.7	148	229	2	1.5	2.0	0.03

Table 4.4. Sensitivity to Total Number of Routes

# of Routes (1)	Average Cost Reduction by Adding Swapping				Average Holding Time Reduction (sec) (6)	Percentage of Runs with a Swap (%) (7)	Average # of Swaps Made When a Swap Occurs (8)	Average # of Swaps per Simulation Run (9)
	Overall		Originating Passengers (Pass-Sec) (4)	Transfer Passengers (Pass-Sec) (5)				
	(Pass-Sec) (2)	(%) (3)						
5	3030	103.2	1035	1995	20	15	2.1	0.31
6	2652	60.3	910	1743	18	13.8	2.1	0.28
7	2342	38.7	830	1512	16	12.7	2.1	0.26
8	2082	29.3	768	1314	14	11.9	2.1	0.25
9	1833	20.9	694	1138	13	10.9	2.0	0.22
10	1766	17.3	686	1081	12	10.6	2.0	0.22

Table 4.5. Sensitivity to Travel Time Variation

Travel Time SD (Sec) (1)	Average Cost Reduction by Adding Swapping				Average Holding Time Reduction (sec) (6)	Percentage of Runs with a Swap (%) (7)	Average # of Swaps Made When a Swap Occurs (8)	Average # of Swaps per Simulation Run (9)
	Overall		Originating Passengers (Pass-Sec) (4)	Transfer Passengers (Pass-Sec) (5)				
	(Pass-Sec) (2)	(%) (3)						
5	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0
15	263	13.4	80	182	1	1.4	2.0	0.03
20	1766	17.3	686	1081	12	10.6	2.0	0.22
25	4390	30.5	1891	2498	32	25.3	2.1	0.54
30	7084	47.3	3225	3858	51	39.1	2.2	0.87

Table 4.6. Sensitivity to Originating Passenger Boardings

# of Originating Passengers (1)	Average Cost Reduction by Adding Swapping				Average Holding Time Reduction (sec) (6)	Percentage of Runs with a Swap (%) (7)	Average # of Swaps Made When a Swap Occurs (8)	Average # of Swaps per Simulation Run (9)
	Overall		Originating Passengers (Pass-Sec) (4)	Transfer Passengers (Pass-Sec) (5)				
	(Pass-Sec) (2)	(%) (3)						
60	1766	17.3	686	1081	12	10.6	2.0	0.22
54	1543	13.7	582	959	12	9.5	2.0	0.19
48	1305	10.5	475	830	11	8.2	2.0	0.17
42	1102	8.0	381	721	10	7	2.0	0.14
36	890	5.9	293	596	9	5.8	2.0	0.12
30	658	4.0	207	451	8	4.5	2.0	0.09

Table 4.7. Sensitivity to Average Route Headway

Headway (Min) (1)	Average Cost Reduction by Adding Swapping				Average Holding Time Reduction (sec) (6)	Percentage of Runs with a Swap (%) (7)	Average # of Swaps Made When a Swap Occurs (8)	Average # of Swaps per Simulation Run (9)
	Overall		Originating Passengers (Pass-Sec) (4)	Transfer Passengers (Pass-Sec) (5)				
	(Pass-Sec) (2)	(%) (3)						
10	488	--	202	285	1	4.7	2.0	0.10
20	1240	53.8	476	764	8	8.9	2.0	0.18
30	1766	17.3	686	1081	12	10.6	2.0	0.22
40	2149	10.8	806	1342	15	11.6	2.1	0.24
50	2545	8.1	901	1643	17	12.0	2.1	0.25
60	2853	6.8	975	1879	19	12.8	2.0	0.26

Table 4.8. Sensitivity to Slack Time at Terminal

Slack Time at Terminal (Sec) (1)	Average Cost Reduction by Adding Swapping				Average Holding Time Reduction (sec) (6)	Percentage of Runs with a Swap (%) (7)	Average # of Swaps Made When a Swap Occurs (8)	Average # of Swaps per Simulation Run (9)
	Overall		Originating Passengers (Pass-Sec) (4)	Transfer Passengers (Pass-Sec) (5)				
	(Pass-Sec) (2)	(%) (3)						
0	2629	11.5	1246	1383	21	17.4	2.1	0.37
30	2189	12.0	948	1241	16	13.9	2.1	0.29
60	1766	17.3	686	1081	12	10.6	2.0	0.22
90	1340	38.0	465	876	8	7.6	2.0	0.16
120	1016	16.9	337	679	7	5.9	2.0	0.12

In Tables 4.2 through 4.8, the overall percentage cost reduction by adding swapping (i.e. column 3 in the tables) is the percentage cost reduction of swapping and holding, using a baseline of only holding. That is, it is the percentage improvement from swapping over the holding-only strategy. Column 7 is the percentage of simulation runs where a swap occurs, out of the total number of simulation runs (i.e. 20,000). Column 8 is the average number of swaps, over all simulation runs when swapping occurs. Column 9 is the average number of swaps, over overall simulation runs, no matter whether swapping occurs. Column 9 is thus the product of columns 7 and 8.

The sensitivity of the average cost reduction achieved by applying real-time swapping into the vehicle dispatching strategy to each of these factors is also shown graphically in Fig. 4.4.

From Tables 4.2 through 4.8 and Fig. 4.4 above, there are several important tendencies that can be observed. First, when a favorable transit operation environment exists, significant passenger cost reduction can be achieved by applying real-time

swapping into the public transit control strategy at the terminal. Herein, the significance of the real-time swapping can be represented by the overall reduction of the passenger cost and the percentage of simulation runs with swapping applied. For example, for the baseline case across all tables, the average overall passenger cost reduction can be increased by more than 17 percent (0.5 passenger-hours) by using vehicle swapping; and, when operating conditions turn more favorable (e.g. lower swapping threshold cost in Table 4.3, smaller number of routes in Table 4.4, more variable arrival time at the terminal in Table 4.5, and so on), this percentage and/or absolute value will increase considerably. For example, when the total number of routes drop to five while keeping the swapping group the same, the absolute improvement from swapping increases to an average of about 50 passenger-minutes per simulation run, which essentially doubles the overall passenger cost reduction achieved by holding only. Also, when the swapping threshold cost drops to zero, a majority of simulation runs use swapping (73.9%).

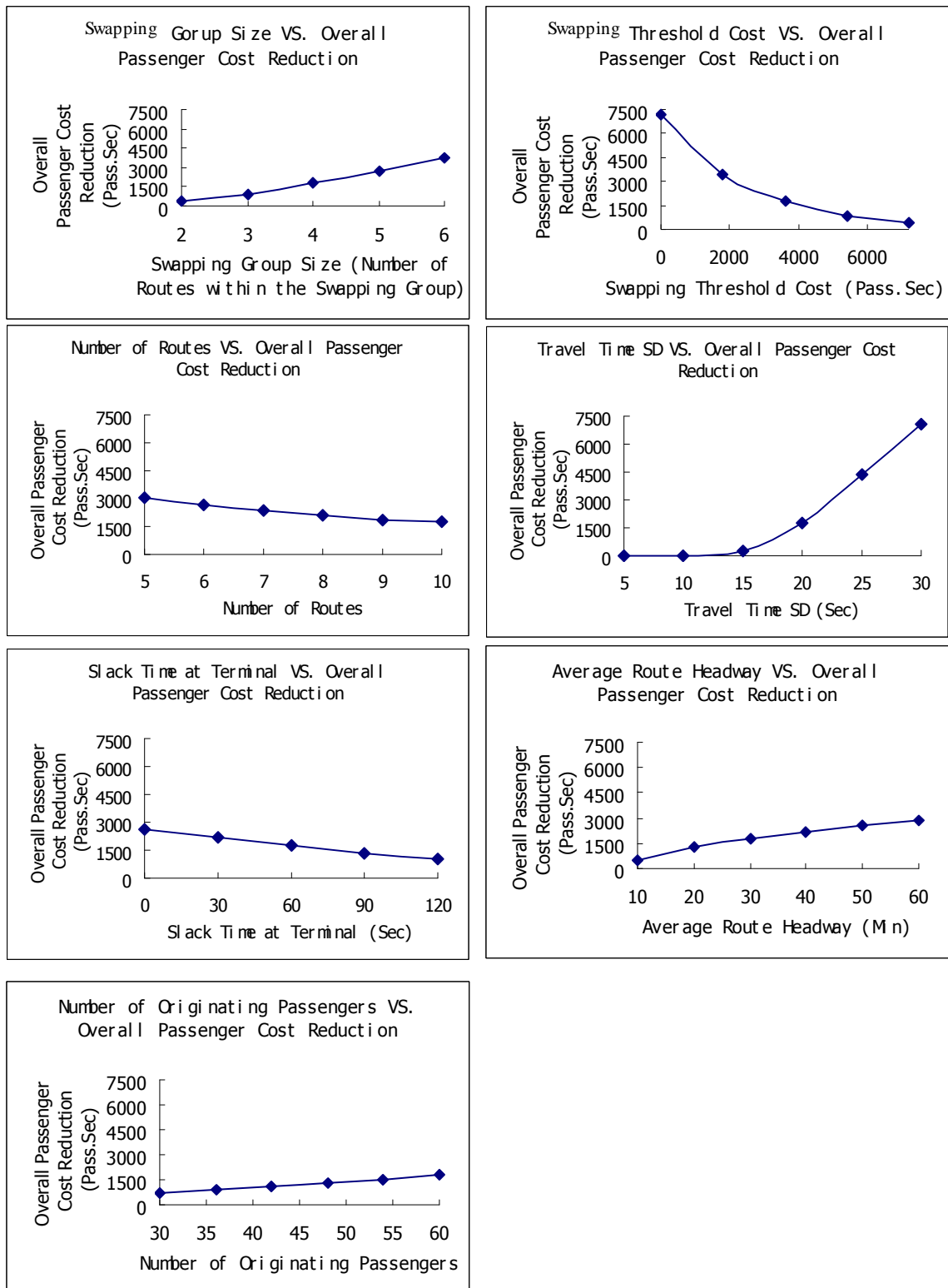


Fig. 4.4. Sensitivity Analysis for Terminal Swapping

Both the overall passenger cost reduction and the percentage of runs swapped increase as the swapping group size (Table 4.2), the variation of travel time (Table 4.5), the number of originating passengers (Table 4.6) and the average route headway (Table 4.7) increase, but decrease as the swapping threshold cost (Table 4.3), the total number of routes (Table 4.4) and the slack time at the transfer terminal (Table 4.8) increase. However, the relative improvement of the combined strategy over the holding-only strategy does not necessary have this tendency, or varies at a different rate. When the average route headway decreases, the overall passenger cost reduction and the benefits achieved by swapping drop quickly, but the percentage improvement with respect to the passenger cost reduction from the holding-only strategy increases. This implies that the holding strategy may be most effective for longer headways, but that swapping may have considerably greater advantage with shorter headways.

The swapping threshold cost may be the most sensitive factor influencing the performance of real-time swapping. As shown in Table 4.3, in each case where the equivalent swapping threshold cost was increased by 1800 passenger-seconds, the average passenger cost reduction was cut in half. As the swapping threshold cost drops to zero, the average passenger cost reduction could be as high as about two passenger-hours, which is about 71 percent higher than of the overall passenger cost reduction incurred by the holding strategy alone. Conversely, at one passenger-hour, the cost reduction drops to under 0.5 hours (17.3%), and to almost nothing at two passenger-hours.

When the travel time variation is relatively low, as for the first two cases in Table 4.5 with 5 sec and 10 sec standard deviations respectively, there is no significant benefit

achieved by applying real-time swapping. As a matter of fact, for these particular situations, even holding has no benefit. For the case of a 5 sec standard deviation, the average overall passenger cost reduction from holding strategy drops to zero and is only 270 passenger-seconds (less than 1.5% of the base overall passenger cost) for the case of a 10 sec standard deviation.

In Table 4.8, the absolute value of the overall passenger cost reduction decreases as the slack time at the terminal decreases. However, the tendency of the improvement of the combined strategy over the holding-only strategy is not as clear, in part because the precision of the factors in the simulation process may raise a large variation on the outcomes when only very few cases are observed with holding and/or swapping applied. This is the case when the slack time at the terminal increases up to more than one and a half minutes. In addition, the AVL data polling rate, every 40 seconds in this particular study, also adds to the difficulty of estimating the vehicle departure times accurately.

In comparing columns 4 and 5 in all tables, swapping tends to reduce both the originating passenger cost and the transfer passenger cost consistently in all cases. Also, the cost reductions on both originating passengers and transfer passengers increase/decrease with those sensitivity factors. This further demonstrates what was argued previously about how originating passengers and transfer passengers function in the control strategy: originating passengers also play an active role to achieve a final tradeoff with the transfer passengers. Also, moderate transit vehicle holding time reductions, ranging from 0 to 51 seconds (Table 4.5), are achieved as swapping is applied,

and this reduction of holding time is certainly contributing a portion of the overall passenger cost reduction.

It appears that integrating real-time swapping into the transit terminal control strategy will achieve greater cost reduction for transfer passengers than for originating passengers, even in the case of relatively frequent service, e.g. for the 10 minute average headway in Table 4.7. This is primarily because only the cases of infrequent service (average headway larger than 10 minutes) are simulated and tested in the sensitivity analysis. For the cases of infrequent service, missing the connection means a large waiting cost will be incurred by the transfer passengers. In this study, the ratio of transfer passengers to originating passengers for the baseline case is about 0.3, so one can expect that the transfer passenger cost reduction may dominate the originating passenger cost in a majority of cases when applying real-time swapping. As a matter of fact, as the number of originating passengers on each route decreases to 30 (transfer/originating ratio is about 0.6) in Table 4.6, the transfer passenger cost reduction more than doubles that for originating passengers.

The average number of swaps for those simulation runs with at least one swap in the control strategy mostly ranges between 2.0 and 2.1, except for a special case in which no cost is incurred by the swapping itself (this value is 2.8 for the case of zero swapping threshold cost in Table 4.3). According to the definition of swapping, two swaps are essentially the minimum if swapping has been applied. Therefore, this narrow range of the number of swaps may demonstrate that the potential for multiple swaps (as in the example for defining swapping in Sec. 4.4.3) are limited. It is unlikely that cases of more

than two swaps contribute significantly to the overall passenger cost reduction, even for the case of relatively large swapping group size: this value is only 2.2 for the case of 6 routes in a swapping group in Table 4.2.

Aside from the observations above, control strategies applied over one swapping group are completely independent of those on other swapping groups. As a result, the passenger cost reduction from the control strategies applied for one swapping group is also entirely independent of others. From Table 4.2, it can be easily observed that the overall passenger cost reduction for a larger swapping group grows faster than linearly. This is because a relatively larger swapping group can always offer more opportunities to make swapping pairs or chains than with several smaller groups. This conclusion is also valid with regard to the components of originating passenger cost and transfer passenger cost reductions. However, a larger swapping group brings much more rigorous requirements for knowledge of routes among the vehicle operators, which is certainly a challenge to the swapping strategy for a public transit agency.

4.5.2 Sensitivity to Uncertainty of Passenger Boardings

As an assumption, the total number of originating passengers over each route in the simulation follows a normal distribution established from historical data. However, only the mean of this distribution is applied for selecting the optimum strategy and generating the results in Tables 4.2 through 4.8. As a result, the real value of the number of originating passengers may deviate from the mean of the distribution applied for selecting the optimum strategy, depending on the variance of the distribution. Consequently, one may wonder whether the conclusions based on the results from Tables 4.2 through 4.8

still hold, if the real number of originating passengers over routes are used to re-estimate the real passenger cost reductions based on the optimum strategies derived from only the mean values of those distributions. Therefore, it is particularly helpful to conduct another simulation study for observing how the uncertainty of the originating passengers, in terms of the standard deviation of the passenger distribution, affects the passenger cost reduction in Tables 4.2 through 4.8. This essentially gives the value of information on originating passengers to the swapping strategy.

In this simulation process, the values of the originating passenger boardings are randomly generated from the normal distribution with the standard deviation shown in Table 4.9, and the mean as used in generating results in Tables 4.2 through 4.8. These randomly generated values are called “real originating passenger boardings”. Then these real passenger boardings are applied to re-compute the passenger cost reductions based on the optimum strategy decided using the mean values only. The simulation results are shown in Table 4.9.

Table 4.9. Originating Passenger Cost Reduction vs. Originating Passenger Variation

SD of Originating Passenger Distribution	‘Real’ Originating Passenger Cost Reduction (Pass.Sec)
0	686
5	675
10	674
15	674
20	671
25	671
30	670

From Table 4.9, it can be easily observed that the ‘real’ originating passenger cost reduction does not vary significantly, in comparison with the baseline case of no

uncertainty with the estimation of the number of originating passengers. Therefore, one can conclude that the uncertainties in the estimate of the originating passenger boardings on each route are not very important for comparing the relative performance of the swapping-integrated control strategy with the holding-only control strategy. In other words, the uncertainty in estimating the number of originating passengers would not affect the conclusions drawn previously.

4.5.3. Characteristics of Swapping-Prone Scenarios

It was argued earlier that it would be helpful to understand why vehicle swapping occurs by comparing the statistics of the two input scenario groups (the gray and white areas in the scenario set in Fig. 4.2). As the basic sensitivity factors in Tables 4.2 through 4.8 are fixed, each input scenario differs from others with regard to three factors: the vehicle arrival times, the originating passenger boardings, and the number of transfer passengers. These three factors and their combination in each specific scenario will affect the use and performance of the control strategies with or without swapping. More specifically, these three factors can be represented by three measures that are defined below.

- *Originating passenger boardings on swapping routes (routes in the swapping group):* In the objective function, we can see that the originating passenger boarding on the transit routes only within the swapping group will determine the optimal control strategy.
- *Transfer passengers attracted by swapping routes:* It is difficult to find one individual measure which can represent completely the overall effects of the

passenger transfer over the swapping decision. However, the transfer passengers attracted by swapping routes may have the largest influence on the swapping decision.

- *Arrival time lateness distribution over the swapping routes:* Essentially, any swapping decision would be determined by the arrival time distribution over all routes. However, only the arrival times on the swapped routes would contribute to the originating passenger cost reduction (the first term in the objective function).

The comparisons of the statistics of the two scenario groups, the swapping-prone group and the non-swapping-prone group, in terms of the three measures, are shown in Table 4.10, which is drawn from the simulation results for the baseline case.

Table 4.10. Statistical Comparison of Swapping-Prone/Non-Swapping-Prone Scenario Groups in the Baseline Case

	Mean When a Swap Occurs	Mean When No Swap Occurs	SD When a Swap Occurs	SD When No Swap Occurs
# of Originating Passengers on Each Route Within Swapping Group	60	59	10	10
# of Transfer Passenger Attracted by Each Route Within Swapping Group	18	17	9	9
Arrival Time Lateness for each Route within Swapping Group (Sec)	127	79	111	60

As seen in Table 4.10, the average number of originating passengers on each route within the swapping group when swapping occurs is only 1 passenger larger than when no swapping occurs, and this is the same for the transfer passengers attracted by each route. Moreover, the standard deviations of both originating and transfer passenger distributions among the routes within the swapping group do not show any difference between the two cases. However, for the arrival time lateness distribution among those routes, substantial differences exist. When swapping occurs, the average vehicle arrival

lateness on each route is about 48 seconds (about 60%) more than when no swapping occurs, and the standard deviation of arrival lateness is 51 seconds (about 85%). Although not shown explicitly here, similar tendencies can be observed across all other cases (beyond this baseline case) from the simulation results. Accordingly, a simple conclusion may be derived:

It may not be necessary that the originating passengers and transfer passenger distribution over the transit routes within the swapping group have a higher average and variance to justify the implementation of real-time swapping. However, a swapping application is preferred when there are more delayed vehicle arrivals and higher variation in the vehicle arrival time distribution. Furthermore, the combination of the three factors, which are represented by the three measures, justifies the real-time swapping strategy.

4.6. Conclusions and Future Study

This chapter describes the study of the problem of vehicle dispatching with swapping, which is the first to investigate the potential of real-time swapping as a vehicle dispatching strategy at a major transfer terminal at which multiple transit routes intersect.

The problem of vehicle dispatching with swapping is formulated as a mixed integer programming problem. An exhaustive search method is employed as the problem solution method to take the advantage of the small problem scale for a limited number of transit routes terminating at a common transfer terminal and the special feature of the vehicle holding times. Again, a simulation study is conducted to compare the holding and swapping strategy and the holding only strategy.

The simulation outcomes show that substantial benefits, in terms of both originating passenger and transfer passenger cost reductions, can be achieved by applying real-time swapping among a properly designed swapping group of moderate size. Further, the sensitivity analysis suggests that the combined strategy with real-time swapping and holding consistently outperforms the holding-only strategy in terms of reducing both originating and transfer passenger cost and the vehicle holding time.

Decreasing the swapping threshold cost is the most effective way to improve the performance of this combined control strategy. As the swapping threshold cost is reduced to zero, a majority of situations (73 percent in Table 4.3) include real-time swapping into the vehicle control strategy, and a substantial passenger cost reduction can be achieved on average. Reduction of this threshold cost reduction may be achieved through careful design of transit routes, advanced technologies which can assist the operator navigating on an unfamiliar route, and well-trained vehicle operators under a rigorous training program.

The passenger cost reduction from real-time swapping appears to be sensitive to all factors examined in this particular study, but to different extents. Generally, real-time swapping favors larger swapping group size, a smaller total number of routes, more variable travel time, more originating passengers, larger average route headway, less slack time at the terminal, and a lower swapping threshold cost. Based on this, the swapping may be considered at a terminal under the following conditions:

- A relatively small number of appropriately utilized routes (the number of originating passengers is expected to be more than 50) terminate at the terminal;

- The average route headway is less than 40 minutes;
- Vehicles often arrive late at the terminal; and,
- It is hard to embed slack time into the service timetables.

In addition, any efforts for reducing the swapping threshold cost and enlarging the swapping group size would be helpful for applying real-time swapping, and a swapping cost of no more than one passenger-hour is strongly suggested to ensure that the overall passenger cost reduction would not be less than 10 percent.

Chained swapping among more than two routes does not show a noticeable advantage over the paired swapping, except for the case of zero swapping threshold cost (Table 4.3). Also, a large swapping group can help achieve substantially more benefits over several small groups, but this also brings greater challenge to the swapping group design.

No evidence suggests that any significant difference with regard to both the originating passenger and transfer passenger distributions exists among the routes within the swapping group for those cases with or without swapping. Higher delay and greater variance in the arrival time distribution over the routes within the swapping group seem to show the greatest advantage for swapping. Ultimately, the combination of the originating passenger, the transfer passenger and the vehicle arrival time would really justify the implementation of real-time swapping.

Finally, the hypothetical study design applied here can offer some evidence of the potential of integrating real-time swapping into the vehicle dispatching strategies. Also, some guidelines about how, where, and when to apply real-time swapping can also be

drawn from this study. However, there are several ways the study of this particular vehicle dispatching problem can be enhanced in the future. The assumption of normally distributed travel times over a short time period may have to be validated and calibrated by real-world evidence, though this is also assumed in the holding control and stop-skipping control problem studies in previous chapters. Proper travel time forecasting models may also be employed at each time of updating the control strategy. A real example can definitely be added to help make the analysis more convincing.

Moreover, the future study can focus on the swapping group design under certain route operating environments, under resource and policy constraints. Specifically, the studies on transit route design, driver training program, and vehicle technology deployment would be particularly helpful for eventually implementing real-time vehicle swapping into real practice, without violating the crew working schedule, the vehicle maintenance schedule as well as many other factors that may have significant meaning to the transit agencies.

CHAPTER 5 AVL DATA ANALYSIS

As introduced in the previous chapters, the transit vehicle travel time forecasting is the key for operations control strategy optimization. The studies on the holding problem in Chapter 2 and stop-skipping problem in Chapter 3 again assume that the vehicle downstream trajectories can be precisely predicted. This chapter develops the methodology to predict the vehicle downstream trajectories using AVL data specifically collected from Level *A* or Level *B* AVL system. Level *A* and Level *B* AVL systems will be introduced in Section 5.1.

5.1. Introduction

Transit vehicle travel time can be determined by a large number of factors, e.g. length of the segment traversed, number of stops and number of intersections on the segment, passenger demand, traffic condition, vehicle maximum speed, vehicle acceleration/deceleration ability, seasonality, time of day, direction of travel, to name but a few. Many previous studies have investigated the behavioral mechanisms and dynamics underlying the transit travel time forecasting. Abkowitz and Engelstein (1983) developed empirical regression models of transit mean travel time and travel time deviation based on data collected in Cincinnati, Ohio. The study results showed that the mean travel time is strongly influenced by trip distance, passenger boarding and alighting, and also the number of signalized intersections. Also, the travel time deviation at an early stop can propagate as the vehicle proceeds further downstream. Alfa, Menzies, Purcha and McPherson (1988) developed both linear and nonlinear regression models for estimating bus travel times, and found that the bus travel time is mostly determined by the number

of stops, number of stop signs, number of traffic lights and length of the segment. Later, Abdelfattah and Khan (1998) also developed both linear and nonlinear vehicle delay regression models for the existing street network, as well as for the cases with one lane blocked. They concluded that the link length, number of bus stops, and the M/T ratio (moving time/ travel time) are the most significant independent variables influencing bus delays; and, the vehicle traffic density variables have also strong effects on bus delays. More recently, Strathman et al. (2001) are among the first who investigated the composite effects of the bus operator behavior and other common factors on transit vehicle running time by using AVL data. Their study suggested that the operator effects on running time appear to be normally distributed and account for 17 percent of running time variation.

However, collecting data for revealing the underlying mechanism of vehicle trajectory evolution can be very laborious, and in a majority of occasions seems monetarily prohibitive for transit agencies. As a result, many previous studies focused on only the relatively simple and straightforward relationships between the impacting factors and phenomena that can be easily observed. Abkowitz and Engelstein (1982) focused on the temporal and spatial dimensions of travel time in transit system, and found that average transit travel time during the afternoon peak is roughly 10 percent higher than during the morning peak, and 25 percent higher than during the evening off-peak. Higher travel time variation is correlated with higher mean travel time on links, and the standard deviation of travel time is greater in the afternoon peak than in the morning peak. The coefficient of variation is lowest during the morning peak and highest during the off-peak.

Finally, the predictability of vehicle arrival time and passenger waiting time definitely deteriorate as the vehicle moves farther away from the route origin. Later, Levinson (1983) reported a detailed analysis of transit speeds, delays, and dwell time based on surveys conducted in a cross section of U.S. cities. The investigated relationships and parameters can provide inputs for planning service changes and assessing their impacts. Seneviratne and Loo (1986) analyzed the vehicle travel time data from two transit routes in Halifax, Nova Scotia, Canada, and found that fundamental to a realistic analysis of an bus route is proper segmentation (to divide the overall route into several separate segments), which can improve both Poisson and negative binomial model estimates of the number of stops for passenger boarding and alighting for an on-call transit service. Also, this study showed that the passenger demand is a very suitable and logical surrogate for many factors that would otherwise be required to explain the increase in bus travel time. This observation is consistent with what is implied by the recursive vehicle trajectory evolution employed in previous chapters.

In parallel with the advent of AVL systems in European countries and in the U.S., new methods to predict the vehicle travel time become more attractive for researchers, with an assumption that the overwhelming amount of historical and real-time AVL data may have captured the composite effects of many underlying factors on the transit vehicle operation. Accordingly, historical vehicle location data can be used to interpret the vehicle's future location, with or without the assistance from other auxiliary data. Kalaputapu and Demetsky (1995) investigated schedule behavior modeling concepts to bus transit systems using Artificial Neural Networks (ANNs) with AVL data, and the

preliminary results showed that the schedule deviation of a bus at a timepoint k is influenced by the schedule deviations at timepoints $k-1$ and $k-2$. Also the schedule behavior methodology using ANNs appeared to be suitable to predict future vehicle trajectories. Lin and Zeng (1999) developed a set of algorithms for estimating the transit vehicle's arrival time at stops using AVL data considering the scheduled arrival time, delay correlation, and waiting time at timepoints. Also, the performance of the algorithms with different levels of information is compared in terms of overall precision, robustness and stability. This study suggested that a reasonable algorithm can only be developed based on the identified delay pattern that reflects the bus operation characteristics of a specific site. Wall and Dailey (1999) presented an algorithm for predicting the arrival time of transit vehicles by using a combination of both real-time and historical AVL data. In the proposed algorithm, a Kalman filter framework is employed to track the vehicle's real-time position, and the statistical estimation is used for predicting the vehicle arrival time at the downstream terminal. Also, the study results showed that the proposed algorithm is sufficiently flexible to many adverse conditions, typically where data are sparse. More recently, Chien and Ding (1999) developed both link-based and stop-based ANNs for predicting bus arrival time in real time. Also, the accuracy of the methodology is assessed through simulating a real-life transit route and conducting an analysis of the predicted bus arrival time. The study results finally showed that the link-based ANN performs well at the stops with few intersections in between, and the stop-based ANN is suitable to work for the converse case.

However, the previous studies on vehicle travel time prediction obviously lack the consideration of the interactions between the previous and the following vehicles, which is the most significant aspect that the operation control strategy intends to control. Therefore, this chapter also presents a vehicle travel time prediction study, assuming AVL data are available. This study differs from the previous studies in that it focuses on using AVL data to estimate the vehicle operating parameters used in the recursive vehicle trajectory evolution functions introduced in previous chapters, e.g. equations (2-13) and (2-14). In equations (2-13) and (2-14), the interactions between the previous and the current vehicle are explicitly modeled. Furthermore, the link travel time $r_{i,j-1}$ and the passenger boarding rate λ_j are the two parameters which cannot be directly obtained from some AVL systems (e.g. Level A and B). Another major contribution of this study is that it proposes the methodologies to derive the vehicle operating parameters using the AVL data collected specifically from Level A and Level B AVL systems. In the report, *Uses of Archived AVL-APC Data to Improve Transit Performance and Management: Review and Potential* (TCRP, 2003), Level A and Level B AVL systems are described as:

Detail level A represents the least detail: infrequent event-independent location records. This level is representative of many older AVL systems, in which information is captured only about the location of the bus when it is polled. The polling interval, depending on system design, is typically 40 to 120 s, though intervals as low as 16 s and as great as 240 s have been implemented. For special purposes, the polling rate can be increased on some buses at the expense of others. Determining the moment at which the bus passed a particular location (e.g., a timepoint) requires interpolation, causing approximation errors that can be as large as half the polling cycle. This level of detail is the simplest to implement in an AVL system because it requires no onboard vehicle tracking intelligence.

Detail level B includes timepoint records. An onboard computer knows when it reaches the timepoint location, and either records it in an onboard computer, or transmits it over the air as a timepoint message. Knowing the onboard location is best done by the onboard computer tracking vehicle location, either using GPS or dead reckoning. With this tracking capability, users can select and change locations of interest. Most new AVL systems with GPS have this capability. King County Metro has implemented this capability in a signpost system by having the central computer tell the bus, about two polls before arriving at a timepoint, what odometer reading will indicate arrival at the coming timepoint. A less flexible way to obtain level B data is to install wayside transmitters at locations of interest which will trigger a record being made on a passing bus. However, this arrangement makes it hard to change specified locations, and the failure of a transmitter means loss of information at that point. (TCRP, 2003, p. 16)

As described above, Level A and Level B AVL system does not directly collect the vehicle running speeds ($\tau_{j-1,j}$) and passenger boarding rates ($r_{i,j-1}$). Furthermore, due to the relatively low frequency for data collecting, Level A and Level B AVL system does not provide sufficiently fine data resolution for directly estimating the vehicle operating parameters. As a result, specific algorithm needs to be developed for deriving the vehicle operating parameters from AVL data, specifically the vehicle running speeds ($\tau_{j-1,j}$) and passenger boarding rates ($r_{i,j-1}$). This is also the focus of the AVL data analysis in this dissertation.

Furthermore, $r_{i,j-1}$ can be equivalently translated into the vehicle running speed $\tau_{j-1,j}$, since the distance between the current vehicle location and the downstream stop or between two stops can be easily estimated with the AVL data and the transit system inventory data.

As argued in Chapter 1, for an infrequent transit service, the transit passengers consult with the schedule and the vehicle dwell time is thus relatively fixed during certain

time periods, e.g. off-peak hours, since a vehicle headway change may not result in the change of the number of passengers boarding the vehicle. However, for a frequent service, this is not true. The passengers missing a vehicle may not have to wait for long time to board the next vehicle. Therefore, for frequent transit service, the passengers arriving to stops are generally assumed random. Accordingly, a vehicle with a larger headway is usually subject to a larger number of passenger boardings, and thus larger dwell time at the stop. Mathematically, if vehicle's current location is at A, for the case depicted in Fig. 5.1, the vehicle arrival time t_B at B would be approximated as:

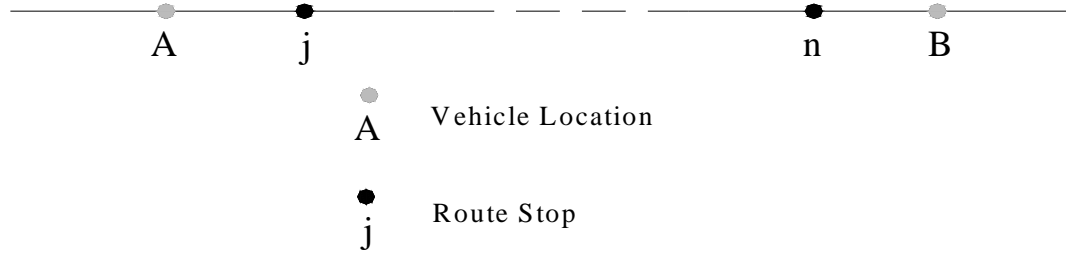


Fig. 5.1. Example of vehicle locations

$$t_B = t_A + \frac{D_{A,j}}{\tau_{j-1,j}} + \sum_{l=j}^{n-1} \frac{D_{l,l+1}}{\tau_{l,l+1}} + \sum_{l=j}^n dw_l + \frac{D_{n,B}}{\tau_{n,n+1}} \quad (5-1)$$

wherein,

t_A : Vehicle arrival time at point A;

t_B : Vehicle arrival time at point B;

dw_l : Vehicle dwell time at stop l ;

$\tau_{l,l+1}$: Vehicle running speed on the link between stops l and $l+1$;

$D_{A,j}$: Distance between location A and stop j ; and,

$D_{n,B}$: Distance between stop n and location B .

Equation (5-1) indicates that the vehicle travel time consists of the vehicle running time (terms 2, 3 and 5 at the right hand side) and the vehicle dwell time (term 4 at the right hand side). Vehicle dwell time consists of both passenger alighting time and boarding time. If passenger boarding and alighting occur simultaneously, the dwell time would be the larger of passenger alighting time and boarding time; otherwise, it would be the total of both. However, for either sequential or simultaneous passenger boarding/alighting process, it is extremely difficult to derive the passenger alighting effect merely from AVL data. For the sake of practicability and simplicity, as assumed in Chapter 2 for the holding control problem, it is assumed again that the vehicle includes the passenger boarding time only. The constant part (α) of the dwell time is also neglected, and its effect is included in $\tau_{l-1,l}$. Furthermore, as argued previously, for frequent service, the number of passengers boarding a particular vehicle is determined by the vehicle headway, provided the passenger boarding rate λ_l is given. Mathematically, for a random passenger boarding process (e.g. the Poisson process assumed in previous chapters), the average vehicle dwell time at stop l is:

$$dw_l = \beta_1 \cdot h_l \cdot \lambda_l \quad (5-2)$$

By adding i as the index for the vehicle, equation (5-1) becomes:

$$t_{i,B} = t_{i,A} + \frac{D_{A,j}}{\tau_{i,j-1,j}} + \sum_{l=j}^{n-1} \frac{D_{l,l+1}}{\tau_{i,l,l+1}} + \sum_{l=j}^n \left(\beta_1 \cdot h_{i,l} \cdot \lambda_l \right) + \frac{D_{n,B}}{\tau_{i,n,n+1}} \quad (5-3)$$

In equation (5-3), $D_{A,j}$, $D_{n,B}$, $t_{i,B}$ and $t_{i,A}$ either are directly offered by AVL or can be precisely calculated with AVL data and the transit system inventory data. Though

$h_{i,l}$ cannot be directly observed or calculated, it may be estimated from an average vehicle running speed V , e.g. 20 mph, as by equations (5-6) and (5-7). As a matter of fact, the inaccuracy in the headway $h_{i,l}$ estimation due to the assumed vehicle running speed may not matter seriously. For example, if the location A is the one closest to stop j among all observed vehicle i 's locations downstream of stop j and upstream of stop $j+1$, by using $t_{i,A}$ and the assumed vehicle running speed, the estimated departure time of vehicle i at stop j seldom deviates from the actual departure time much, especially when the data polling is relatively frequent. The difference between the estimated headway and the actual headway seldom exceeds 30 sec by doing so, and this may not affect the passenger boarding time estimation seriously if the passenger boarding rate is not extremely large.

Other than the five variables ($D_{A,j}$, $D_{n,B}$, $t_{i,B}$, $t_{i,A}$ and $h_{i,l}$), the only unknown factors left in equation (5-3) are λ_l and $\tau_{i,j-1,j}$ ($\tau_{i,l,l+1}$), namely the passenger boarding rate at each stop and vehicle running speed on each segment between two adjacent stops along the transit route. This chapter will be spent to investigate the methods for estimating these two unknown factors using AVL data collected by a Level A or B AVL system such as that operated in Tucson, Arizona.

In the following part of this chapter, the AVL data used for analysis and the procedure to process the raw AVL data will be briefly introduced. Furthermore, three assumptions are made on the vehicle running speed for facilitating deriving vehicle operating parameters from AVL data as follows:

Assumption 1: A vehicle's running speed is trip-specific. The vehicle running speed on each separate route segment for any particular trip is similar from day to day, but the speed may differ from trip to trip within the same day;

Assumption 2: A vehicle's running speed is day-specific. In this case, the assumption is made that the current vehicle's running speed resembles the previous vehicle's on the same route segment; and,

Assumption 3: The combination of assumption 1 and assumption 2, considering both trip-specific and day-specific effects.

With these three assumptions, linear regression models based on the recursive vehicle trajectory function are employed to derive vehicle running speeds and the passenger boarding rates. Also, statistics are presented to evaluate the regression results and to suggest the rationality of the assumptions made on the vehicle running speed. Furthermore, a comparison study is conducted to compare the performance of the reasonable vehicle running speed assumptions in terms of their prediction error, using an independent AVL data set.

5.2. AVL Data

AVL Data Source

The AVL data under investigation in this chapter is collected from eastbound route 8, which is currently operated by Sun Tran (Tucson, Arizona) primarily on Broadway and 6th Avenue as in Fig. 5.2. The Sun Tran AVL system basically polls vehicle location data to the dispatch every 40 sec for all operating vehicles in fleet. For either Level A or Level B AVL systems, they do not directly collect the passenger boarding data and vehicle

running speeds. These parameters generally need to be estimated using interpolation, regression, or some other approaches.

The AVL data collected for analysis spans from May 4, 2004 through August 5, 2004. The AVL data of three trips, starting from Laos transit center at 6:55 AM, 7:05 AM and 7:15 AM respectively, are collected, but only the trip starting at 7:15 AM will be analyzed. However, the AVL data of the other two trips are indispensable in the sense that they will be used to derive the vehicle headways ($h_{i,t}$, one of the decisive variables of the vehicle travel time) for the 7:15 AM trip and to account for the impacts of the previous trips (6:55 AM and 7:05 AM) on the following trip (7:15 AM).

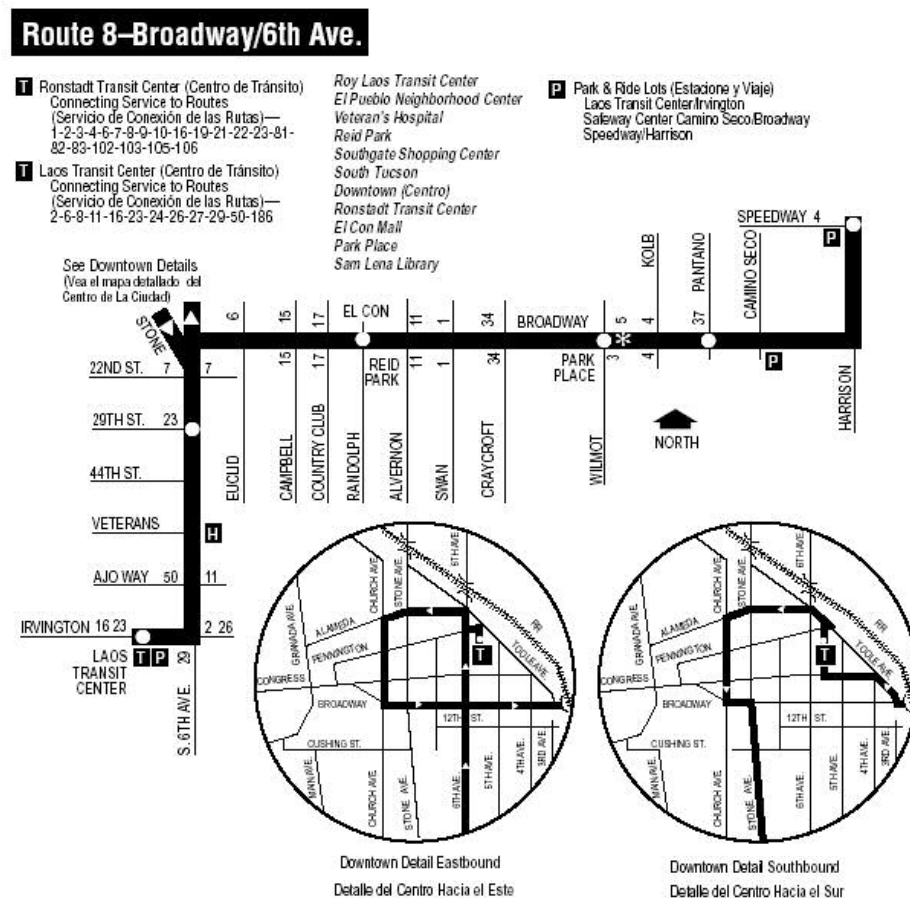


Fig. 5.2. Route 8 of Sun Tran in Tucson, AZ

Data Processing

The original AVL data in Sun Tran's AVL database are processed in the following steps before they are analyzed.

Step 1:

Retrieve the vehicle location data of three consecutive eastbound trips starting from Laos transit center at 6:55 AM, 7:05 AM and 7:15 AM respectively.

Step 2:

For each vehicle location, identify the stop that the vehicle just passed before it recorded its location, and estimate the distance between this location and the identified stop.

Algorithm: Find the perpendicular point of each vehicle location on the straight line defined by each pair of adjacent stops. If the perpendicular point falls between two adjacent stops, pick the upstream one as the stop just passed by the vehicle.

Step 3:

Screen out the outliers. In this study, outlier is defined as the vehicle location more than 50 meters away from its perpendicular point on the route segment. 50 meters is assumed as the upper limit of the system error of Sun Tran AVL, which might be a bit conservative in the sense that maximum error of the Sun Tran AVL may be far less than assumed here.

Step 4:

Truncate the data set. There are 78 stops along route 8 in total, with downtown Ronstadt Center as an intermediate stop. Downtown Ronstadt Center is special in the sense that a large amount of slack time and frequent operations controls may be applied at this particular stop. To avoid the impacts of the operations control on the AVL data analysis, only the vehicle locations and stops downstream of Ronstadt Center are used to form the data set for analysis. Furthermore, over-congested traffic condition and curved street segments in the downtown area could make the data analysis appreciably more complicated. Therefore, the data set is further truncated to comprise only the location data downstream of the intersection of Euclid and Broadway. In addition, for eastbound route 8, half of the trips terminate at the Wilmot and Broadway terminal, and another half operates beyond Wilmot along Broadway and terminates at Speedway/Harrison Park 'N' Ride terminal. On the route segment between Euclid and Wilmot, within the peak hours, the vehicle headway is 10 minutes on average according to the schedule, and 20 minutes on the segment between the Wilmot terminal and the Speedway/Harrison Park 'N' Ride terminal. This data analysis focuses on only frequent transit service (average operation headway is equal to or less than 10 minutes). Therefore, the vehicle location data from the Wilmot terminal to the Speedway/Harrison Park 'N' Ride terminal is removed from the data set. Moreover, the number of observed vehicle locations (samples) between Craycroft and Wilmot along Broadway is surprisingly small. Therefore, only the vehicle locations falling on the segment between Euclid and Craycroft along Broadway is kept in the data set.

Step 5:

Estimate the vehicle headway at each stop. The vehicle headway can be estimated with the following algorithm, assuming average vehicle running speed as 20 mph.

With Fig. 5.1, given that point B is the one closest to stop n among all observed locations of vehicle i falling between stops n and $n+1$, the departure time $d_{i,n}$ of vehicle i at stop n will be:

$$d_{i,n} = t_{i,B} - \frac{D_{n,B}}{V} \quad (5-4)$$

' V ' in (5-4) is the assumed average vehicle running speed, it is assumed as 20 mph in this study.

If there is no observed location between stops n and $n+1$, simply give $d_{i,n}$ a negative value. For example,

$$d_{i,n} = -1 \quad (5-5)$$

Do exactly the same thing for vehicle $i-1$, and the headway of vehicle i at stop n can be calculated with equation (5-6) below, given both $d_{i,n}$ and $d_{i-1,n}$ are positive.

$$h_{i,k} = d_{i,n} - d_{i-1,n} \quad (5-6)$$

If either $d_{i,n}$ or $d_{i-1,n}$ is negative, the vehicle headway at stop $n-1$ is used to represent the vehicle headway at stop n .

$$h_{i,n} = h_{i,n-1} \quad (5-7)$$

By going through these five steps, the AVL data related to 7:05AM trip and 7:15AM trip may have the form as in Table 5.1.

Table 5.1. AVL Data Format after Preliminary Processing

Time	Stop	Distance (m)	Headway (sec)
20040504075147	1	182	567
20040504075149	1	227	567
20040504075222	2	262	662
20040504075540	6	143	708
20040504075752	8	808	694
20040504080012	12	311	771
...

In Table 5.1, each row represents an observed vehicle location with the first column standing for the time that the location is recorded by AVL, the second column representing the index of the stop just passed, the third column being the distance between the location and the stop in the second column headed by ‘*Stop*’, and the fourth column representing the vehicle headway at the stop just being passed. Furthermore, in Table 5.1, records are sorted in the order in which the vehicle locations have been observed.

The data form in Table 5.1 can also be converted into any specific form for estimating the passenger boarding rates and vehicle running speeds, and developing the travel time forecasting models, under different circumstances.

5.3. Regression Methodology and Parameter Estimation

Similar to the assumptions made on the vehicle running speed in Section 5.1, the passenger boarding rate can also be assumed either trip specific or not within certain time period, say peak hours. Both assumptions on the passenger boarding rate have been adopted in this particular study. Specifically, for vehicle running speed assumptions 1 and 2, passenger boarding rates are assumed the same for all trips of interest, and the trip-specific boarding rates are assumed for the vehicle running speed assumption 3. The

remaining part of this chapter presents the models for deriving the vehicle operating parameters under three vehicle running speed assumptions separately. Before proceeding to investigate vehicle running speed assumption 3, the study on the first two vehicle running speed assumptions is necessary.

5.3.1. Assumption 1: Trip-Specific Vehicle Operating Behavior

Trip-specific vehicle operation assumes the vehicles of the different trips (e.g. 7:05 AM trip and 7:15 trip) behave differently. It also assumes that the vehicles on the same trip (e.g. 7:15 AM trip) on different days behave the same operationally, so that they have common operational parameters, e.g. vehicle running speeds and passenger boarding rates. Accordingly, the operational parameters derived from the trip (e.g. 7:15 trip) on previous days may best represent the parameters of the same trip (e.g. 7:15 trip) of the current day, and they can be used directly to predict the vehicle travel time for the same trip in the future. The derivation of the vehicle running speeds and the passenger boarding rates for a particular trip (e.g. 7:15 AM trip) can rely on the method of linear regression.

To estimate a regression, equation (5-3) is further generalized into the following form:

$$T_{i,A,B} = t_{i,B} - t_{i,A} = \sum_{j=2}^N \frac{p_{i,j-1,j} \cdot D_{j-1,j}}{\tau_{i,j-1,j}} + \sum_{j=1}^N (\delta_{i,j} \cdot h_{i,j} \cdot \beta_1 \cdot \lambda_j) \quad (5-8)$$

wherein,

$T_{i,A,B}$: The travel time from location A to location B for vehicle i ;

$\delta_{i,j}$: Binary variable representing whether or not stop j is located between locations A and B . If yes, $\delta_{i,j} = 1$; otherwise $\delta_{i,j} = 0$;

$p_{i,j-1,j}$: The proportion of the link (between stops $j-1$ and j) falling between locations A and B ; and,

N : total number of stops of concern.

In equation (5-8), the dependent variable is the vehicle travel time ($T_{i,A,B}$), and the independent variables are $p_{i,j-1,j} \cdot D_{j-1,j}$ and $\delta_{i,j} \cdot h_{i,j}$. The coefficients, namely the reciprocal of the vehicle running speeds ($\tau_{i,j-1,j}$) and passenger boarding rates (λ_j), will be calibrated through regression. Accordingly, the data form in Table 5.1 will be converted into the form as in Table 5.2.

Table 5.2. AVL Data Form for Regression (Assumption 1)

Link 1-2	Link 2-3 ...	Link (N-1)-(N)	Link (N)-(N+1)	Stop 1	Stop 2	...	Stop N	Time (Sec)
45	0	0	0	0	0	0	0	2
183	262	0	0	567	0	0	0	33
...

The ‘link’ columns accommodate the $p_{i,j-1,j} \cdot D_{j-1,j}$, and the ‘stop’ columns include the $\delta_{i,j} \cdot h_{i,j}$ respectively. Each record in Table 5.2 is made up of the information from two successively observed vehicle locations. The two records in Table 5.2 are essentially derived directly from the first three records (three vehicle locations) in Table 5.1. For example, the first record in Table 5.1 represents a location on the link between stops 1 and 2, and so does the second record. Therefore, the distance between the first location and second location is $227-182 = 45$ meters in Table 5.2. Since only the link between stops 1 and 2 is involved, the other cells are all zeros in the first row in Table 5.2.

The third record in Table 5.1 represents the third observed vehicle location, and it is on the link between stops 2 and 3. Therefore, ‘262’ (distance from stop 2) appears in this cell in Table 5.2. The bus location 2 is 227 meters away from stop 1, and equivalently 183 meters away from stop 2 (the length of the link between stops 1 and 2 is 410 meters), so 183 is the distance from bus location 2 to stop 2 as recorded in Table 5.2. Stop 2 is between the second and the third locations. Therefore, the vehicle headway (567) at stop 2 appears in the column ‘stop 2’ in Table 5.2.

Furthermore, since the Sun Tran AVL data is polled by the dispatch every 40 sec, the vehicle travel time (including running time and dwell time) between any two successively observed vehicle locations is almost exactly a multiple of 40 sec (40, 80, 120, ... seconds). This may be a problem for the regression method. Therefore, a data transformation is made by dividing each cell in Table 5.2 with the summation of all ‘link’ cells of the same row, which is essentially the distance between the two locations. This transformation can help to make the dependent variable continuous. For example, each cell within record 2 in Table 5.2 will divide $183+262 = 445$, and then the record 2 in Table 5.2 becomes the only record in Table 5.3:

Table 5.3. Transformed AVL Data Form for Regression

Link 1-2	Link 2-3 ...	Link (N-1)-(N)	Link (N)-(N+1)	Stop 1	Stop 2	...	Stop N	Time/Dist (Sec)
...
0.411236	0.588764	0	0	1.274157	0	0	0	0.074157
...

With the data in the form of Table 5.3 and the variable definitions mentioned previously, 1669 records in total for the trip starting from Laos transit center at 7:15 AM

are generated to calibrate 41 unknown parameters (21 vehicle running speeds and 20 passenger boarding rates). The linear regression was run in Minitab 14.0 (Minitab, 2004) and the regression results are presented in Tables 5.4 and 5.5. In these two tables, the regression results with and without speed constraints (35 mph is assumed as the maximum speed) are listed separately.

Table 5.4. Regression Results – Vehicle Running Speed (Assumption 1, 7:15 AM Trip)

Segment	Without Speed Constraints			With Speed Constraints		
	Speed (mph)	t-statistic	P value	Speed (mph)	t-statistic	P value
Euclid - Fremont	17	5.3	0.00	17	5.4	0.00
Fremont - Highland	30	2.8	0.01	30	2.7	0.01
Highland - Cherry	48	1.5	0.14	35 ¹	-	-
Cherry - Campbell	23	4.5	0.00	23	4.5	0.00
Campbell - Plumer	28	4.8	0.00	28	4.8	0.00
Plumer - Tucson	23	4.6	0.00	23	4.5	0.00
Tucson - Treat	45	2.0	0.04	35 ¹	-	-
Treat-Country Club	27	4.7	0.00	28	4.7	0.00
C Club-Randolph	33	7.3	0.00	33	7.3	0.00
Randolph-Dodge	34	2.8	0.01	34	2.8	0.01
Dodge-Alvernon	14	14.4	0.00	14	14.4	0.00
Alvernon-Irving	17	8.1	0.00	17	8.1	0.00
Irving-Columbus	27	5.7	0.00	27	5.7	0.00
Columbus-Belvedere	43	3.0	0.00	35 ¹	-	-
Belvedere-Swan	42	2.6	0.01	35	2.5	0.01
Swan-Swan(Inter)	3	7.7	0.00	3	7.7	0.00
Swan - Niven	64	1.8	0.08	35 ¹	-	-
Niven – Rosemont	14	5.5	0.00	14	5.4	0.00
Rosemont-Williams	55	2.3	0.02	35 ¹	-	-
Williams-Craycroft	14	9.4	0.00	14	9.3	0.00
Craycroft-Leonora	27	2.9	0.00	27	2.9	0.00

Note¹: Since the speed derived directly from the original data is larger than 35 mph, it is forced to be 35 mph to make the regression to be the one with speed constraints.

Table 5.5. Regression Results – Passenger Boarding Rate (Assumption 1, 7:15 AM Trip)

Stop	Without Speed Constraints			With Speed Constraints		
	λ_j (Pass/Min) ¹	t-statistic	P value	λ_j (Pass/Min) ¹	t-statistic	P value
Fremont	- ²	- ²	- ²	- ²	- ²	- ²
Highland	0.9	5.4	0.00	0.8	5.6	0.00
Cherry	0.5	2.0	0.04	0.4	2.0	0.04
Campbell	0.9	7.8	0.00	0.9	7.8	0.00
Plumer	0.6	3.9	0.00	0.6	4.0	0.00
Tucson	0.8	5.2	0.00	0.7	5.5	0.00
Treat	0.4	3.3	0.00	0.3	3.7	0.00
Country Club	0.8	6.9	0.00	0.9	6.9	0.00
Randolph	0.5	3.2	0.00	0.5	3.2	0.00
Dodge	- ²	- ²	- ²	- ²	- ²	- ²
Alvernon	1.2	22.7	0.00	1.2	22.7	0.00
Irving	- ²	- ²	- ²	- ²	- ²	- ²
Columbus	1.2	13.4	0.00	1.2	15.8	0.00
Belvedere	0.6	4.2	0.00	0.6	4.4	0.00
Swan West	1.0	5.4	0.00	1.0	5.4	0.00
Swan East	0.4	2.2	0.03	0.3	1.8	0.07
Niven	0.5	2.0	0.04	0.3	1.5	0.14
Rosemont	1.0	7.5	0.00	1.0	7.5	0.00
Williams	0.5	2.9	0.00	0.4	2.5	0.01
Craycroft	0.9	12.0	0.00	0.9	12.1	0.00

Note¹: The passenger boarding rate is estimated assuming the average passenger boarding time is 3 sec.

Note²: “-” means the value is not statistically significant.

Though, statistically, both Table 5.4 and Table 5.5 show that a majority of the independent variables do affect the vehicle travel time significantly, a validation exercise is useful. Assuming the operation schedule of route 8 is appropriate to ensure that most trips of route 8 can operate on schedule at all time points, the vehicle travel time from the timepoint Broadway at Randolph to another timepoint Wilmot at Broadway (NE) should be 14 minutes. According to the regression results, i.e. vehicle running speeds and passenger boarding rates, the travel time between the timepoint Broadway at Randolph and the stop Broadway at Craycroft would be forecasted as 11.2 minutes (7.7 minutes for running time and 3.5 minutes for dwell time), with the assumption that the vehicle headway is 10 minutes at all stops along the route as scheduled. The average travel time

for the remaining distance up to the next timepoint Wilmot at Broadway (NE) is approximately 2.75 minutes according to the AVL data. In total, the travel time between the two timepoints is 13.95 minutes. This number is very close to the scheduled 14 minutes. Furthermore, the regression results above also imply an average operating speed of 13.6 mph on the segment under investigation, which also appears reasonable.

Though the regression results may give relatively accurate vehicle travel time estimation, a few link speeds from the regression without speed constraints are surprisingly high (e.g. 64 mph between Swan and Niven along Broadway). It is also suspicious that there are three (Fremont, Dodge and Irving) stops having no passenger demand at all. A further investigation of the methodology and appropriateness of the relevant sampling context may be necessary and helpful.

5.3.2. Simulation Analysis of Sampling

Although it has been statistically shown that the vehicle running speeds and the passenger boarding rates do together provide reasonable estimates of the vehicle travel time, there is no guarantee that the derived vehicle running speeds and the passenger boarding rates from regression would be the same as the actual ones. On the contrary, a simulation example can specify the ‘real’ parameters beforehand, simulate the vehicle operations, and then compare the ‘real’ parameters with the derived parameters to see any discrepancy. In this sub-section, a simulation example is deliberately designed to mimic the simplified vehicle operations of route 8 and the Sun Tran AVL data collection process to examine the appropriateness of assumption 1.

The AVL data polling is actually a data sampling process. The data sampling process is often subject to both/either sample rate and/or sampling method problem(s). Both sample rate and sampling method are determined by many factors. The purpose of the AVL data analysis in this study is to derive the vehicle operating parameters, specifically vehicle running speeds and passenger boarding rates, from which to develop vehicle travel time forecasting models. This is also the purpose of the data collection process simulated in the example. Furthermore, the simulation example also intends to examine how the appropriateness of the data sample rate and sampling method can be affected by the characteristics of the data.

In the simulation example, a hypothetical route segment is designed. This hypothetical route segment has seven stops in total, and the relevant average vehicle running speeds as well as the passenger boarding rates are given in Table 5.6.

Table 5.6. Average Vehicle Running Speeds and Passenger Boardings (10 min vehicle headway)

Stop	Vehicle Running Speed (mph)	Average # of Pass Boardings
1	30	---
2		5
3	25	8
4	22	4
5	28	6
6	30	4
7	34	---

It is further assumed that the actual vehicle running speed is normally distributed with a COV of 0.15; passenger arrivals are a Poisson process; passenger alighting is negligible; average passenger boarding time is 3 sec; and, the distance between any two adjacent stops is constant (400, 500 and 800 meters). The combinations of different between-stop distances and sample sizes (the number of simulated trips) constitute different scenarios. The vehicle locations are sampled for different scenarios, with the same sampling time interval (40 sec) as employed by Sun Tran AVL. Furthermore, the data polling time interval of 20 sec is also employed for comparison purpose.

The hypothetically sampled vehicle location data goes through the same procedure as the actual AVL data, and similar regression results are given in Tables 5.7 and 5.8. From these two tables, one may see how the difference between the given parameter values and the derived values varies with the sample rate (data polling time interval), average between-stop distance and the sample size.

Table 5.7. Regression Results – Passenger Boardings (Simulation Example)

Stop	Derived Average Number of Passenger Boardings							
	Data Polling Time Interval (DPTI) = 40 sec							DPTI=20sec
	Between-stop distance = 800m		500		400			400
	50 simulated trips	100	50	100	50	100	1000	100
2	6	6	6	5	3	3	4	5
3	9	8	8	8	12	10	9	8
4	4	4	4	3	13	10	8	4
5	6	6	7	7	14	13	12	6
6	2	3	3	3	13	13	12	4

Table 5.8. Regression Results – Vehicle Running Speed (Simulation Example)

Link	Derived Vehicle Running Speed (mph)							
	Data Polling Time Interval (DPTI) = 40 sec							DPTI=20sec
	Between-stop distance = 800m		500		400			400
	50 simulated trips	100	50	100	50	100	1000	100
1-2	32	31	29	29	-151	21	31	29
2-3	26	25	25	25	21	23	23	24
3-4	23	22	22	22	45	30	25	23
4-5	29	28	27	27	62	47	40	28
5-6	29	30	34	33	72	62	56	30
6-7	33	32	29	29	56	68	75	34

According to Tables 5.6 through 5.8, one may see that, in comparison with the number of simulated trips, the average between-stop distance has a significant impact on the regression results. For the case with the average between-stop distance of 800 meters, both the vehicle running speed and the number of passenger boardings derived from regression are very close to the ‘real’ parameter values in Table 5.6. For the case with the between-stop distance of 500 meters, the discrepancy between the derived and the ‘real’ parameter values is still acceptable; however, when the average between-stop distance drops to 400 meters, which is similar to the Sun Tran case (370 meters on average), the

estimated parameter values are much different than the ‘real’ parameter values, even when a sufficient number of samples (1000 trips) have been collected. Yet with the average between-stop distance of 400 meters, when the data is sampled more frequently (20 sec), the estimated and the ‘real’ parameter values are much closer.

The reasons underlying these observations may be very simple. With an interval of 40 sec and the average between-stop distance being 800 meters, the data polling is sufficiently frequent to use the sampled data to directly estimate the vehicle running speeds. For such case, more than 50 percent of the sampled vehicle locations have one or more vehicle locations falling on the common route link for any particular trip. This implies that, in Table 5.3, more than 50 percent of the records have only one non-zero cell and it resides in the ‘*link*’ column. This forces the regression to directly estimate the average vehicle running speeds without much interference from another parameter (passenger boarding rate), and this results in a relatively accurate vehicle running speed estimation. Accurate vehicle running speed estimation further helps the regression to derive the passenger boarding rates precisely.

For the case with the average between-stop distance of 500 meters, the 40 sec data polling time interval is still acceptable. There are still more than 30 percent of the sampled vehicle locations having one or more other vehicle locations falling on the same route link for each particular trip. However, this percentage sharply decreases to less than 20 percent when the between-stop distance drops to 400 meters. For this case, the derivation of vehicle running speeds depends on the estimation of the passenger boarding rates, and vice versa.

The AVL data polling process is basically a time sampling process, and the dependent variable in the regression depends on the time sampling cycle. Due to this fact, the general linear regression method may not be perfectly appropriate, even though the estimated parameter values fits the data well. For instance, for the case of average between-stop distance of 400 meters and data polling time interval of 40 sec, the derived parameter values are appreciably different from the ‘real’ ones in Tables 5.6. Nonetheless, the estimated average travel time from stop 1 to stop 6 using the derived parameter values is not much different from the travel time estimated by using the ‘real’ parameter values. This is shown in Table 5.9.

Table 5.9. Difference between the Estimated Travel Time Using Derived Parameters and Using Actual Parameters

	Total Travel Time from Stop 1 to 6 (Sec)							
	Data Polling Time Interval (DPTI) = 40 sec							DPTI=20sec
	Between-stop distance = 800m		500		400			400
	50 simulated trips	100	50	100	50	100	1000	100
Using Derived Parameters ¹	472	472	325	325	277	277	277	277
Using Actual Parameters ²	463	473	332	327	265	306	290	277
Percentage Difference $(\frac{2-1}{1})$	-1.8	0.3	2.2	0.6	-4.2	10.6	4.8	0.3

In Table 5.9, for the worst case (column 7, data polling time interval 40 sec and average between stop distance 400 meters), the percentage prediction error of the total travel time from stop 1 to stop 7 is no more than 11 percent; for all others, the prediction error is negligible. However, as indicted in Tables 5.7 and 5.8, the derived vehicle running speeds and the number of passenger boarding are appreciably different from the

‘real’ parameters for the cases represented by columns 5 through 7 in Tables 5.7 through 5.9.

Furthermore, with the average between-stop distance of 400 meters, as the vehicle location data is sampled more frequently, e.g. every 20 sec (DPTI = 20 sec in Tables 5.7 and 5.8), the percentage of the sampled vehicle locations having other sampled vehicle location(s) on the common route link goes back to as high as 58 percent, and this explains why the derived parameter values from this scenario are fairly close to the ‘real’ parameter values in Table 5.6 again.

With this simple simulation example, one may conclude that, given the particular data sampling method employed by Sun Tran AVL, the data polling rate may need to be compatible with many factors, among which the average between-stop distance is the most significant one. Sun Tran AVL data are polled every 40 sec, and the average between-stop distance of the route segment under investigation is close to 400 meters. These together make up the situation that might not give precise parameter estimation, as suggested by the simulation example. Furthermore, the real vehicle operation can be considerably more complicated than simulated in the example. Accordingly, the regression results presented in Tables 5.4 and 5.5 may need further examination before being used in practice.

5.3.3. Assumption 2: Day-Specific Vehicle Operating Behavior

The day-specific assumption assumes that the current vehicle’s operating behavior is similar to the previous vehicle’s, so that the previous vehicle’s operating behavior can be incorporated to forecast the current vehicle’s operating behavior. In specific, the

previous vehicle's running speed may reflect the traffic condition that the following vehicle will also experience, and thus can be a good surrogate to the following vehicle's running speed. Based on this assumption, a simple hypothesis can be made that the current vehicle and the previous vehicle share the common running speed $\tau_{i,j-1,j}$.

There are two methods to examine this hypothesis. One is to let:

$$\tau_{i,j-1,j} = \psi_{j-1,j} \cdot \tau_{i-1,j-1,j} \quad (5-9)$$

and plug equation (5-9) into equation (5-3). The regression will be employed to give statistical evidence to either accept or reject the hypothesis that $\psi_{j-1,j}$ is 1.0. However, the derivation of $\tau_{i,j-1,j}$ (this will be introduced a bit later in this sub-section) will eventually make equation (5-3) polynomial in nature, and the regression becomes particularly difficult. Other methods need to be sought.

Another approach is to let $\tau_{i-1,j-1,j} = \tau_{i,j-1,j}$, then to examine statistically whether the derived passenger boarding rates are reasonable to be used to determine the vehicle travel time. If passenger boarding rates are reasonable, the hypothesis $\tau_{i-1,j-1,j} = \tau_{i,j-1,j}$ may not be rejected; otherwise, the hypothesis $\tau_{i-1,j-1,j} = \tau_{i,j-1,j}$ would be rejected. The logic of this method is as follows. A majority of the previous studies on operations control and the observations from many others on vehicle travel time forecasting have all shown that the vehicle travel times are determined in part by the passenger boarding rates and the vehicle headways. Therefore, if it is found that assumption 2 leads to unreasonable passenger boarding rates, $\tau_{i-1,j-1,j} = \tau_{i,j-1,j}$ must not be true, and assumption 2 should be rejected.

With this approach, equation (5-3) becomes:

$$T_{i,A,B} = t_{i,B} - t_{i,A} = \frac{D_{A,j}}{\tau_{i-1,j-1,j}} + \sum_{l=j}^{n-1} \frac{D_{l,l+1}}{\tau_{i-1,l,l+1}} + \sum_{l=j}^n (h_{i,l} \cdot \beta_1 \cdot \lambda_l) + \frac{D_{n,B}}{\tau_{i-1,n,n+1}} \quad (5-10)$$

In equation (5-10), $\tau_{i-1,j-1,j}(\tau_{i-1,l,l+1})$ is assumed known (it can be derived from the observation over vehicle $i-1$). Therefore, λ_l is the only unknown coefficient to be calibrated from the regression.

Though $\tau_{i-1,l,l+1}$ is assumed known, since the AVL data collection is a time sampling process, the vehicle running speed can seldom be estimated directly from the AVL data. Approximation has to be employed, and some derivation work needs to be done under the assumption that vehicles $i-1$ and i share the same running speed on the common route segment.

In all the three situations indicated in Fig. 5.3, it is assumed that the vehicle i ($i-1$) runs at speed $\tau_{i-1,j-1,j}$ (the running speed of vehicle $i-1$) between locations C (A) and D (B).

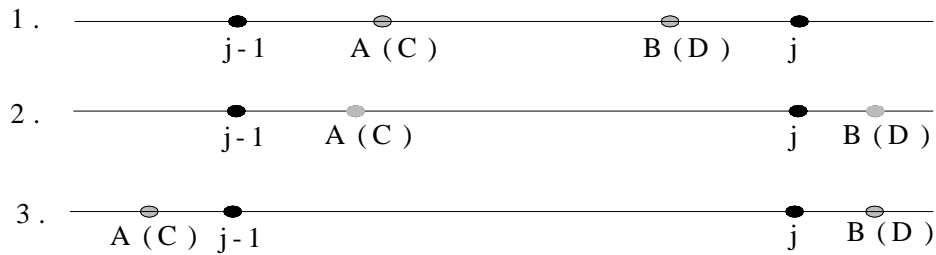


Fig. 5.3. Three Typical Situations for Running Speed Derivation

There are nine cases (9 combinations of situations) in which both the current vehicle i and the previous vehicle $i-1$ have two successively observed locations falling within the three situations depicted in Fig. 5.3. These cases lead to different specification

of the regression function, i.e. different relationship of the dependent variable (travel time between C and D) and independent variables (relevant vehicle headways), as follows:

Case 1: Vehicle $i-1$ is in situation 1 with two locations A and B , and vehicle i is also in situation 1 with two locations C and D .

For this case, $\tau_{i-1,l,l+1}$ can be directly calculated from the distance and the travel time between locations A and B for vehicle $i-1$. However, the only unknown factor λ_l is not involved in either vehicle $i-1$'s travel time between locations A and B or vehicle i 's travel time between locations C and D . Therefore, this case will not be transformed into a data record for the regression.

Case 2: Vehicle $i-1$ is in situation 1 and vehicle i is in situation 2.

The vehicle $i-1$'s speed on link between stop $j-1$ and stop j can be directly observed.

$$\tau_{i-1,j-1,j} = \frac{D_{A,B}}{T_{i-1,A,B}} \quad (5-11)$$

$$T_{i,C,D} = \frac{D_{C,D}}{\tau_{i-1,j-1,j}} + \beta_1 \cdot h_{i,j} \cdot \lambda_j \quad (5-12)$$

Equations (5-11) and (5-12) together yield equation (5-13).

$$T_{i,C,D} - \frac{D_{C,D}}{D_{A,B}} \cdot T_{i-1,A,B} = \beta_1 \cdot h_{i,j} \cdot \lambda_j \quad (5-13)$$

Case 3: Vehicle $i-1$ is in situation 1 and vehicle i is in situation 3.

Similarly, equation (5-14) can be specified.

$$T_{i,C,D} - \frac{D_{C,D}}{D_{A,B}} \cdot T_{i-1,A,B} = \beta_1 \cdot h_{i,j-1} \cdot \lambda_{j-1} + \beta_1 \cdot h_{i,j-1} \cdot \lambda_{j-1} \quad (5-14)$$

Case 4: Vehicle $i-1$ is in situation 2 and vehicle i is in situation 1.

The speed of vehicle $i-1$ on link $(j-1)-j$ can be specified using equation (5-15).

$$T_{i-1,A,B} = \frac{D_{A,B}}{\tau_{i-1,j-1,j}} + \beta_1 \cdot h_{i-1,j} \cdot \lambda_j \quad (5-15)$$

and, the equation (5-16) can be specified for regression.

$$\frac{D_{C,D}}{D_{A,B}} \cdot T_{i-1,A,B} - T_{i,C,D} = \frac{D_{C,D}}{D_{A,B}} \cdot \beta_1 \cdot h_{i-1,j} \cdot \lambda_j \quad (5-16)$$

Similarly, for other cases, equations (5-17) through (5-21) hold.

Case 5: Vehicle $i-1$ is in situation 2 and vehicle i is in situation 2.

$$\frac{D_{C,D}}{D_{A,B}} \cdot T_{i-1,A,B} - T_{i,C,D} = \left(\frac{D_{C,D}}{D_{A,B}} \cdot h_{i-1,j} - h_{i,j} \right) \cdot \beta_1 \cdot \lambda_j \quad (5-17)$$

Case 6: Vehicle $i-1$ is in situation 2 and vehicle i is in situation 3.

$$\frac{D_{C,D}}{D_{A,B}} \cdot T_{i-1,A,B} - T_{i,C,D} = \left(\frac{D_{C,D}}{D_{A,B}} \cdot h_{i-1,j} - h_{i,j} \right) \cdot \beta_1 \cdot \lambda_j - h_{i,j-1} \cdot \beta_1 \cdot \lambda_{j-1} \quad (5-18)$$

Case 7: Vehicle $i-1$ is in situation 3 and vehicle i is in situation 1.

$$\frac{D_{C,D}}{D_{A,B}} \cdot T_{i-1,A,B} - T_{i,C,D} = \frac{D_{C,D}}{D_{A,B}} \cdot \beta_1 \cdot h_{i-1,j-1} \cdot \lambda_{j-1} + \frac{D_{C,D}}{D_{A,B}} \cdot \beta_1 \cdot h_{i-1,j} \cdot \lambda_j \quad (5-19)$$

Case 8: Vehicle $i-1$ is in situation 3 and vehicle i is in situation 2.

$$\frac{D_{C,D}}{D_{A,B}} \cdot T_{i-1,A,B} - T_{i,C,D} = \frac{D_{C,D}}{D_{A,B}} \cdot \beta_1 \cdot h_{i-1,j-1} \cdot \lambda_{j-1} + \left(\frac{D_{C,D}}{D_{A,B}} \cdot \beta_1 \cdot h_{i-1,j} - \beta_1 \cdot h_{i,j} \right) \cdot \lambda_j \quad (5-20)$$

Case 9: Vehicle $i-1$ is in situation 3 and vehicle i is in situation 3.

$$\frac{D_{C,D}}{D_{A,B}} \cdot T_{i-1,A,B} - T_{i,C,D} = \left(\frac{D_{C,D}}{D_{A,B}} \cdot \beta_1 \cdot h_{i-1,j-1} - \beta_1 \cdot h_{i,j-1} \right) \cdot \lambda_{j-1} + \left(\frac{D_{C,D}}{D_{A,B}} \cdot \beta_1 \cdot h_{i-1,j} - \beta_1 \cdot h_{i,j} \right) \cdot \lambda_j \quad (5-21)$$

Equations (5-13) through (5-21) indicate that vehicles $i-1$ and i experience exactly the same running speed on the link between stops $j-1$ and j , and the reason why the travel time is not directly proportional to the travel distance is only attributed to the headway difference and thus the dwell time difference of the two vehicles. There are also many other methods to estimate the running speed experienced by vehicle $i-1$, such as a composite speed considering more than two successively observed locations, but they are more complicated than the approach given here. Furthermore, it is worthwhile to point out that, in equations (5-13) through (5-21), the dependent variable is just the expression on the left-hand side of the equations, and the independent variables are the linear expressions multiplied by boarding rates λ_j on the right-hand side of the equations. For instance, in equation (5-18), the independent variables are $\left(\frac{D_{C,D}}{D_{A,B}} \cdot h_{i-1,j} - h_{i,j} \right) \cdot \beta_1$ and $h_{i,j-1} \cdot \beta_1$, which multiply with λ_j and λ_{j-1} respectively.

By searching through all vehicle locations and matching them with the three situations in Fig. 5.3, the AVL data can be organized into a form as in Table 5.10.

Table 5.10. Regression Data Form (Assumption 2)

Stop 1	Stop 2	...	Stop N	Time
...
0.775	0	0	0	0.074157
...

The cells in each ‘stop’ column accommodate the observed values of the independent variable (linear function of headways) with λ_j ’s as the coefficient as in equations (5-13) through (5-21). The regression results are presented in Table 5.11.

Table 5.11. Regression Results – Passenger Boarding Rate (Assumption 2, 7:15 AM trip)

	λ_j (Passenger/Min)	<i>t</i> -statistic	<i>P</i> value
Fremont	-0.2	-0.06	0.952
Highland	0.8	0.42	0.671
Cherry	0.8	0.88	0.381
Campbell	0.1	0.12	0.903
Plumer	-0.3	-0.13	0.897
Tucson	0.2	0.08	0.934
Treat	-0.1	-0.06	0.955
Country Club	0.9	1.16	0.246
Randolph	0.4	0.6	0.546
Dodge	0.0	0.01	0.99
Alvernon	-4.3	-2.87	0.004
Irving	-21.2	-6.06	0
Columbus	-0.6	-0.43	0.67
Belvedere	-0.1	-0.05	0.96
Swan West	0.8	0.84	0.402
Swan East	-0.3	-0.27	0.784
Niven	-0.3	-0.09	0.931
Rosemont	0.4	0.33	0.741
Williams	0.1	0.04	0.966
Craycroft	0.1	0.09	0.925

In Table 5.11, the calibrated values of passenger boarding rates (λ_j) are either unreasonable (e.g. λ_j is -21.2 passengers/min at Irving stop) or not statistically significant (e.g. *P* value is 0.952 at Fremont stop). Therefore, one may conclude that Table 5.11 presents no statistical evidence to support the hypothesis that the current vehicle and the previous vehicle share the common running speeds. There might be two ways to interpret this. One possibility is that the vehicle running speed is more likely to be trip specific, rather than day specific; or equivalently, the vehicle running speed of different trips of the same day may be quite different. This possibility may be reasonable for this study, in the sense that there is very small likelihood that serious traffic congestion occurs on Broadway that it propagates to 10 minutes later when the next vehicle comes. Vehicle running speed over any short segment on Broadway is primarily

subject to randomness, rather than following any particular pattern. Another possibility could be that the vehicle dwell time might not be sensitive to the vehicle headway, i.e. vehicle dwell time is fixed, though also stop-specific. This possibility is examined by simply using 10 minutes as the headways for both vehicle i and $i-1$ at all stops of concern. The same regression models were estimated. Though not presented, the regression results do not suggest any improvement over what is shown in Table 5.11.

5.3.4. Assumption 3: Combined Vehicle Operating Behavior

Though the day-specific assumption may not be appropriate to predict vehicle travel time properly, the combined assumption considering both day-specific and trip-specific effects can be more relevant for estimating vehicle operating parameters, and is worth further examination.

In the view of equations (5-11) through (5-21), any attempt to consider both the trip-specific and the day-specific assumptions simultaneously may drive the regression functions to be polynomial in nature. However, another approach is available. Two-stage regression method can be employed to integrate the trip-specific and the day-specific effects, while all mathematical functions remain linear.

A simple comparison of the regression results from both trip-specific and day-specific assumptions may suggest that the former is relatively superior to the latter for the particular case of route 8 in Tucson. Accordingly, the two-stage method is constructed as follows:

Stage 1: Based on the trip-specific assumption, derive the average vehicle running speeds $\tau_{i,j-1,j}$ and passenger boarding rates $\lambda_{i,j}$ (to note that subscript “ i ” is added here

as the index to denote vehicle or trip, and to differentiate the passenger boarding rates for different vehicles or trips) for each particular trip. At this stage, the regression function is the same as equation (5-8) except the trip-specific notation for the passenger boarding rates, and its mathematical form is:

$$T_{i,A,B} = t_{i,B} - t_{i,A} = \sum_{j=2}^N \frac{P_{i,j-1,j} \cdot D_{j-1,j}}{\tau_{i,j-1,j}} + \sum_{j=1}^N (\delta_{i,j} \cdot h_{i,j} \cdot \beta_1 \cdot \lambda_{i,j}) \quad (5-22)$$

Stage 2: Use the vehicle running speeds $\tau_{i,j-1,j}$ and passenger boarding rates $\lambda_{i,j}$ derived in Stage 1 to predict the travel time for both the previous and the current vehicles. Then, derive the relationship of the prediction errors for both previous and current vehicles.

To complete Stage 1, the trip-specific assumption is also applied to the trip starting at 7:05 AM (the regression results for the trip starting at 7:15 AM have been presented earlier in Tables 5.4 and 5.5), and the regression results are given in Tables 5.12 and 5.13.

As argued previously, simply treating the running speeds of both the previous vehicle and the current vehicle the same may not be appropriate. However, a more reasonable assumption can be drafted: the vehicle travel time prediction error is entirely due to the vehicle running time variation; and, the current vehicle's running time variation is proportional to the previous vehicle's running time variation.

Table 5.12. Regression Results – Vehicle Running Speed (Assumption 1, 7:05 AM Trip)

	<i>Speed (mph)</i>	<i>t-statistic</i>	<i>P value</i>
Euclid - Fremont	15	9.4	0.00
Fremont - Highland	32	4.3	0.00
Highland - Cherry	34	3.2	0.00
Cherry - Campbell	17	10.4	0.00
Campbell - Plumer	20	9.4	0.00
Plumer - Tucson	28	5.2	0.00
Tucson - Treat	35 ¹	-	-
Treat-Country Club	33	3.4	0.00
C Club-Randolph	28	12.5	0.00
Randolph-Dodge	35 ¹	-	-
Dodge-Alvernon	18	11.9	0.00
Alvernon-Irving	23	6.2	0.00
Irving-Columbus	21	7.4	0.00
Columbus-Belvedere	35 ¹	-	-
Belvedere-Swan	25	4.7	0.00
Swan-Swan(Inter)	5	2.6	0.01
Swan - Niven	29	6.3	0.00
Niven – Rosemont	13	10.3	0.00
Rosemont-Williams	35 ¹	-	-
Williams-Craycroft	21	7.1	0.00
Craycroft-Leonora	32	3.2	0.00

Note¹: See Table 5.4.

Table 5.13. Regression Results – Passenger Boarding Rate (Assumption 1, 7:05 AM Trip)

	λ_j (Passenger/Min) ¹	<i>t-statistic</i>	<i>P value</i>
Fremont	- ²	- ²	- ²
Highland	0.9	9.1	0.00
Cherry	0.4	2.3	0.02
Campbell	0.8	12.2	0.00
Plumer	0.4	4.0	0.00
Tucson	0.9	9.4	0.00
Treat	0.5	2.5	0.01
Country Club	0.9	11.1	0.00
Randolph	0.7	12.0	0.00
Dodge	0.4	3.4	0.00
Alvernon	1.1	23.6	0.00
Irving	- ²	- ²	- ²
Columbus	1.0	15.5	0.00
Belvedere	0.5	4.0	0.00
Swan West	0.6	4.1	0.00
Swan East	0.4	1.7	0.10
Niven	- ²	- ²	- ²
Rosemont	1.1	16.5	0.00
Williams	0.5	5.4	0.00
Craycroft	1.0	15.5	0.00

Note¹: See Table 5.5.Note²: “-” means the value is not statistically significant.

Specifically, let the percentage difference of the predicted and the observed running times on the link between stops $j-1$ and j for the previous vehicle $i-1$ be $\varpi_{j-1,j}$, the percentage difference for the current vehicle i on the common link will be $\eta_{j-1,j} \cdot \varpi_{j-1,j}$. $\eta_{j-1,j}$ can be seen an augment/discount parameter, and it is both trip- and link-specific. Mathematically, equation (5-23) is assumed. Equation (5-23) is also the regression function at Stage 2 with $\eta_{j-1,j}$ as the only unknown factor to be calibrated.

$$T_{i,A,B} - \bar{T}_{i,A,B} = t_{i,B} - t_{i,A} - \left[\sum_{j=2}^N \frac{P_{i,j-1,j} \cdot D_{j-1,j}}{\bar{\tau}_{i,j-1,j}} + \sum_{j=1}^N \left(\beta_1 \cdot \delta_{i,j} \cdot h_{i,j} \cdot \bar{\lambda}_j \right) \right] \quad (5-23)$$

$$= \sum_{j=2}^N \frac{\eta_{j-1,j} \cdot \varpi_{j-1,j} \cdot P_{i,j-1,j} \cdot D_{j-1,j}}{\bar{\tau}_{i,j-1,j}}$$

wherein,

$\bar{T}_{i,A,B}$: Predicted vehicle i 's travel time from locations A to B with the parameters estimated from stage 1.

$\bar{\tau}_{i,j-1,j}$: Estimated vehicle running speed on the link between stops $j-1$ and j for vehicle i from Stage 1;

$\bar{\lambda}_j$: Estimated passenger boarding rate at stop j for vehicle i from Stage 1;

$\varpi_{j-1,j}$: The relative difference between the estimated running time and the observed running time on link between stops $j-1$ and j for vehicle $i-1$; and,

$\eta_{j-1,j}$: Parameters to be derived from AVL data.

If the AVL data is polled frequently, for one trip, there may be multiple observations related to the link between stops $j-1$ and j . For such case, a composite

estimation of $\varpi_{j-1,j}$ is preferred. For example, four vehicle locations a , b , c and d are observed successively as in Fig 5.4.

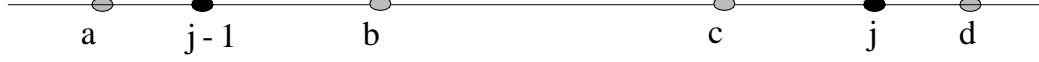


Fig. 5.4 Example for Estimating the Composite Difference of Running Time

The vehicle running speed $\bar{\tau}_{i,j-1,j}$ on the link between stops $j-1$ and j is involved in the vehicle travel times between a and b , b and c as well as c and d , and three vehicle running time relative differences $\varpi_{j-1,j}$ can be estimated for the link between stop $j-1$ and j . For instance, the percentage change $\bar{\varpi}_{1,j-1,j}$ based on locations a and b can be expressed mathematically as:

$$\bar{\varpi}_{1,j-1,j} = \left[T_{i-1,a,b} - \beta_1 \cdot h_{i-1,j-1} \cdot \bar{\lambda}_{j-1} - \left(\frac{D_{a,j-1}}{\bar{\tau}_{i-1,j-2,j-1}} + \frac{D_{j-1,b}}{\bar{\tau}_{i-1,j-1,j}} \right) \right] / \left(\frac{D_{a,j-1}}{\bar{\tau}_{i-1,j-2,j-1}} + \frac{D_{j-1,b}}{\bar{\tau}_{i-1,j-1,j}} \right) \quad (5-24)$$

The numerator in equation (5-24) is the vehicle travel time prediction error, and the denominator is the net vehicle running time from location a to location b . Furthermore, one may have noted that, in equation (5-24), the segment between location a and stop $j-1$, and the segment between stop $j-1$ and location b are treated the same in terms of the percentage change in vehicle running time. Similarly, $\bar{\varpi}_{2,j-1,j}$ and $\bar{\varpi}_{3,j-1,j}$ can be estimated from location pairs (b and c) and (c and d), and the composite percentage change $\varpi_{j-1,j}$ for the link between stops $j-1$ and j may be estimated by weighing $\bar{\varpi}_{1,j-1,j}$, $\bar{\varpi}_{2,j-1,j}$ and $\bar{\varpi}_{3,j-1,j}$ with $D_{j-1,b}$, $D_{b,c}$ and $D_{c,j}$ respectively:

$$\varpi_{j-1,j} = \frac{\overline{\varpi}_{1,j-1,j} \cdot D_{j-1,b} + \overline{\varpi}_{2,j-1,j} \cdot D_{b,c} + \overline{\varpi}_{3,j-1,j} \cdot D_{c,j}}{D_{j-1,b} + D_{b,c} + D_{c,j}} \quad (5-25)$$

In equation (5-25), more terms may be added or removed in a similar way.

With the estimated $\varpi_{j-1,j}$, equation (5-23) can be used to derive $\eta_{j-1,j}$ through regression, and the regression results for the trip starting at 7:15 AM (the vehicle operating on this trip is considered as current vehicle i) are presented in Table 5.14.

Table 5.14. Regression Results – $\eta_{j-1,j}$ (Assumption 3, 7:15 AM Trip)

	$\eta_{j-1,j}$	<i>t-statistics</i>	<i>P value</i>
Euclid - Fremont	- ¹	-	-
Fremont - Highland	-0.4207	-1.78	0.075
Highland - Cherry	- ¹	-	-
Cherry - Campbell	-0.3663	-1.99	0.047
Campbell - Plumer	- ¹	-	-
Plumer - Tucson	- ¹	-	-
Tucson - Treat	0.4523	2.06	0.040
Treat-Country Club	- ¹	-	-
C Club-Randolph	-0.1822	-1.76	0.078
Randolph-Dodge	-0.7037	-2.17	0.031
Dodge-Alvernon	- ¹	-	-
Alvernon-Irving	- ¹	-	-
Irving-Columbus	- ¹	-	-
Columbus-Belvedere	- ¹	-	-
Belvedere-Swan	- ¹	-	-
Swan-Swan(Inter)	0.5353	5.03	0.000
Swan - Niven	-0.8109	-4.02	0.00
Niven – Rosemont	-0.7469	5.67	0.00
Rosemont-Williams	-1.0258	-4.95	0.00
Williams-Craycroft	-0.3189	-2.90	0.004
Craycroft-Leonora	- ¹	-	-

Note¹: “-” means the value is not statistically significant.

According to Table 5.14, for about 50 percent of the links under analysis, the previous vehicle’s running time prediction error shows statistically significant impacts on the current vehicle’s running time prediction error on the same link. Relative to the travel time forecasting model developed solely under the understanding of trip-specific vehicle

operating behavior, the models incorporating $\eta_{j-1,j}$ are expected to improve the travel time prediction.

Furthermore, one may note that, for assumption 3, it is implicitly assumed that the derived passenger boarding rates based on the trip-specific assumption are perfectly accurate. However, this may not be true, and the actual number of the passengers boarding either vehicle $i-1$ or vehicle i is essentially a random factor. Less or more passengers boarding vehicle $i-1$ than the average may result in more or less passengers boarding the vehicle i than the average. This may partly explain why some of the $\eta_{j-1,j}$ values are negative, which implies the travel time prediction errors of the previous vehicle and current vehicle are negatively correlated. On the other hand, the positive $\eta_{j-1,j}$ values, or the positively correlated prediction errors for the previous and the current vehicles, can be interpreted by the fact that the current vehicle and previous vehicle may have experienced similar traffic situations, which increase/decrease the running time for both the previous and the current vehicles. Also in many situations, the passenger boarding effect and the traffic condition effect exist at the same time.

5.3.5. Comparison of the Methodologies

AVL data for the same three trips (6:55 AM, 7:05 AM and 7:15 AM) operating from August 5, 2004 through September 30, 2004, excluding the weekend days, is collected to illustrate the relative performance of the methodologies introduced previously.

The newly collected data went through the same data processing procedure as introduced previously, and was eventually organized into the form as Table 5.1. For the

trip starting at 7:15 AM from Laos Transit Center, with the vehicle location between the Euclid stop and the Fremont stop (location *A* in Fig 5.1), the vehicle travel times to all other downstream locations (location *B* in Fig 5.2) falling on the segment between Fremont and Leonora (excluding this stop) are predicted based on both assumption 1 and assumption 3. The prediction error is calculated as the difference between the observed vehicle travel time and the predicted travel time between locations *A* and *B*. The performance of assumptions 1 and 3 in terms of prediction error is presented in Table 5.15.

In Table 5.15, the first column represents the vehicle locations whose arrival times have been predicted; the second column and the fourth column are the root mean square error of the vehicle travel time prediction based on assumptions 1 and 3 respectively. The third and the fifth columns are the percentage prediction errors (RMSE / Average Vehicle Travel Time from location *A* to location *B* indicated in the first column) based on assumptions 1 and 3 respectively; and, the last column is the percentage difference of the RMSE presented in Columns 2 and 4.

Table 5.15. Model Performance Comparison

Vehicle Location	Prediction Error				Percentage Difference of RMSE (Understanding 1-Understanding 3) / Understanding 1
	Understanding 1		Understanding 3		
	RMSE (Sec)	Percentage	RMSE (Sec)	Percentage	
Fremont - Highland	20	37.0	20	37.0	0
Highland - Cherry	30	26.3	30	26.3	0
Cherry - Campbell	47	31.8	47	31.8	0
Campbell - Plumer	65	29.8	65	29.8	0
Plumer - Tucson	84	31.3	84	31.3	0
Tucson - Treat	83	25.1	84	25.4	-1
Treat-Country Club	95	25.7	89	24.1	6
C Club-Randolph	79	18.5	79	18.5	0
Randolph-Dodge	94	18.5	84	16.5	11
Dodge-Alvernon	84	15.8	87	16.4	-4
Alvernon-Irving	59	9.1	63	9.8	-7
Irving-Columbus	79	11.5	74	10.8	6
Columbus-Belvedere	73	9.6	77	10.1	-5
Belvedere-Swan	58	7.2	66	8.2	-14
Swan-Swan(Inter)	69	8.1	58	6.8	16
Swan - Niven	80	8.7	71	7.8	11
Niven – Rosemont	70	7.3	60	6.2	14
Rosemont-Williams	54	5.1	46	4.4	15
Williams-Caycroft	74	6.8	63	5.8	15
Caycroft-Leonora	84	7.1	78	6.6	7

According to Table 5.15, assumption 3 performs slightly better than assumption 1 overall. Relative to assumption 1, assumption 3 gives more than a 10 percent reduction in error for many locations, especially for the locations relatively far downstream. However, for both models, the RMSE is fairly large. This may imply that the parameters calibrated from the data collected in May, June and July may not be appropriate to used to predict

the vehicle travel time in August and September. This can be true due to several other facts. In Tucson, the transit passengers generally do not make as many trips in summer as in autumn due to the severe weather condition in summer. Also, May, June and July are the summer break time of the largest employer in Tucson, the University of Arizona. This may imply significantly fewer transit passengers in May, June and July than in August and September.

5.4. Conclusion

Although many models have been developed in previous studies for using AVL data to predict the transit vehicle travel time, how to use the AVL data collected specifically from Level *A* and Level *B* AVL systems to predict the vehicle downstream trajectories has not received much attention. Under three different assumptions of vehicle operating behavior, three methodologies are proposed in this chapter to use AVL data collected from a Level *A* or *B* AVL system to derive the vehicle running speeds and the passenger boarding rates. Relatively, the trip-specific assumption of vehicle travel speeds and boarding rates is superior to the day-specific assumption, according to the regression results. However, a simple simulation example points out that, under the trip-specific assumption, even when the parameters from the AVL data can predict the vehicle travel time with an acceptable precision, the derived parameters may not be similar to the actual ones. For the particular purpose to derive the vehicle running speeds and the passenger boarding rates using AVL data, extra requirements may be imposed for the AVL data polling frequency and the data's spatial resolution.

The day-specific assumption simply assumes the running speeds of the previous vehicle and the current vehicle are the same, but this assumption appears unreasonable in view of the regression results.

A two-stage regression method can integrate both the trip-specific and the day-specific vehicle operating behavior effects to yield a combined assumption on the vehicle operating behavior. Regression results based on this combined assumption suggest that, with the parameters calibrated from the trip-specific model, the vehicle travel time prediction error for the previous vehicle shows significant impacts on the travel time prediction for the current vehicle. Such a model is expected to improve the precision of the vehicle travel time prediction.

A performance comparison of the trip-specific model and the combined model using the AVL data collected in August and September 2004 shows slight superiority of the combined model over the trip-specific model. However, the vehicle travel time prediction errors are all fairly large for both models. This may imply that the significant temporal variation of the vehicle operating parameters exists.

REFERENCES

- Abdelfattah, A. and Khan, A., 1998. Models for Predicting Bus Delays, Transportation Research Record, No. 1623, 8-15.
- Abkowitz, M. and Engelstein, I., 1982. Temporal and Spatial Dimensions of Running Time in Transit System, Transportation Research Record, No. 877, 64-67.
- Abkowitz, M. and Engelstein, I., 1983. Empirical Methods for Improving Transit Scheduling, Research for Transport Policies in a Changing World, Vol. 2, 884-885.
- Abkowitz, M. and Tozzi, J., 1986. Transit Route Characteristics and Headway-Based Reliability Control, Transportation Research Record, No. 1078, 11-16.
- Abkowitz, M., Eiger, A. and Engelstein, I., 1986. Optimal Control of Headway Variation on Transit Routes, Journal of Advanced Transportation, Vol. 20, No. 1, 73-78.
- Abkowitz, M., Josef, R., Tozzi, J and Driscoll, MK., 1987. Operational Feasibility of Timed Transfer in Transit Systems, Journal of Transportation Engineering, Vol. 113, No. 2, 168-177.
- Abkowitz, M. and Lepofsky, M., 1990. Implementing Headway-Based Reliability Control on Transit Routes, Journal of Transportation Engineering, Vol. 116, No. 1, 49-63.
- Alfa, A., Menzies, W., Purcha, J. and McPherson R., 1988. A Regression Model for Bus Running Times in Suburban Areas of Winnipeg, Journal of Advanced Transportation, Vol. 21, No. 3, 227-237.
- Andersson, P-A. and Scalia-Tomba, G-P., 1981. A Mathematical Model of An Urban Bus Route, Transportation Research – B, Vol. 15B, No. 4, 249-266.
- Banks, J., 1984. A Model of Stoppings on Urban Bus Routes, Journal of Advanced Transportation, Vol. 18, No. 3, 207-228.
- Barnett, A., 1974. On Controlling Randomness in Transit Operations, Transportation Science, Vol. 8, No. 2, 101-116.
- Bookbinder, J.H. and Desilets, A., 1992. Transfer Optimization in a Transit Network, Transportation Science, Vol. 26, No. 2, 106-118.
- Chien, S. and Ding, Y., 1999. Applications of Artificial Neural Networks in Prediction of Transit Arrival Times.” Proceedings, 1999 Annual Conference of ITS America.
- Chien, S. and Schonfeld, P., 1998. Joint Optimization of a Rail Transit Line and Its Feeder Bus System, Journal of Advanced Transportation, Vol. 32, No. 3, 253-284.
- Dessouky, M., Hall, R., Nowroozi, A. and Mourikas, K., 1999. Bus Dispatching at Timed Transfer Transit Stations Using Bus Tracking Technology, Transportation Research: Part C, Vol. 7, No. 4, 187-208.

Dessouky, M., Hall, R., Zhang, L. and Singh, A., 2003. Real-Time Control of Buses for Schedule Coordination at a Terminal, *Transportation Research: Part A*, Vol. 37, No. 2, 145-164.

Ding, Y. and Chien, S., 2001. Improving Transit Service Quality and Headway Regularity with Real-Time Control, *Transportation Research Board 80th Annual Meeting Proceedings*.

Eberlein, X., 1995. Real-Time Control Strategies in Transit Operations: Models and Analysis, Ph.D. Dissertation, Department of Civil and Environmental Engineering, MIT.

Eberlein, X., Wilson, N. and Barnhart, C., 1998. The Real-Time Deadheading Problem in Transit Operations Control, *Transportation Research Part B: Methodological*, Vol.32 No. 2, 77-100.

Eberlein, X., Wilson, N. and Bernstein, D., 2001. The Holding Problem with Real-Time Information Available, *Transportation Science*, Vol. 35, No. 1, 1-18.

Ercolano, J., 1984. Limited-Stop Bus Operations: An Evaluation, *Transportation Research Record*, No. 994, 24-29.

Federal Transit Administration, 2000, *Advanced Public Transportation Systems: The State of the Art Update 2000*, Final Report.

Fu, L., and Liu, Q., 2003, A Real-Time Optimization Model for Dynamic Scheduling of Transit Operations, *Transportation Research Board 82th Annual Meeting Proceedings*.

Fu, L. and Yang, X., 2002. Design and Implementation of Bus-Holding Control Strategies with Real-Time Information, *Transportation Research Record*, No. 1791, 6-12.

Furth, P.G., 1986. Zonal Route Design for Transit Corridors, *Transportation Science*, Vol. 20, No. 1, 1-12.

Furth, P.G., 1995. A Headway Control Strategy for Recovering from Transit Vehicle Delays, *Transportation Congress: Civil Engineers--Key to the World Infrastructure*, Vol. 2, 2032-2038.

Furth, P.G., and Nash, A.B., 1985. Vehicle Pooling in Transit Operations, *Journal of Transportation Engineering*, Vol. 111, No. 3, 268-279.

Hall, R., 1985. Vehicle Scheduling at a Transportation Terminal with Random Delay en Route, *Transportation Science*, Vol. 19, No. 3, 308-320.

Hall, R., Dessouky, M. and Lu, Q., 2001. Optimal Holding Times at Transfer Stations, *Computers and Industrial Engineering*, Vol. 40, 379-397.

Hickman, M., 2001. An Analytical Stochastic Model for the Transit Vehicle Holding Problem, *Transportation Science*, Vol. 35, No. 3, 215-237.

Jarrah, AIZ., Yu, G., Krishnamurthy, N. and Rakshit, A., 1993. A Decision Support Framework for Airline Flight Cancellations and Delays, *Transportation Science*, Vol. 27, No. 3, 266-280.

Jordan, W.C. and Turnquist, M.A., 1979. Zone Scheduling of Bus Routes to Improve Service Reliability, *Transportation Science*, Vol. 13, No. 3, 242-268.

Kalaputapu, R. and Demetsky, M., 1995. Modeling Schedule Deviations of Buses Using Automatic Vehicle Vehicle-Location Data and Artificial Neural Networks, *Transportation Research Record*, No. 1497, 44-52.

Kikuchi, S. and Vuchic, VR., 1982. Transit Vehicle Stopping Regimes and Spacings, *Transportation Science*, Vol. 16, No. 3, 311-331.

Knoppers, P., and Muller, T., 1995. Optimized Transfer Opportunities in Public Transport, *Transportation Science*, Vol. 29, No. 1, 101-105.

Lee, K., and Schonfeld, P., 1991. Optimal Slack Time for Timed Transfers at a Transit Terminal, *Journal of Advanced Transportation*, Vol. 25, No. 3, 281-308.

Lesley, L. J. S., 1975. The Role of the Timetable in Maintaining Bus Service Reliability, *Proceedings of the 6th Public Transport Symposium*, University of Newcastle upon Tyne.

Levinson, H., 1983. Analyzing Transit Travel Time Performance, *Transportation Research Record*, No. 915, 1-6.

Li, Y., Rousseau, J-M. and Gendreau, M., 1991. Real-Rime Scheduling on a Transit Bus Route: a 0-1 Stochastic Programming Model, *Proceedings of the Thirty-Third Annual Meeting*, *Transportation Research Forum*, 157-166.

Lin, G., Liang, P., Schonfeld, P., and Larson, R., 1995. Adaptive Control of Transit Operations, Final Report for Project MD-26-7002, University of Maryland.

Lin, W-H. and Zeng, J., 1999, Experimental Study of Real-Time Bus Arrival Time Prediction with GPS Data, *Transportation Research Record*, No. 1666, 101-109.

Minitab, Inc., *Minitab 14.0*, 2004. Information at www.minitab.com, accessed March 15, 2005.

NCHRP Project A-15. 1999. Transit Capacity and Quality of Service Manual.

O'Dell, S. and Wilson, N., 1999. Optimal Real-Time Control Strategies for Rail Transit Operations During Disruptions, *Lecture Notes in Economics and Mathematical Systems #471: Computer-Aided Scheduling of Public Transport*, 299-323.

PAG (Pima Association of Governments), 2001. 2001-2025 Regional Transportation Plan, Final Report.

Powell W. and Sheffi Y., 1983. A Probabilistic Model of Bus Route Performance, *Transportation Science*, Vol. 17, No. 4, 376-404.

Strathman, JG., Kimpel, TJ., Ducker, KJ., Gerhart, RL., Turner, K., Griffin, D. and Callas, S., 2001. Bus Transit Operations Control: Review and an Experiment Involving TRI-MET's Automated Bus Dispatching System, *Journal of Public Transportation*, Vol. 4, No. 1, 1-26.

Sacramento Regional Transit, 2003. Glossary of Terms. http://www.sacrt.com/Short_Range/glossary.html, Accessed June 28, 2003.

Seneviratne, P. and Loo, C., 1986. Bus Journey Times in Medium Size Urban Areas, *Journal of Advanced Transportation*, Vol. 20, No. 3, 259-274.

Strathman, J., Kimpel, T., Dueker, K., Gerhart, R. and Callas, S., 2001. Evaluation of Transit Operations: Data Applications of Tri-Met's Automated Bus Dispatching System, Final Report TNW2001-04.

Suh, W., Chon, K-S. and Rhee, S-M., 2002. Effect of Skip-Stop Policy on a Korean Subway System, *Transportation Research Record* No. 1793, 33-39.

Sun, A. and Hickman, M., 2004. The Real-Time Stop-Skipping Problem, accepted by the Special Issue of the *Journal of Intelligent Transportation Systems*, June 2004.

Sun, A. and Hickman, M., 2004. The Holding Problem at Multiple Holding Stations, accepted by the 9th International Conference on Computer-Aided Scheduling of Public Transport, May 2004.

Talluri, K.T., 1996. Swapping Applications in a Daily Airline Fleet Assignment, *Transportation Science*, Vol. 30, No. 3, 237-248.

Transit Cooperative Research Program (TCRP), 2003. Uses of Archived AVL-APC Data to Improve Transit Performance and Management: Review and Potential, Final Report, Transportation Research Board of the National Academies.

Turnquist, M.A., 1978. A Model for Investigating the Effects of Service Frequency and Reliability on Bus Passenger Waiting Times, *Transportation Research Record*, No. 663, 70-73.

Turnquist, M.A. and Blume S., 1980. Evaluating Potential Effectiveness of Headway Control Strategies for Transit Systems, *Transportation Research Record*, No. 746, 25-29.

United States Urban Mass Transit Administration, 1983. Timed Transfer: An Evaluation of Its Structure, Performance and Cost, UMTA-MA-06-0049-83-6, Final Report.

Vuchic, V.R., 1973. Skip-Stop Operation as a Method for Transit Speed Increase, *Traffic Quarterly*, Vol. 27, No. 2, 307-327.

Vuchic, V.R., Clarke, R., and Molinero, A.M., 1981. Timed Transfer System Planning, Design and Operation, UMTA-PA-11-0021-82-2, Final Report.

Wall, Z., and Dailey D., 1999. An Algorithm for Predicting the Arrival Time of Mass Transit Vehicles Using Automatic Vehicle Location Data, Transportation Research Board 78th Annual Meeting Proceedings.

Welding, P.I., 1957. The Instability of Close Interval Service, Operational Research Quarterly, No. 8, 133-148.

Wilson, N., Macchi R., Fellows R. and Deckoff A., 1992. Improving Service on the MBTA Green Line through Better Operations Control, Transportation Research Record, No. 1361, 296-304.

Zhao, J., Dessouky, M., and Bukkapatnam, S., 2001. Distributed Holding Control of Bus Transit Operations, 2001 IEEE Intelligent Transportation Systems Conference Proceedings, 978-983.