ANALYSIS OF DYNAMIC TRAFFIC CONTROL AND

MANAGEMENT STRATEGIES

KHOO HOOI LING @ LAI HOOI LING

(B.Eng. (Hons.), MSc Eng., University of Malaya, Malaysia)

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF CIVIL ENGINEERING

NATIONAL UNIVERSITY OF SINGAPORE

2008

In Memory of My Late Grandmother

<u>Madam Gooi Siew Hong</u>

"I love you, Grandma. You are always in my heart and memory"

ACKNOWLEDGEMENTS

I would like to express my deepest appreciation to my supervisor, Assistant Professor Dr. Meng Qiang for his guidance, constructive suggestions and continuous support throughout my Ph.D. study in National University of Singapore. Many a time when I met with bottlenecks in my research, he always stands beside me, giving me valuable comments, advice and encouragement. With this, I am able to step through all the difficulties that I met in my research and study. Also, through his meticulous reviews and keen observations, the quality of my research is enhanced. I feel indebt to him.

I would like to utter my greatest appreciation to my Ph.D. study committees, Associate Professor Dr. Lee Der-Horng and Associate Professor Dr. Chan Weng Tat. Their continuous encouragement has made me progress well in my research study.

I would like to specially thank the National University of Singapore for providing the research scholarship for me during the course of research. Thanks are also extended to Mr. Foo Chee Kiong, Madam Yap-Chong Wei Leng, Madam Theresa Yu-Ng Chin Hoe for their assistance in handling the tools and software I required for my research study. Their kind co-operation has allowed me to complete my research smoothly.

I would like to thank my research mates: Dr. Huang Yikai, Dr. Raymond Ong, Jenice Fung Chau Ha, Cao Jinxin, Huang Yongxi, Dr. Alvina Kek and Dr. Wang Huiqiu for all kind of support and assistance they have provided me throughout my study in NUS.

Last but not least, the most sincere gratitude goes to my family and relatives for their endless love and long time support.

TABLE OF CONTENTS

SUMMARY

The imbalance between supply and demand in transportation system has caused traffic congestion to exacerbate for the past few decades. The worsening traffic congestion has resulted in negative impacts to the environment, society and economy. This problem has grown to an extent that it is now too complex for only one technology or technique to be "the solution". Hence, a set of dynamic traffic control and management strategies has been proposed to mitigate traffic congestion from various perspectives. However, methodologies used by many of these strategies required further improvement to ensure their effectiveness. In addition, proper modeling methods have to be adopted and more analyses need to be carried out to study the efficiency and effectiveness of strategies before implementation. This thesis serves to fulfill these purposes by proposing new strategies, enhancing current methodologies and mitigating the shortcomings of current models and algorithms.

Three traffic control and management strategies are studied in detail, namely contraflow operation, advanced traveler information system (ATIS)-based traffic control operation and ramp metering operation. Contraflow operation involves the reversal of travel lanes to cater for traffic demand and has been put in practice in many countries. In this thesis, two decision problems arisen from the operation, namely contraflow scheduling problem and contraflow lane configuration problem, are investigated. These two decision problems are formulated as bilevel programming models, which allow the capturing of the drivers' route choice decision during the optimization process. To solve the models, a hybrid meta-heuristics-microscopic simulation solution method is proposed. The numerical results show that the proposed

methodology is useful and allow the determination of better results compared to initial solutions.

Second, a novel ATIS-based online dynamic traffic control operation for urban expressway-corridor systems is proposed. The operation aims to maintain a certain level of service on the expressways by discharging additional vehicles to the arterial streets from the off-ramps. This could be achieved by deployment of ATIS tools to disseminate traffic congestion information to drivers. In addition, the thesis shows how drivers' compliance rate can be incorporated to ATIS under a microscopic traffic simulation environment. It is shown that the proposed methodology could bring a significant improvement in total travel time savings. A sensitivity analysis is performed to study how parameters such as the drivers' compliance rate can affect the performance of the proposed control operation.

Third, ramp metering operation is studied. A single level optimization model is developed to optimize the efficiency of the operation. A Probit-based ideal dynamic stochastic user optimal (DSUO) model is added as one of the constraints, allowing the drivers' route choice decision to be considered in the expressway-arterial network system. It is shown that, a significant of drivers divert to the arterial streets when ramp metering is applied. In addition, a fair ramp metering operation, which can balance both efficiency and equity, is examined. An equity index is defined to quantitatively measure the degree of equity of ramp metering operation. By maximizing the equity index, an equitable ramp metering can be attained. Furthermore, a multi-objective optimization model is developed to evaluate the fair ramp metering operation. Solving the proposed model gives a set of Pareto solutions, which indicates that the efficiency and equity issue is partially contradicted. The modified cell transmission model is employed in the analysis to simulate the dynamic traffic flow in the network. This is

advantageous since shockwave phenomenon and horizontal queue phenomenon can be modeled while the first-in-first-out (FIFO) principle is fulfilled.

LIST OF TABLES

LIST OF FIGURES

NOMENCLATURE

t A time interval in a discretized time period $[1,2,\cdots,T]$.

i A non-weaving or weaving expressway section.

j An on-ramp or an off-ramp.

J_i^{on} Set of on-ramps upstream of the traffic bottleneck Section i, on which the OnC mechanism will be implemented.

J_i^{off} Set of off-ramps upstream of the traffic bottleneck Section i, on which the OffC mechanism will be implemented.

l_i Total number of lanes in section i.

l_j^{off} Total number of lanes at off-ramp j.

l_j^{on} Total number of lanes at on-ramp j.

$N_i^k(t)$ Cumulative arrival of vehicles at lane k at section i at time t.

$O_i^k(t)$ Cumulative departure of vehicles from lane k at section i at time t.

$N_{j,k}^{\text{off}}(t)$ Cumulative arrival of vehicles at lane k of off-ramp j at time t.

$O_{j,k}^{\text{off}}(t)$ Cumulative departure of vehicles from lane k of off-ramp j at time t.

$N_{j,k}^{\text{on}}(t)$ Cumulative arrival of vehicles at lane k of on-ramp j at time t.

$O_{i,k}^{\text{on}}(t)$ Cumulative departure of vehicles from lane k of on-ramp j at time t at time t.

$a_i(t)$ Average number of vehicles per lane in section i at time t.

$a_j^{off}(t)$ Average number of vehicles per lane at off-ramp j at time t.

$a_j^{on}(t)$ Average number of vehicles per lane at on-ramp j at time t.

Π_i Target average number of vehicles per lane in section i.

Π_j^{off} Target average number of vehicles per lane that at off-ramp j.

Π_j^{on} Target average number of vehicles per lane that at on-ramp j.

A_{on}^{+} metered onramps.

A_{on}^{-} un-metered on-ramps.

A_{off} off-ramps.

A_E expressway stretches

A_{art} arterial road segments

$Q_{a_i}(t)$ The maximum possible number of vehicles through cell (a,i) in arc $a \in A$ when the clock advances from t to $t+1$.

$N_{a_i}(t)$ The maximum number of vehicles that can be presented in cell (a,i) in arc $a \in A$ at time t

$\beta_{a_i b_j}(t)$ Percentage of vehicles in cell (a,i) of arc $a \in A$ turning to cell (b,j) of arc $b \in A$ at time t.

$Z_a(t)$ The ramp metering rate imposed at onramp $a \in A_{on}^{+}$ when the clock advances from t to $t+1$.

v Free-flow speed.

w Backward wave speed.

τ Waiting time

$n_{a_i}(t)$ The number of vehicle in cell (a,i) of arc $a \in A$ at time t.

$n_{a_i \tau}(t)$ The number of vehicles in cell (a,i) of arc $a \in A$ that entered the cell in the time interval immediately after clock tick $(t-\tau)$.

$y_{a_i b_j}(t)$ The number of vehicles on link (a_i, b_j) of the cell-based network, which

connects cell (a,i) of arc $a \in A$ to cell (b,i) of arc $b \in A$, from clock tick

t to clock tick $t+1$.

$R_{a_i}(t)$ The maximum number of vehicles than can be received by cell (a,i) of

arc $a \in A$ in the time interval between t and $t+1$.

$S_{a_i}(t)$ The maximum number of vehicles than can be sent by cell (a,i) of arc

$a \in A$ in the time interval between t and $t+1$.

$H_a(\Delta_{l-1})$ Average number of vehicles at the onramp in the previous metering period

$(\Delta_{l-2}, \Delta_{l-1}]$.

$\rho(\Delta_l)$ Decision variable to be determined in the range $0 \le \rho_a(\Delta_l) \le 1$ for

metering period Δ_l.

$\overline{\chi}_a(\Delta_l)$ Lower bound of the ramp metering rate in the period of $(\Delta_l, \Delta_{l-1}]$.

$\widehat{\chi}_a(\Delta_l)$ upper bound of the ramp metering rate in the period of $(\Delta_l, \Delta_{l-1}]$.

$\overline{D}_a(C)$ average delay incurred by vehicles on arc a for metering rate C.

$I_k(C)$ equity index for group k.

$V_a(T)$ number of vehicles experiencing at least one time interval delay at on-

ramp a during time interval $[0,T]$

M total number of ramp metering period.

μ pre-determined parameter

z_{c_i} binary variable, $\{0,1\}$

CHAPTER 1 INTRODUCTION

1.1 Background

Traffic congestion is an inevitable problem faced in the metropolitan areas and urban agglomerations. It stems from the desires of individuals to pursue their daily activities at almost the same time on the same facilities. This has overloaded the available infrastructure and network systems. Traffic congestion has increased at a relatively constant rate since 1980s. If congestion in the early days only affects the area with crowded population, statistics has shown that, congestion today is more severe in terms of scope and scale compared to that in the 1980s. It has affected most of the roads, trips and time of day (Schrank and Lomax 2007). It is believed that in the coming few decades, this problem will further deteriorate with the ever-increasing number of people living on the Earth.

Proper planning on increasing transportation capacity (e.g. highway expansion) is essential to accommodate the increment of travel demand from growing population and rapid motorization. For this, much research effort were devoted in urban land-use planning and transportation infrastructure expansion. Research was focused on whether or not to build additional roadway given the budget constraints, the location of the new infrastructure, and the capacity expansion by adding additional lanes and so on. These problems were greatly studied by researchers since 1970s and can be termed as network design problems (Abdulaal and LeBlanc 1979; Boyce 1984; Yang and Bell 1998a; Meng 2000). Network design problems involve transportation system planning over a long period, for instance, 10 to 15 years. Static traffic flow functions/models were adopted to describe the traffic flow in the region during the planning process.

Such models are ineffective in the study of transient traffic congestion problem. Congestion lasts for an hour to two, often requiring a minute-by-minute analysis of traffic flow condition. In addition, building new roadways to meet current traffic demands often attract the occurrence of additional induced traffic (Hills 1996). Goodwin (1996) reported that an average road improvement would induce an additional 10% of base traffic in the short term and 20% in the long run.

Transient traffic congestion is caused by temporary inefficiencies in the control and management of traffic flow. It can happen when there is a sudden change in traffic demand or the traffic flow is changing abruptly by unexpected events. For example, a bottleneck is created during an accident that resulted in some lanes being blocked. An efficient traffic system should be responsive to changes quickly and adjust itself to avoid congestion. Hence, there is a need to better coordinate and manage the existing infrastructure at the operational level in order for the system to be efficient. In the old days, the most effective way of controlling traffic is through traffic signal. Much research effort has been devoted to improve the efficiency of traffic signal and to minimize queuing delay. Allsop (1974) was the pioneer in the study of traffic signals. He suggested that the traffic equilibrium theory should be embraced in the signal optimization process. Over the years, different control and optimization theories have become the main subject of the traffic signal study. To name a few, mixed-integer-linear-programming (Little 1966), hierarchical optimal control theory (Park et al. 1984), reserved capacity (Wong and Yang 1997), multivariable regulator approach (Diakaki et al. 2002), bilevel programming model (Chen et al. 2004) and meta-heuristics method (Ceylan and Bell 2004) were proposed. On the other hand, much effort was devoted to develop a better traffic flow model for optimizing traffic signal timings. Due to these fruitful research efforts, the efficiency of traffic signal has

improved. The system has evolved from the use of fixed traffic signal timing to pre-timed signals, then to traffic responsive signal control and to the current traffic adaptive control system. In the latest development, a prediction of traffic arrival is embedded in the control algorithm in adaptive traffic signal setting (Mirhandani and Head 2001).

Recent traffic management strategies employed real time traffic information to influence drivers' route choice decisions so that traffic can be distributed evenly within the network to avoid peak-hour congestion on certain roadways. Variable message signs are installed at the network bifurcation point to inform drivers about the traffic condition and the travel time on the network. Given this piece of information, drivers may choose to use alternative roads that are less congested and congestion levels on the busy roads are alleviated. In addition to this, congestion pricing (Yang and Huang 2005) is adopted to influence the drivers' route choice decision in order to ameliorate the traffic congestion level on some heavily used roadways.

Certainly, the above-mentioned strategies are not the only countermeasures that engineers can rely on to cope with traffic congestion. Indeed, the traffic congestion problem has become so severe that it is now too complex for only one technology to be "the solution". Schrank and Lomax (2007) recommended that a balanced and diversified approach should be adopted to ameliorate traffic congestion problem. In their report, they mentioned that different mix solutions depending on the type of development, the level of activity and geographic constraints may be utilized to solve the problem. Perhaps, the use of Intelligent Transportation Systems (ITS) technologies, which cover broad range of systems, could better coordinate, control and manage traffic for the sake of overall network improvement. Various countermeasure strategies are proposed in ITS, ranging from arterial management: contraflow operations (ITS

America 2008), freeway management: ramp metering (Chen and Miles 1999), infrastructure management and vehicle management. These various strategies could be adopted to mitigate the traffic congestion on the cities with different development level, level of activity and geographical constraints. They can be applied individually or integrated together to mitigate traffic congestion. In addition, with the advancement in computing and communications technologies, ITS have shown its potential in modern traffic control and management.

However, there is still much rooms for improvement in the current research and development of ITS. Many of the systems are still under development. Researchers are still studying the applicability of the dynamic route guidance system and the automatic vehicle control system (Peeta et al. 2000a; Lo and Szeto 2002a). More in-depth study is necessary to guarantee the feasibility of the strategies. Besides, some of the existing algorithms and strategies need to be improved. For example, critiques on the issue of ramp metering inequity have initiated the reinvestigation of the existing algorithms (Levinson and Zhang 2004). More importantly, proper traffic modeling and analysis tools are essential to evaluate the applicability of strategies and ensure that the desired outcome is reached without having to put drivers at risk. Currently, there are still many unresolved issues on dynamic traffic flow modeling. All these have highlighted the importance and the urgent need for research to be carried out with regards to the effectiveness and efficiency of traffic control and management strategies.

1.2 Research Objectives

The objectives of this research are:

1. To develop models and algorithms for alleviating traffic congestion on arterial roads and expressways.

2. To enhance existing methodologies to improve the efficiency of dynamic traffic management strategies.

3. To evaluate and analyze the applicability of the proposed models and algorithms.

4. To mitigate shortcomings of existing models and methodologies used in tackling traffic congestion problem.

1.3 Research Scope

There are many traffic control and management strategies in ITS system that can be adopted in alleviating traffic congestion. In this thesis, three of the most important strategies are investigated, namely contraflow operations, advanced traveler information system (ATIS)-based expressway traffic control operations and ramp metering operations. These strategies are practical and have been applied successfully on the field. However, this thesis is focused on the theoretical modeling of these strategies. Dynamic traffic network modeling is adopted to allow traffic to propagate over time and space. By using the dynamic traffic network analysis model, real-time traffic condition can be captured. Bottlenecks that cause congestion, queue propagation and spillback can be modeled. This enables the evaluation of various traffic control and management strategies.

There are two types of dynamic traffic network models, namely analytical model and simulation-based model. Both models have their own advantages. Analytical models have the advantage of providing specific and precise solutions. By using the analytical models, one can check whether a solution exists, unique and stable. In addition, one can examine the convergence properties of the solution algorithms and devise it. Hence, analytical models are favored by researchers. On the other hand, traffic engineers preferred simulation-based models. These simulation-based models

are able to capture vehicle interactions in detail, which enables the practical deployment for realistic networks. Proper choice of the dynamic traffic flow models is thus necessary so that the advantages of these models can be fully utilized. In view of this, this study adopted both approaches as the tools to model the dynamic traffic flow condition.

1.4 Organization of Thesis

Chapter 1 provides a general introduction to the traffic control and management strategies adopted to mitigate the traffic congestion. The importance and the need for current research are discussed. In addition, the objectives and the scope of the study are highlighted.

Chapter 2 is divided into three parts. The first part presents the review of strategies used to alleviate traffic congestion, namely contraflow operation, ATIS-based expressway control operation and ramp metering operation. The existing state-of-the-art of these strategies is outlined. The second part presents the dynamic modeling approach adopted to describe the dynamic of traffic flow over time. Both traffic simulation models and analytical models used in the literature are discussed in detail. The third part highlights the limitations of the existing studies, which inspired the need for this research.

Chapter 3 presents the scheduling and the lane configuration decision problems that arisen during contraflow operation. Both problems are considered individually and are formulated as bilevel programming models. The model consists of two levels; the upper level is an integer programming model with the objective function to minimize the total travel time, while the lower level describes the drivers' route choice decision following the solution given in the upper level. In addition, this chapter also proposed

a solution algorithm that consisted of a hybrid genetic algorithm and microscopic traffic simulation model to solve the bilevel programming models. A unique string repairing procedure is embedded in the solution algorithm to repair the infeasible solutions found by the genetic algorithm.

Chapter 4 presents a novel online ATIS-based dynamic traffic control operation for an urban expressway-arterial corridor system. Assuming that the real-time vehicle arrival and departure can be obtained from the loop detectors mounted on the expressway, the proposed traffic control operation aims to maintain a pre-determined level of service on a congested expressway section by disseminating the congestion information to the drivers through various ATIS means. The strategy consists of three components, namely Expressway Mainline Control (EMC), On-ramp Control (OnC) and Off-ramp Control (OffC) mechanisms. A simulation-based modeling methodology is developed to evaluate the proposed online dynamic traffic control operation. The evaluation is carried out by performing sensitivity analysis to the different control parameters used in the strategy.

Chapter 5 presents a detailed discussion on an analytical traffic flow model, namely cell transmission model (CTM). It was introduced by Daganzo (1994, 1995). The model shows added advantage compared to other analytical models because it could capture the shock wave and horizontal queue phenomenon and fulfill the first-in-first-out principle. In light of this, it is adopted as the traffic model for the analysis of the ramp metering operation in the subsequent chapters. A modification on the original version of CTM is carried out in order to accommodate the modeling of the merging situation of the on-ramp and expressway mainline segments. It is then termed as modified cell transmission model (MCTM) for the sake of presentation. The detail description of CTM and MCTM is highlighted in this chapter. An expressway-ramp-

arterial network system is employed to facilitate the elaboration on the modification done. The set of new formulas and rules derived for the merging cells are also presented.

Chapter 6 aims to optimize the efficiency of the ramp metering operation by taking into consideration the drivers' route choice decision. A Probit-based ideal dynamic stochastic user optimal assignment model (DSUO), which incorporates the MCTM model presented in Chapter 5 is developed to capture the route choice of drivers. An optimization model with the objective to improve the efficiency of the ramp metering operation by minimizing the total travel time is proposed. In the model, DSUO condition is defined as one of the constraints. This could be done since the DSUO assignment model is formulated as a fixed point problem. Genetic algorithm is employed to solve the proposed optimization model.

Chapter 7 presents a fair ramp metering operation that takes into consideration the efficiency and the equity issue. In this chapter, a novel equity index inspired from the make-span problem is proposed to capture the equity issue of ramp metering operation. If the equity index is maximized, a perfect equitable ramp metering system can be obtained. The equity issue of ramp metering is characterized by the average delay suffered by on-ramp drivers. Instead of deriving the equity index for each individual on-ramp, the methodology classifies the on-ramps into groups. A fair ramp metering is defined by considering both efficiency and equity issue simultaneously. A multi-objective optimization model is developed to optimize the fair ramp metering. MCTM is employed to simulate the dynamic traffic flow condition on the expressway-ramp network. The Non-dominated Sorting Genetic Algorithm (NSGA-II) embedded with MCTM is used to solve the proposed multiobjective optimization model. A set of Pareto solution is obtained for the fair ramp metering system.

Chapter 8 summarizes the main findings drawn from the current research and highlights their contribution to the state-of-the-art. It also provides directions and recommendations for future research.

CHAPTER 2 LITERATURE REVIEW

This chapter shall present a review of the literature on a few major aspects of this research. It is divided into three parts. The first part describes the past studies on the three important traffic control and management operations, namely contraflow operations, ATIS-based control strategies and ramp metering operations. The state-of-the-art of these strategies is discussed in detail. The second part of the chapter is devoted to discuss the dynamic traffic modeling approaches. Both the traffic simulation and analytical approaches are discussed. Finally, the last part highlights the weaknesses of the existing literature and the need for this research.

2.1 Dynamic Traffic Flow Control and Management Strategies

2.1.1 Contraflow Operations

Contraflow operations deal with temporarily reversing some lanes on one side of a two-way road to cater for the busy side of the road. It is a highly cost-effective dynamic traffic management strategy since significant capacity gains can be obtained without the need to construct additional lanes. The strategy is adopted to handle unforeseen events such as evacuation during hurricane attack (Theodoulou and Wolshon 2004), as well as to mitigate recurring congestion on busy roadways (Zhou et al. 1993). Generally speaking, the strategy is essential to alleviate traffic congestion arisen from insufficient roadway capacity due to unbalanced demand on both sides of a roadway. For example, during morning peak hour, more traffic will congest roadways that lead to the central business district (CBD) while the opposite trend is observed

during evening peak. This is similar to the traffic patterns observed during disaster evacuation. Prior to the implementation of contraflow operations, engineers need to solve two decision problems, namely the optimal contraflow scheduling problem (OCSP) and the optimal contraflow lane configuration problem (OCLCP).

2.1.1.1 Optimal Contraflow Scheduling Problem (OCSP)

OCSP is concerned with the determination of the start time and the duration of contraflow operations. According to a recent study carried out by Wolshon and Lambert (2006), the scheduling of the operation differs with the objectives of the strategy. If the contraflow lanes are implemented for the emergency related events, the implementation period can last for 2-3 days. Nonetheless, for the commuter traffic, planned events (like Washington Redskins football games) and transit bus, the implementation is carried out for a shorter duration, typically about 2-3 hours.

There is limited study on OCSP in the literature. Zhou et al. (1993) tried to compute an online optimal schedule for the George Massey Tunnel in Vancouver by adopting a fast optimization algorithm. The cost (objective) functions for the optimization are total delay and queue length, which can only be obtained from traffic data. To do this, they developed an intelligent controller comprised of a fuzzy logic pattern recognizer to estimate the flow demand pattern from partially available flow demand data obtained online from the detectors mounted in the Tunnel. They then adjusted the online optimal schedule according to the estimated pattern. However, the model is unable to predict traffic demand accurately as errors could occur during traffic congestion or when there is a lack of traffic counters. Xue and Dong (2002) made efforts to improve the estimation accuracy by incorporating a least square curve-fitting method to remove the random traffic noise. In both studies, no traffic flow

models are embedded. However, the schedule is amended according to the real time traffic demand prediction using some intelligent algorithms.

2.1.1.2 Optimal Contraflow Lane Configuration Problem (OCLCP)

OCLCP involves the determination of the number of lanes to be reversed from the opposite side of the road. Similarly, its application is also much dependent on the purpose of implementation. For example, Wolshon and Lambert (2006) pointed out that the number of lanes reversed may range from 1 to 3 lanes to cope with the increase commuter traffic due to pre-planned events (like football games). For emergency usage, such as during hurricane evacuation, all lanes of the opposite side can be reversed.

How one decides on the number of lanes to have their directions reversed therefore becomes a key issue in this problem. However, this issue received limited attention. Drezner and Wesolowsky (1997) made an initial attempt to determine the optimal configuration of one-way and two-way routes with flow-independent link travel times. In their study, the static traffic flow model is used to obtain transportation costs, which is required for the objective function. They used a pre-defined fixed link travel time for their network. A shortest route algorithm is then applied to calculate the shortest route costs given the contraflow configuration. Although the methodology proposed to determine the optimum configuration is feasible, the results are questionable as the traffic flow dynamic is not captured with the use of the pre-defined fixed link travel time.

Another study related to OCLCP was studied by Cova and Johnson (2003). They presented a network flow model for identifying optimal lane-based evacuation routing plans in a complex road network. The model is an integer extension of the minimum-cost flow problem. It is used to generate routing plans that trade total vehicle-distance against merging, while preventing traffic crossing-conflicts at

intersections. A mixed-integer programming solver is used to derive optimal routing plans for a sample network. The study concluded that a 40% reduction of total travel time is found, which is likely to vary depending on the road-network context and scenario.

Theodoulou and Wolshon (2004) presented an interesting case study regarding expressway contraflow evaluation on the westbound I-10 Freeway out of the City of New Orleans. They used CORSIM, a microscopic traffic simulation program, to evaluate two alternative contraflow lane configurations for one road segment along I-10 Freeway. The study concluded that proper planning and design of contraflow entry points is crucial. These studies could capture the dynamic of traffic flow well, but there is no proper methodology to search for the optimum lane configuration. The trial-and-error methodology was adopted in their study.

Tuydes and Ziliaskopoulos (2004) formulated the OCLCP as a continuous network design problem. In their standard continuous optimization formulation, capacities of reversible lanes are treated as continuous decision variables and a system optimum dynamic traffic assignment model without consideration of the selfish behavior of drivers in route choice is involved. They adopted a modified cell transmission model to simultaneously calculate the system optimal solution of the dynamic traffic assignment and optimal capacity reallocation in the contraflow operation. However, one shortcoming of the proposed method is the high computational cost associated with the analytical nature of the methodology, which prevents its use for actual urban networks. Due to this shortcoming, Tuydes and Ziliaskopoulos (2006) adopted a heuristic approach using Tabu Search to determine the (near) optimal solution. The authors performed a system optimal traffic assignment instead of a user equilibrium approach.

Lim and Wolshon (2005) studied a variation of contraflow configuration problem. They examined the contraflow termination point on the expressways during the hurricane evacuation. This study is essential because the merging of the traffic flow after contraflow can easily cause bottlenecks if it is not treated carefully. In the study, they used microscopic traffic simulation model to model and assess the contraflow evacuation termination points. A few assumptions had been made regarding the modeling of driver behavior under contraflow operations, such as travel speed of the drivers and lane choices.

Cantarella and Vitetta (2006) optimized the network layout and link capacity for different criteria considered by users, non-users and public system managers. Elastic travel demand is considered with respect to mode choice and the time-dependent departure time choice. In addition, the choice of parking location is also simulated in their study. The proposed formulation is solved by genetic algorithms. Another study done be Russo and Vitetta (2006) applied the topological method to solve the same problem. The methodology utilizes a "cluster" formation in relation to the solution topology and a "best" solutions extraction in relation to the criteria values. Promising results were obtained from the case study presented that confirms the applicability of the proposed method.

Kim et al. (2008) conducted a research study to find the optimal contraflow network configuration in order to minimize the evacuation time. In their study, a macroscopic traffic simulation model is adopted to describe the traffic flow condition on the network. They formulate the OCLCP using the graph theory assuming that the capacity of roadways is constant. Besides, their formulation restricts the occurrence of partial reversal, i.e. partial number of lanes is not allowed. The formulation is then

14

proved to be NP-complete problem and is solved by adopting greedy heuristics algorithms.

2.1.2 ATIS-based traffic management operations

Information about the traffic condition on the network is very important, to both the authority and drivers. This is because traffic condition on the roads can change abruptly due to unplanned events, such as accidents or vehicles break-down. Provision of travel information to drivers may allow them to choose alternative routes to avoid congestion. ATIS can be broadly divided into two categories, namely the pre-trip and en-route information. Pre-trip communication possibilities include internet, phone services, mobile devices, television, and radio. This type of information aims to influence the mode choice and departure time choice of a traveler. For a driver already on the road, he may receive traffic information through radio, variable message signs (VMS), or special in-car equipment to help him make rational routing decision at bifurcation points of the network (McQueen et al. 2002). En-route information can further divided into two sub-categories: prescriptive and descriptive information. Descriptive information provides information on traffic conditions only, with no routing advice while prescriptive information is used to advise drivers on routes without giving information on prevailing traffic conditions (Watling and Van Vuren 1993). ATIS-based traffic management operations refer to traffic management strategies that directly or indirectly use ATIS to manage traffic. Providing information to drivers with the aim to influence their route choice decision is considered an indirect way of managing traffic using ATIS. On the other hand, route guidance that provides shortest route guidance to drivers is a direct way of ATIS-based management operation. The subsequent sub-sections present the studies in both categories.

2.1.2.1 Descriptive Strategy: Traffic information only without route guidance

This group of strategies aims to disseminate traffic information to drivers on the road without suggesting detailed alternative routes. It is anticipated that by obtaining the travel information, drivers can decide the shortest route based on their knowledge of the network. The effectiveness of the method is thus relied on the drivers' behavior such as their ability to discern the information disseminated to them, their willingness to comply with those information and their ability to make quick decision on route-switching. There are two types of strategies used in collecting and disseminating travel time information to drivers, namely the reactive strategies and predictive strategies.

Reactive strategies disseminate travel time information to drivers based on real-time travel data collected from the network. Under such strategy, no prediction on the subsequent traffic condition is made. A practical example of this type of strategy is the Expressway Monitoring and Advisory System (EMAS), implemented in Singapore (LTA 2006). EMAS depends on camera systems installed on the expressways to collect real-time traffic data, and disseminate the information through the travel time display monitors installed on the side of the expressway. No additional prediction strategy is embedded in such a system. Many control rules and methods have been proposed for this strategy, for example, the P-regulator proposed by Messer and Papageorgiou (1994). In their study, they used the regulator to calculate the percentage of splitting rate, i.e., drivers' diversion rate at the expressway bifurcation point to achieve the user optimum flow condition. Although their model incorporates the drivers' compliance rate in determining the splitting rate, this method is found to be insensitive to the compliance rate. Other methods have been suggested to control the traffic density on the expressway. These include artificial neural network-based control

(Shen et al. 2003), iterative learning control (Hou and Xu 2003) and learning approach (Xu and Xiang 2005). To accomplish the objective to alleviate the traffic congestion on the expressway, these methods pre-specified density values that consistent to the desired traffic condition on the expressway. They then control the speed of the vehicles and the on-ramp traffic signals timing (Chien et al. 1997) in order to achieve the desired level. This means that such control methods assume drivers totally comply with the advice and opt to reroute. This poses a strong limitation to their implementation in practice. In the process to evaluate the proposed control algorithms, Shen et al. (2003), Hou and Xu (2003) and Xu and Xiang (2005) adopted macroscopic traffic simulation approach.

Balakrishna et al (2004) outlined a detailed simulation-assignment model to evaluate the ATIS with the flexibility to analyze the impacts of various design parameters and modeling errors on the quality of the strategy. The hybrid model used MITSIMLab, the microscopic traffic simulation model to model the ATIS-based operation while adopting DynaMIT, the mesoscopic traffic simulation model, to generate the driver response to the ATIS advice. The results confirm existing findings on overreaction, while providing valuable insights into the roles played by critical parameters that control simulation-based ATIS systems. The results also show that several exogenous factors such as traffic demand levels, incident characteristics, network structure and connectivity, availability of alternative routes, and assumed route choice behavior would affect the effectiveness of the system.

Predictive strategies attempt to predict traffic conditions sufficiently far in the future in order to improve the quality of the provided recommendations. Such strategy deploys current traffic state, control inputs and predicted future demand to provide traffic conditions. Such control schemes are known as Internal Model Control (IMC)

strategies in automatic control theory (Papageorgiou et al. 1994; Morin 1995). A real life example of IMC implementation can be found in Messmer et al. (1998), where the scheme is implemented on a Scottish highway network.

Ahmed et al. (2002) proposed a simulation approach to evaluate expressway diversion route plan which integrated with incident management systems using real-time traffic data. The proposed approach employed an anticipatory technique to estimate demand and incident severity based on current data and a library of historical traffic volume and incident data. Using the anticipated volume and incident data as inputs to CORSIM for the expressway and arterial systems network, an optimal decision about expressway diversion plans can be reached. The results show that the proposed framework produced optimal diversion route plans. The anticipatory technique used to predict the expressway demand during the incident provides accurate estimates and allows for a realistic representation of the traffic condition throughout the incident duration.

Some research studies were carried out to evaluate the effectiveness of the ATIS-based management operation. Chartterjee et al. (2002) carried out a questionnaire survey to study the response of drivers to information provided by the variable message signs in London. According to their study, only one third of drivers saw the information presented to them and few of these drivers diverted, although they found the information to be useful. An interview survey conducted in Paris found that 97% of drivers were aware of the existence of VMS, 62% of drivers completely understood the information presented on VMS, 84% of drivers considered the information presented to be useful and 46% has at least one occasion diverted in response to travel time information (MV2 1997). Another study in Scotland found that drivers diverted in 16% of the cases when a message indicated there was a problem on

their route (Swann et al. 1995). Through these studies, one can observe that the effectiveness and the benefits expected vary with driver behavior.

One important factor that can influence drivers' route choice decision is the information quality of VMS. This is because an important assumption on the effectiveness of the system is that drivers understand the message convey to them. A questionnaire survey done by Benson (1996) indicated that drivers are well disposed toward new types of VMS message that are simple, reliable and useful. Information about the exact locations of accidents and time-lagging traffic information are viewed favorably by drivers. Delay time estimates, on the other hand, was difficult to deal with. Furthermore this information can be inaccurate and might be presented in a format difficult to understand. Peeta et al (2000b) had developed a logit model to study the diversion behavior or drivers' compliance to the VMS message using data collected from stated preferences survey. Their study indicated that there is a strong correlation between VMS message type and driver response, and the message content is an important control variable for improving system performance without compromising the integrity of the information provided. Besides, they also found that there is a significant difference between the attitudes of truck and non-truck drivers. This finding proved that the effectiveness of the system is very much dependent on individual driver behavior.

Besides the information quality, the location of VMS is also a crucial factor that affects the effectiveness of the system. Abbas and McCoy (1999) proposed the use of genetic algorithm (GA) in order to optimize the placement of VMS to maximize potential benefits. The benefits are defined as the sum of the reduction in delay and accidents on the expressway upstream of the incident, the increase in delay and accidents on the alternate routes, and the change in delay and accidents on the

expressway sections downstream of the incident. Huynh et al. (2003) formulated the VMS location problem as an integer programming model with the objective of minimization of the network travel time in response to actual incidents. A greedy heuristics was then proposed to solve the problem.

2.1.2.2 Prescriptive Strategy: Route Guidance System

Route guidance system involves suggesting a detailed route for drivers who have vehicles equipped with navigation devices. Most of the system providers impose a certain amount of charge on the drivers for the service provided. This differs from the descriptive strategies in which the traffic information is provided free of charge to the drivers. However, route guidance has an advantage over descriptive strategies because it does not require drivers to have full knowledge of the network as the alternative route is proposed by instruction. The effectiveness of the system relies on the drivers' compliance rate and the market penetration of the service (Lo and Szeto 2002a).

The aim of route guidance systems is to achieve system optimal. Given the objective, the main task in route guidance is to find the dynamic shortest route from any origin to any destination. An important issue is then the modeling of the link travel time. If the travel time is assumed to be deterministic and time-dependent, classic labeling algorithms (Dijkstra 1959; Gallo and Pallottino 1984) can be adopted. However, if the link travel time is considered as random variables, the problem becomes more complicated. Most of the studies then tried to minimize the expected travel time (Mirchandani and Soroush 1986; Murthy and Sakar 1996). However, Hall (1986) mentioned that if the link travel time is assumed to be both time-dependent and stochastic, the classic route algorithms might fail to find the expected shortest route for passenger route choice problem in a transit network. Fu and Rilett (1998) found out

that this is the same case for road network. Some of the research studies adopted robust optimization technique to model the drivers' stochastic routing behavior (Kouvelis and Yu 1997; Bertsimas and Sim 2000; Lu et al. 2005a).

In addition, some studies adopted traffic simulation models or analytical models to evaluate the effectiveness of the route guidance system. Tsuji et al. (1985) developed mathematical models to guide vehicles through the shortest travel time routes in order to improve the effectiveness of the route guidance system. The models proposed in the paper are formulated by considering the stochastic nature of travel time. The parameters which characterize the models are defined and a quantitative analysis of the model is given for various values of parameters. It is shown that the effectiveness of route guidance can be estimated by the number of nodes along the route. The model is validated using field data collected in Tokyo.

Al-Deek et al. (1988) investigated the potential of route guidance under two congestion scenarios, i.e. recurrent and non-recurrent congestion along the Santa Monica Freeway in California. The FREQS and TRNSYT-7F models were used as macroscopic simulation tools to derive link travel time saving for the shortest route calculation. Shortest route between OD is determined based on different criteria such as travel-time shortest route, expressway-biased route, arterial-biased route and user-specified route. The result shows that with the route guidance system, there is no significant difference in travel time saving between travel time shortest route and other types of shortest route. However, it is evident that when the travel time shortest route is adopted, significant travel time saving can be achieved.

Jayakrishnan and Mahmassani (1990) evaluated the effectiveness of route guidance on the performance of a congested network using hybrid of simulation-assignment model. They integrated three main components of a traffic network under

real time traffic information into a single modeling framework. The main components were traffic flow, driver behavior and traffic information dissemination. Travelers' route choice decisions in response to traffic information were modeled individually. In addition, the proposed framework was capable in handling incidents in the traffic network.

Ben-Akiva et al. (1991) developed a dynamic network modeling framework that can be used to generate predictive information for a dynamic route guidance system and predict the effects of travel decisions made by informed drivers on the overall traffic conditions. The approach is based on a dynamic network modeling framework that incorporates driver behavior and network performance. It further extends the framework to incorporate drivers' acquisition and processing of traffic information which aims to capture the potential effects of new information services on individual drivers and overall traffic conditions. However, the model is deterministic and does not consider drivers' perception of error. Kachroo and Ozbay (1998) proposed a dynamic traffic routing and assignment as a real-time feedback control problem for the ATIS model. They represented the dynamic of traffic flow with partial differential equations. The information released by ATIS is then determined from the feedback of the traffic model, i.e. the flow. Wang et al. (2002) adopted the dynamic deterministic user equilibrium model to describe the route choice behavior of drivers in route guidance. The control strategy proposed then can be adjusted accordingly to the response of drivers.

Chen et al. (1994) developed a hybrid simulation-assignment model to evaluate the benefits of route guidance system. This model consists of three modules, namely multiple driver class traffic assignment module, traffic signal control module and route guidance module. There were five factors investigated, which were density of roadside

beacons, proportion of equipped vehicles, proportion of equipped vehicles in compliance with route guidance, levels of traffic demands and severity of incidents. They noticed that optimal proportion of equipped vehicles to gain maximum benefits was about 60%. They also concluded that route guidance could benefit incident management and minimized incident impact by diverting vehicles onto alternative routes.

Although the route guidance system is benefited, it is constrained by the requirement not to suggest routes that would dis-benefit complying drivers. This is because by suggesting unattractive routes to complying drivers could affect the credibility of the guidance system. Eventually the impact of the whole system may be jeopardized. However, if system optimal is the objective, this poses a challenge on determining ways to guide the drivers to achieve this objective. This is because under such objective, some drivers might need to suffer a longer travel time for the sake of the overall system. However, if user optimal is considered, it defeats the purpose of route guidance. More research is thus required to study this issue in detail.

2.1.3 Ramp Metering Operations

Ramp metering regulates vehicles at the on-ramps from entering into the expressways by a proper metering rate pattern and is a practical traffic control strategy to mitigate traffic congestion on the expressway system. A survey study carried out by Cambridge Systematics (2000) have demonstrated benefits of ramp metering, such as increasing expressway's throughput, reducing total system travel time and enhancing traffic safety. Papageorgiou and Kotsialos (2002) gave a comprehensive study on how and why ramp metering improves traffic flow. They showed that by using ramp

metering, the expressway mainline throughput can be increased for about 10-15%, which in turn decrease the travel time.

Theoretically, ramp metering is effective only if the traffic volume on the mainline expressway at the section immediately upstream of the ramp is less than the capacity. Under this condition, the application of ramp metering could ensure that the traffic volume delivered downstream of the ramp does not create a bottleneck due to the excessive demand from the upstream. While it is impossible to control the vehicles traveling on the expressway, the best choice is to regulate the entry vehicles from ramp. By ramp metering, it could break the "platooning" of entering vehicles for a more efficient merging. It could also reduce the demand of the expressway-ramp system by encouraging the diversion of vehicles to the surface streets (Wu 2001).

2.1.3.1 Ramp metering strategies

Fixed-time ramp metering strategies defined the ramp metering rate pattern using the historical flow and demand pattern of the expressway-ramp system, without real-time measurement. Under this strategy, the flow of vehicles allowed to move into the expressway is expressed by a ratio of the total demand of on-ramp, subjected to some constraints to ensure the feasibility of the metering rate. Wattleworth and Berry (1965) were the first researchers to propose this type of ramp metering algorithm. The drawbacks of this type of strategy can be easily pointed out. Demands on the network are not constant but change with different time period. The occurrence of unexpected events may perturb the traffic condition on the road, which lead to travel time variance with the historical data.

Reactive or traffic responsive ramp metering can address these limitations well. The metering rate pattern will be adjusted according to the real-time traffic condition on the expressway-ramp network system. If each on-ramp is assumed to be isolated in

the network, the metering strategy is termed as the local ramp metering. ALINEA, proposed by Papageorgiou et al. (1991) is a typical example of this type of metering strategy. It applied a classical feedback theory in optimizing the ramp metering rate. The basic idea behind the algorithm is to maximize the throughput on the expressway. The metering rate is adjusted automatically according to the traffic condition, by defining the occupancy term in the control logic. If the measured occupancy at certain cycle is found to be higher or lower than the desired occupancy, the algorithm can adjust by itself to decrease or increase the metering rate. Hence, the metering rate is responsive to the traffic condition on the expressway-ramp system.

If a few of vicinity ramp meters are considered and coordinated, the strategy is termed as area wide coordination strategy. ZONE algorithm (Lau 1997) is a typical example of this type of application. The algorithm divides a directional expressway into zones, and may contain several metered or non-metered on-ramps and off-ramps. The upstream end of a zone is a free-flow area and the downstream end of a zone generally is a critical bottleneck. METALINE (Papageorgiou et al. 1997) is another example of this strategy.

Besides these two categories, there are some algorithms that could produce the local metering rate as well as the system-wide metering rate. FLOW (Jacobson et al. 1989) algorithm is one example. The algorithm consisted of three components: calculation of metering rates based on local conditions called local metering rate (LMR), calculation of metering rates based on system capacity constraints called bottleneck metering rate (BMR), and adjustment to the metering rate based on on-ramp queue lengths. Pre-determined metering rates are selected on the basis of occupancy level upstream of the metered ramp. The LMR is obtained based on the traffic condition vicinity to the on-ramp. Coordinated BMR accounts for the independencies

among entrance ramp operations. BMR is calculated based on demand-capacity relationships. Finally, the metering rate is adjusted accordingly in order to take into account the minimum metering rate and ramp queues. BOTTLENECK (Jacobson et al. 1989) is another similar example for this type of strategy.

Since there are so many different ramp metering strategies available, researches have been carried out to study and compare the effectiveness of these strategies. Most of the studies adopted traffic simulation models for this purpose because field study is very difficult to be carried out.

Hellinga and Van-Aerde (1995) was one of the first researchers to evaluate the effectiveness of ramp metering implementation using simulation. They evaluate the time-of-day ramp control for a test network using INTEGRATION. They found that by implementing ramp metering, there was a slight reduction i.e. 0.39% in total network travel time. Based on sensitivity analysis, they also discovered that the traffic condition was influenced by the timing of ramp metering implementations. Matson and Daniel (1998) used CORSIM for the evaluation of time of day, fixed time metering in the Atlanta Metropolitan area. They simulate and compare the no-metering and with-metering case for the I-75 northbound corridor and found out that there is a decrement of 16.5% in total travel time and 19.7% increase in average speed for the freeway mainline after the implementation of ramp metering. Lahiti et al (2002) attempted to quantify the impact of simple ramp metering on average speed using CORSIM at the merge influence area under different traffic and geometric conditions. Using a hypothetical network, they found out that the average speed increase when on-ramp junction is metered and the increase is most prevalent under high traffic volumes, short acceleration lane and low number of mainline lanes. They used CORSIM in the simulation and constructed a hypothetical network to conduct the test.

Hasan et al. (2002) used MITSIM to evaluate and compare the effectiveness of ALINEA and FLOW algorithms. The purpose of their study was to evaluate how the level of traffic demand, queue spillback handling policy and downstream bottleneck conditions affect the performance of the algorithms when it was applied on the I-93 North expressway at Boston. They proposed a systematic evaluation methodology consisted of two stages. In the first stage, key input parameters for the algorithms were identified and calibrated. The calibrated parameters were then used for the second stage, where the performance of the algorithms was compared with respect to the variables using an orthogonal fraction of experiments. Regression analysis is performed to identify the impacts of some of the interactions among experimental factors on the algorithms' performance. It was concluded that the queue override strategy needs to be activated to suspend the ramp metering when it is no more warranted. FLOW algorithm is able to account for unspecified congestion spots than ALINEA. Ben-Akiva et al. (2003) employed the same network and same simulation model to evaluate a combination of dynamic traffic management system which consisted of ramp metering, route guidance and incident management algorithm. However, only ALINEA is evaluated in this paper.

Chu et al. (2004a) employed PARAMICS to evaluate three metering algorithms, namely ALINEA, BOTTLENECK and ZONE. The evaluation was conducted in a simulation environment over a stretch of the I-405 expressway in California, under both recurrent and incident scenarios. Simulation results show that ALINEA shows good performance under both recurrent and non-recurrent congestion scenarios. BOTTLENECK and ZONE are found to be more efficient in reducing traffic congestion while BOTTLENECK performs robustly under all scenarios.

Ozbay et al. (2004) proposed a new ramp metering strategy that takes into consideration the ramp queue. They modeled an isolated "feedback" based ramp metering strategy using PARAMICS on a calibrated test network located in Hayward, California. They tested and compared the results of this strategy, termed as Mixed Control Model with the ALINEA and New Control Model (Kachro and Ozbay 2003). Mixed Control Model adopts the concept of traffic responsive ramp metering strategy where both queue on the ramp and mainline condition is taken into consideration. Their results showed that Mixed Control Model can significantly improve system performance compared to the other two algorithms. Nevertheless, the queue override system proposed here is more to the *ad hoc* scheme. Sun and Horowitz (2006) proposed a more systematic traffic responsive ramp metering controller. In their model, a queue length regulator was designed to prevent the on-ramp queue from exceeding the storage capacity. In addition, a queue length estimator was designed to provide feedback to the queue length regulator, using the queue detector speed data that are available in the field. Results show that their model outperforms ALINEA.

Some parameters that need to be calibrated and validated before ramp metering can be applied on site. Chu and Yang (2003) used Genetic Algorithm (GA) hybrid with PARAMICS to calibrate the ALINEA parameters. Four parameters calibrated are: update cycle of metering rate, a constant regulator, the location and the desired occupancy of the downstream detector station. The fitness function used is the total travel time. Simulation results show that GA is able to be applied to find a set of parameter values that can optimize the performance of ALINEA algorithm. While GA is good for optimization, Beegala et al. (2005) proposed a general methodology for the optimization to Minnesota's Stratified Zone Metering (SZM). The proposed methodology, Response Surface Methodology (RSM) is claimed to be less

computational intensive and simpler. It consisted of two steps: the first step is to identify the parameters of the ramp control algorithm that most strongly influence its performance. This is done through detailed sensitivity analysis. The second step is to obtain optimal values for the parameters identified in step one. Three optimization thresholds, namely system total travel time, mainline total travel time and ramp total travel time were adopted as the optimization objectives. They employed AIMSUN in their paper. Results showed that RSM is suitable for the optimization in the case of SZM which involving multiple optimization objectives.

2.1.3.2 Optimization of ramp metering operation

In another group of ramp metering studies, the main focus was to optimize the effectiveness of ramp metering operation. The initial attempt in optimization ramp metering control was done by Wattleworth and Berry (1965) where linear programming model is proposed to optimize the ramp metering algorithm. The linear program proposed is a time-invariant optimization program which aims to maximize the total input to the system at all on-ramps constrained by a set of restrictions. In the study, however, a static traffic flow model with the link travel time function is employed to describe the traffic flow condition on the network. This has imposed a limitation to the study.

Papageorgiou (1983) proposed a hierarchical control system consisting of three control layers for the expressway control problem. The first layer is the optimization layer, where a simplified optimization problem is solved for the overall expressway system. Using the results obtained from the first layer, the second layer, which is an adaptation layer, updated the prediction of changes to the parameter exogenous to the system, for example, the demand variation. The third layer is the direct control layer in

which the disturbance of traffic conditions, for example, the traffic incidents can be accounted for by performing feedback control.

Besides, there is a wide variety of literature on the use of optimal control theory for ramp metering optimization. Papageorgiou et al. (1990) adopted the linear quadratic (LQ) optimization technique, a well known method of Automatic Control Theory to derive control strategies for traffic responsive and coordinated ramp metering. The control strategy adopted the typical LQ approach whilst the second one includes integration for some bottleneck densities. Both strategies are shown to be effective in ramp metering control. However, the approach adopted in the study linearizes the nonlinear traffic flow conditions, which made the control algorithms sensitive to traffic flow variations. The performance of the control strategy deteriorates when traffic disturbances is large. Zhang et al (2001) extended the study by introducing the nonlinear feature in the control strategy. This nonlinearity is realized by a series of neural networks. It was found that this type of control strategy outperforms the LQ optimization strategy by reducing total travel times, particularly at situations where drastic changes in traffic demand and road capacity occur.

Zhang et al. (1996) provided an in-depth study on the benefit of ramp metering by formulating the ramp metering system using the optimal control theory. They found out that ramp metering does not improve total travel time for a uniformly congested expressway system. However, this conclusion might not hold if vehicles can divert from ramp queues to alternative routes and the diverted traffic do not rejoin the controlled expressway section. Kotsialos and Papageorgiou (2004) adopted a constrained discrete-time nonlinear optimal control problem in the formulation of coordinated ramp metering system. In their formulation, the queue length limitation is embedded as a weighted factor in the objective function. It is observed that these

studies have similar objectives. They either minimize the total travel time or maximize throughput which concentrate only to the efficiency issue of ramp metering.

Besides optimal control theory, bilevel programming was proposed by Yang and Yagar (1994) to solve the ramp metering optimization problem. In their study, the upper level problem is to determine ramp metering rates that optimize a system performance criterion while the lower level problem is a traffic assignment where drivers' route choice and ramp queuing condition can be expressed. However, they adopted a static-based traffic flow model to describe the traffic flow condition and the queue at on-ramp.

During the optimization process of the ramp metering in these studies, macroscopic traffic simulation models are employed to describe the dynamic of traffic flow conditions. For example, Papageorgiou et al. (1990) and Zhang et al. (2001) adopted LWR model, Chang and Li (2002) adopted Payne model, and Kotsialos and Papageorgiou (2004) adopted second order model, Metanet. Although these models are much appropriate to capture traffic dynamic compared to the static models but they are point queue models, which are limited in describing the real queue situations especially at the on-ramps. In addition, most of these studies do not consider the effect of on-ramp control on drivers' route selection (Bellemans et al. 2003; Hegyi et al. 2002; Kotsialos et al. 2002). Chang and Li (2002) found that by incorporating the OD estimation using the Payne model in the optimization of ramp metering, the efficiency of the control could improve. However, by incorporating the OD information, Erera et al. (1999) has shown that segregating flows by destination introduces the problem of having to manage FIFO queue at the on-ramps, which makes the ramp metering problem intractable. Zhang and Levinson (2004) on the other hand suggested using the off-ramp exit percentages or splitting ratio as the replacement to the OD information

for ramp metering operation. They showed that by doing this, the same efficiency can be achieved.

2.2 Dynamic Traffic Flow Models

Dynamic traffic flow models are models developed to describe the minute-by-minute change of traffic flow condition on the network. The major purpose of any traffic control and management strategy is to control and manage the traffic in a more efficient way. At the operational level, the embedment of the dynamic traffic flow models is essential and necessary during the modeling and evaluation process of any traffic control and management strategies. How well the traffic is represented plays a crucial part in determining the performance and the efficiency of the strategy. This could be observed from the literature presented in the previous section, where each of the traffic control and management strategies is evaluated and analyzed by dynamic traffic flow models. Dynamic traffic flow models are more appropriate compared to the static traffic flow models for the operational-based strategies evaluation. This is because the dynamic change in the traffic flow conditions can be tracked and represented in detail. Dynamic traffic flow models have received great attention recently. Much research effort has been devoted to the development of traffic simulation models as well as analytical models. Both types of traffic models have their pros and cons and are described in the following sections.

2.2.1 Simulation Models

Traffic simulation models are computer software programs of traffic flow models, such as car following model, lane changing model and gap acceptance model, which can be adopted to describe the dynamic of traffic as well as drivers' behavior on

the network. There are many available software packages, which can be employed by engineers or researchers as evaluation tools. There are also many types of classification for traffic simulation models. It can be categorized according to the facilities modeled such as intersections, arterials, urban networks, expressways, and expressway corridor. It can also be classified according to the dynamic modeling system which is either continuous or discrete. The classification method adopted in this thesis is according to the level of detail, used in representing the system, i.e. macroscopic, microscopic and mesoscopic models.

2.2.1.1 Microscopic Traffic Simulation Models

Microscopic traffic simulation models describe the traffic behavior of individual vehicles and their interaction with each other and the road infrastructure. This behavior is captured in sets of rules which determine when a vehicle accelerates, change lane, and choose route. Car following models (Gazis et al. 1961; Newell 1961) enable drivers in the model to determine the accelerating and breaking patterns that result from the lead vehicle. Lane changing models (Gipps 1986) govern the driver's decision to perform the mandatory change to avoid obstacle or discretionary lane change to achieve his desired speed. Gap acceptance models (Daganzo 1981; Kita 1993) determine whether the lane change is executable or not by evaluating the available gap, either lead gap or lag gap. In addition, gap acceptance models are also applicable at the intersections when traffic from different directions merges. Microscopic traffic simulation models are also embedded with drivers' route choice models. Route choice models describe the decision of the drivers in choosing the route to travel from their origin to destination (O-D). Some of the traffic simulation models have the extended functions in route choice models in which the user equilibrium flow is calculated.

PARAMICS (PARAllel MICroscopic Simulation) (Quadstone 2005) is one of the most advanced microscopic traffic simulation tools available in the market. First, it allows the specification of the OD matrix and profile to simulate the time dependent traffic demand. This is of utmost importance especially in the framework of dynamic modeling where demand might change according to different time periods. Second, PARAMICS is embedded with three types of assignment methods to describe the drivers' route choice decision. It can perform all-or-nothing assignment, stochastic assignment and dynamic feedback assignment. All-or-nothing assignment is a type of static assignment in which all the drivers choose the shortest time/distance to use. Stochastic assignment characterizes the variations of drivers' behavior in route choice. This might cause some drivers to choose longer route. The dynamic feedback assignment allows the simulation of the response of drivers to the real time traffic information. There is no pre-determined route for drivers at the point of generation, but the route is chosen at real time based on the congestion level on the network at each time interval. Hence, the route trees of the drivers are updated at every time interval. All of this is governed by the general cost function embedded in the model, which has parameters that can be specified by the users. In addition, users may specify the percentage of drivers that will response to the dynamic feedback assignment. Drivers in the model are divided into two categories: familiar and unfamiliar drivers. The familiar drivers are assumed to have full knowledge of the network and the traffic flow condition on the roadways. Hence, they could divert to alternative roads when there is congestion on the main roads. This is carried out using the dynamic feedback assignment in PARAMICS. The percentage of familiar drivers is one of the inputs of the model.

PARAMICS is not a black-box model where users simply employ it to do analysis. It provides additional functions that allow users to interact with the core models of PARAMICS. The functionality of the model could be enhanced by Application Programming Interface (API). Extra commands can be set into PARAMICS to extend the existing functions provided by the model. This is done through Dynamic Link Library (DLL) created using C-programming. There are four types of functions available for the users, namely getting functions, extending functions, override functions and setting functions. Getting functions allow users to collect information from PARAMICS, for example, the simulation time; extending functions allow users to extend the function of the model. For example, after the simulation starts, users can extend its function by asking PARAMICS to read some file. Override functions allow users to override the original setting or logic of the model. For example, users can embed their own car following model in PARAMICS. For setting functions, users can set value to the parameters in the model. For instance, driver reaction time can be set. This has offered great flexibility for the users to do research using the tool. Due to its additional features, PARAMICS is adopted as the traffic simulation tool in this thesis. A brief comparison of some selected microscopic traffic simulation software package is shown in Table 2.1. The column with a cross means that the model has the mentioned features.

Table 2.1 Comparison of microscopic traffic simulation model

	PARAMICS	AIMSUN2	VISSIM	CORSIM	HUTSIM
Features Microscopic driver behavior*	X	X	X	X	X
Integrated network system#	X	X	X	With limitation of length modeled	Intersections
Queue spillback	X	X	X	X	X
ATIS/Route guidance	X	X^	N.A.	N.A.	N.A.
Route choice model /assignment	X	X	N.A.	N.A.	N.A.
Animation	X	X	X	X	Online animation
Extra functionality	Allow user defined functions	Linkage with TRANSYTS (Robertson, 1969) EMME 2,GIS	Has unique feature to model transit operation	Explicit modeling of expressway and arterial intersection design	Flexible and versatile object oriented

* Lane changing model, gap acceptance model and car following model
# Expressway and arterial network system
^ Carried out by external system

2.2.1.2 Other Traffic Simulation Models

Macroscopic traffic simulation models assume that the aggregate behavior of drivers depends on the traffic conditions in the drivers' direct environment. Traffic is modeled in analogy to fluids or gases in which its behavior is governed by sets of differential equations. Continuum models such as Lighthill-Whitham-Richards (LWR) model (Lighthill and Whitham 1955; Richard 1956) and Payne model (1971) are employed to derive the traffic condition, i.e. speed, density and flow over time and space. No individual vehicle or driver behavior is modeled in detail. METACOR

(Elloumi et al. 1994) and TRANSYT (TRAffic Network StudY Tool) are some of the examples of this type of traffic simulation model.

Mesoscopic traffic simulation model falls between microscopic and macroscopic model. It describes traffic characteristics at a high level of detail but their behavior and interactions are described at a lower level of detail. It models individual vehicles or platoon of vehicles but does not model the interaction between them. There is no car following model to describe how the vehicles follow their lead vehicle. Instead it employs headway distribution model to describe the headway between individual of vehicles. The vehicles can be model as individual vehicles or as a cluster of vehicles. The size of a cluster (the number of vehicle in a cluster) and the velocity of a cluster are of dominant important. The size of the cluster is dynamic in which it can grow or decay according to the traffic condition. The individual driver behavior within a cluster is not modeled. Hence, the cluster is homogenous in this sense. The velocity of the vehicles in the model is modeled using the gas-kinetic continuum model, which is macroscopically. The speed of the vehicles is defined by the speed-density relationship of the link. If the link has higher density, a lower speed will be assigned for the vehicles. Lane changing of the vehicles is not modeled. At nodes, the 'additional delay' for packets is calculated based on signal timing plans, average give way delays and others. The capacities at the node servers follow from saturation flows and their variance is calculated. Signal controlled intersection can be modeled by replacing the queue servers with gates that open and close according to the signal timing. DYNASMART (Jayakrishnan et al. 1994) and DynaMIT (Ben-Akiva 1996) are two typical mesoscopic traffic simulation models.

2.2.1.3 Limitations of Traffic Simulation Models

Traffic simulation models are like a black box to the users if the users do not understand the underlying concept behind the model. This is due to the complex nature of the traffic simulation models which makes the derivation of the associate mathematical properties is impossible. This causes traffic simulation models to lack mathematical properties in their model development. Calibration, verification and validation are very important steps in any simulation models in order to ensure that the system is represented correctly and the outcome from the models are reliable.

Hellinga (1996) had defined the terminologies of calibration, verification and validation. Model calibration is defined as the process by which the model user establishes input parameter values in order to reflect the local traffic conditions being modeled while model validation is defined to be the process of determining if the model logic proposed by the model developer is correctly represented by the computer code. Verification simply ascertains that the outputs from the computer code are consistent with the model logic. Model validation is defined to be the process of determining to what extent the model underlying theory and logic reflects reality. Calibration and validation are to be performed to ensure the reliability and credibility of models.

Although simulation models have emerged as an important tool in traffic control and management, the issue of calibration and validation still remains a point of concern to researchers. To date, there are no manual or special guidelines that can be adopted by the researcher when dealing with these issues. There are many on-going procedure proposed by researchers regarding this issue. Nevertheless, most of the calibration and validation effort is concentrated on microscopic model. This is simply

because microscopic simulation models are more complex and it is widely use in the modeling especially research pertaining to Intelligent Transportation System (ITS).

Many types of methodologies are proposed in literature. Many of the researches proposed systematic step-based and flow chart-based procedure in the calibration and validation process. Hourdakis et al. (2003) proposed a general and complete three-step calibration procedure, i.e. volume-based, speed-based and objective based calibration. Dowling et al. (2004) also proposed a three-step procedure which consisted of capacity and network calibration as the first step, route choice as second and system performance as the third step. Chu et al. (2004b) proposed a four-step systematic procedure, i.e. driving behavior (network), route choice, OD estimation calibration and model fine tuning. Optimization technique has also being employed to ease the calibration process. Cheu et al. (1998) adopted Genetic Algorithm (GA) to do the calibration on an expressway in Singapore using FRESIM model. Subsequently, Lee et al. (2004a) adopted the same approach but using PARAMICS as the simulation model. Kim & Rilett (2003) engaged the Simplex method to solve the optimization problem. Others method used are such as statistical method (Meritt 2003) and failure detection method (Wan et al. 2006).

Another fundamental issue is regarding the usage of traffic simulation models is calibration. Generally speaking, it can be classified into two categories: network based and demand based calibration. Basic network geometry such as the accuracy of the mapping of the simulation network to the real network needs calibration effort. This is extremely important for microscopic simulation model. The location of the traffic signal, the stop line, the road kerb point, the VMS location and etc may influence the drivers' behavior. In addition, parameters that govern the driver's behavior such as target mean headway and mean reaction time also need calibration.

The second component is the OD demand and the route choice model. In reality, it is very difficult to obtain these types of information.

Most of the calibration and validation effort is concentrated on the first component that is the network calibration. Gardes et al. (2002) calibrated the network and the model general configuration such as time step per second using PARAMICS for a freeway corridor. For route choice model calibration, Jayakrishnan et al. (2001) calibrated the PARAMICS route choice model using a hybrid model with DYNASMART while Mahut et al. (2004) calibrated the route choice model for a road network using method of successive average. Chu et al. (2006) recognized the importance of the OD matrix calibration in the validation procedure and calibrated the OD matrix for a freeway corridor using PARAMICS. Ma et al. (2006) also proposed a five-step procedure for micro-simulation calibration differencing the driving behavior, the departure time choice and the route choice model by using GA.

Calibration parameters and performance measures are another issue in the calibration and validation process. From the existing literature, it is observed that, the calibration parameters chosen for expressway calibration usually are volume, flow and speed while for the arterial road segment with intersection; maximum queue length, queue time and travel time are chosen. Nevertheless, the performance measure chosen in evaluation is inconsistent. Dowling et al. (2004) used mean square error, Brockfeld et al. (2005) adopted Theil's U statistic; Cheu et al. (1998), Lee et al. (2004a) used average relative error; Chu et al. (2006) adopted chi square statistic (GEH) and Kim and Rilett (2003) used mean absolute error ratio.

2.2.2 Analytical Models

Mathematical models are adopted to represent, describe and predict the time-varying traffic dynamics on the network in the process of network loading and assignment. The following section outlined these models. Focus was given to the cell transmission model (CTM), which will be used explicitly in the following chapters. Nevertheless, the other three functions: exit-flow functions, exit-time functions, and entrance-exit flow function are discussed in brief.

2.2.2.1 Cell Transmission Model (CTM)

The CTM developed by Daganzo (1994) is an approximate method solving the hydrodynamic or kinematic wave model of traffic flow (Lighthill and Whitham 1955; Richard 1956). The model assumes the fundamental relationship of traffic flow to follow piecewise linear functions as shown in Figure 2.1. Given that the roadway network is divided into cells and time is discretized into time steps, the vehicle flow among the cells within the time horizon could be calculated by two procedures. The first procedure calculates the flow of vehicle on a link from the current time interval to the next time interval. The second procedure updates the vehicle volume of the cell according to the flow conservation equations.
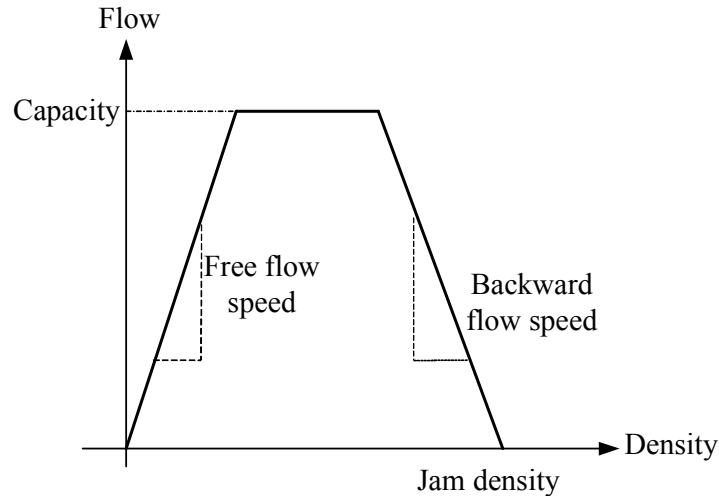
Figure 2.1 Piecewise linear function of fundamental flow-density relationship for CTM

The determination of vehicle flow across cells in the first procedure is the minimum of the three important parameters of the cells, namely current cell occupancy, capacity of the cell and availability of the cell occupancy subjected to shock waves. Daganzo (1995) extended the simple model to a general network. The cells in the network are divided into three main categories: ordinary if one link enters and one leaves. In the case of a network with some nodes whose degree is greater than three, Daganzo (1995) demonstrated that the cell-based network can be constructed by shortening the discretizing time interval. For an ordinary cell, the three parameters mentioned above are grouped into two groups to describe the vehicle flow. Two new terms are defined for each cell, the sending and receiving capacities respectively. The sending capacity determines the vehicle volume that a cell can send. It takes the minimum of cell capacity and cell occupancy. The receiving cell determines the receiving capability of a cell. It is the minimum of the cell capacity and the remaining cell occupancy subjected to shock waves.

For merge cells, if the receiving capacity of the receiving cell is greater than the sum of the sending capacity from both cells, all of the vehicle in the cells can flow into the merge cell. However, if the receiving capacity is lower, Daganzo (1995) showed that with the known proportions of the merge from each cell, the solution is the middle point of the parameters (refer Daganzo, 1995 for detail). For diverging cells, if the diversion proportion is known, the vehicle flow takes the minimum of the sending cell capacity, and the proportion of the receiving cell capacity. If the diverging proportions are not known, Daganzo (1995) suggested the computation of the minimum link waits using the cumulative curve of vehicles. Once the minimum wait time is determined, vehicles which waited longer than the minimum time are allowed to flow, and vice versa. Only some portion of the vehicles with the waiting time same as the minimum wait time are allowed to flow.

The advantages of CTM compared to other analytical models are its ability to capture shockwaves, model horizontal queue formation and dissipation, maintaining FIFO rule and dynamic traffic interactions across multiple links. However, a major difficulty associated with the model is that the recursive equation is non-differentiable. Szeto (2003) showed that the route travel time obtained using the model is non-differentiable and non-monotone. This poses a great challenge during the network loading process when performing equilibrium assignment. However, considering the advantages of the model, it is adopted in this thesis to model the traffic flow conditions in the expressway-ramp-arterial network system.

2.2.2.2 Other functions

Link exit-flow functions are built to express mathematically the rate at which traffic exits a link as a function of the volume of traffic presents on the link. The change of traffic volume on a link at the time interval is measured by the difference

between the entry flow and exit flow of the link at the time interval. This relationship was first used by Merchant and Nemhauster (1978) to describe the propagation of traffic. They used a static link performance function to represent the travel cost as a function of link volume in order to achieve a system optimal dynamic traffic assignment. In addition, they introduced the flow-conservation equation for a single destination and multiple origins as the constraints. Subsequently, Carey (1987) reformulated the M-N model as by introducing some extension to the exit function. This method has a serious flaw: it violates the first-in-first-out (FIFO) queue discipline as discussed by Carey (1986, 1992).

A better version of exit-flow function is to treat both link entrance and exit flows as control variables, instead of considering the exit flow only. However, by doing so, the FIFO violation problem has not been solved completely. In addition, a phenomenon of instantaneous flow propagation, in which the shock wave speed is greater than the free flow speed, may occur. Ran et al. (1993) tried to resolve this limitation by proposing an additional constraint to the cumulative number of vehicles entering and leaving the link.

Another method, link exit-time functions were proposed by Friesz et al. (1993). By specifying the relationship between the exit time of the flow of certain route at current time and its inverses, one can derive the flow-propagation constraints. The constraints are hence consistent with the chosen exit-time function dynamics. The applicability of the functions has been studied by Wu et al. (1998). However, it needs to be pointed out that the flow-propagation constraints assumed a point queue model.

By adopting proper models or functions to describe the traffic propagation on the network, one can perform network loading to find the traffic flow and travel time on the network. There are two types of user optimal flow, namely deterministic and

stochastic user optimal flow. Deterministic user optimal assumes that drivers have full knowledge of network and are identical in making choice. For the stochastic user optimal, it is assumed that drivers have partial knowledge of the network condition and the route choice behavior varies for different drivers.

2.2.2.3 Dynamic Deterministic User Optimal (DDUO)

By adopting the analytical traffic flow model, one can perform the route choice assignment to find the optimal flow on the network. According to Wardrop (1952) first principle, dynamic user optimal flow on the network is achieved when no driver can improve their travel time by switching route or change departure time. There are two types of travel time mentioned herein, namely the instantaneous travel time and the actual travel time. These two types of travel time constitute to the different type of optimal models. If actual (real) travel time is considered, actual (instantaneous) user optimal is achieved. The concept behind these types of models is same. The only different depends on the type of travel time derived from the traffic flow models. In the network loading process, DDUO assumes that drivers have full knowledge about the traffic condition on the network and behave identically.

Extensive research studies have been carried out related to DDUO. Ran and Boyce (1996a) discussed in detail the definition, formulation and solution algorithm to solve the model. According to Tong and Wong (2000), there are four approaches adopted to formulate the DDUO, namely mathematical programming formulations (see Merchant and Nemhauser 1978; Carey 1987; Janson 1991; Birge and Ho 1993; and Ziliaskopoulos and Lee 1996), optimal control formulation (see Friesz et al. 1989; Ran and Shimazaki 1989; Ran et al. 1993; and Ran and Boyce 1994), variational inequality (VI) approach (see Fiestz et al. 1993; Ran and Boyce 1996a; Ran and Boyce 1996b; Wie et al. 1995 and Chen and Hsueh 1998), and simulation-based (Peeta and

Mahmassani 1995a; Peeta and Mahmassani 1995b; Smith 1994; Ziliaskopoulos and Waller 2000).

Besides different approaches adopted, a review from the literature would observe the improvement in the traffic flow model adopted for describing traffic flow condition on the network. As mentioned in the previous sub-sections, the exit flow function is adopted in a few studies, for example, Merchant and Nemhauster (1978). Ran and Shimazaki (1989) adopted the entrance-exit flow functions to describe the traffic flow condition. In their study, the inflow is treated as a control variable while the exit flow is treated as the nonlinear function of the number of vehicle on the link. In addition, Ran and Boyce (1996a) defined sets of constraints, such as flow conservation constraints, flow propagation constraints, FIFO constraint and link capacity constraint to ensure that the traffic flow theory is obeyed and modeled correctly in their models. Chen and Hsueh (1998) adopted the same approach but without consideration of FIFO and link capacity constraints.

It is known that by using these *ad hoc* functions or constraints, a near-to-reality traffic flow condition cannot be guaranteed. These functions faced some flaw according to the traffic flow theory. They are point models and could not fulfill the FIFO discipline. Hence, in order to capture the traffic flow condition more accurately and to tackle the abovementioned limitations, Lo (1999) proposed to use CTM as the traffic flow model when perform DDUO assignment. By considering a general transportation network, it is shown how the actual travel time experienced by drivers can be calculated from the cumulative count at the origin and destination cells by using CTM. He formulated the ideal DDUO as a mathematical program and solved the formulation using the gap functions.

Subsequently, Lo and Chen (2000) and Lo and Szeto (2002b) also adopted the gap functions method to perform the dynamic user optimal. Lo and Szeto (2002c) proposed the VI formulation for ideal DDUO and considered the traffic flow generated from CTM as a functional mapping. This avoids the need to express the traffic flow conditions using the non-linear and non-convex constraints. If the mapping function could be guaranteed its continuity and co-coercivity, the VI formulation can be solved by the projection method. While acquiring the CTM as traffic flow model could address the traffic flow characteristics more realistically, Szeto and Lo (2004) show that in some traffic flow conditions, especially, during congestion, the actual route travel time function may become non-differentiable. In such a case, the user optimal condition could not be guaranteed. As a result, Szeto and Lo (2005) proposed a relaxation method to mitigate the problem. A tolerance-based DUO is introduced. This allows the algorithm to terminate earlier when the route travel time for used routes of the same OD is perceived no significant difference between routes, which is not necessary to be exactly the same.

2.2.2.4 Dynamic Stochastic User Optimal (DSUO) and Equilibrium (DSUE)

DSUO has an advantage compared to DDUO because the strong assumption of drivers having full knowledge on network conditions is eliminated. In addition, by formulating the drivers' route choice decision according to DSUO allows the specification of drivers' variation in perception of route travel time. If only the route choice decision is considered, the DSUO flow is obtained. However, if the departure time choice is considered simultaneously, DSUE flow is obtained. From the literature, it can be observed that most of the studies aim to find the DSUE flow. Horowitz (1984) was one of the first researchers to study the DSUE. Although the study only involves a simple two-link network with fixed O-D, it shows that when the equilibrium is unique,

link volumes may converge to their equilibrium values, oscillate about equilibrium perpetually, or converge to values that may be considerably different from the equilibrium ones, depending on the route choice decision-making process. Three route choice models are formulated and the stability of the equilibrium was analyzed. De Palma et al. (1983) proposed a stochastic equilibrium model to predict the traffic pattern and travel time on the road network during peak period. The study adopted the departure time model to describe the departure time choice of drivers. In addition, they used bottleneck model and waiting time model to consider queues and delays at a single point of insufficient capacity.

Ben-Akiva et al. (1986) modeled the between days dynamic of travel demand by using a hierarchical-based choice model. Furthermore, it was assumed that there is no transition costs imposed for changes of route and departure time choices. The day-to-day adjustment of the distribution of traffic is derived from a dynamic Markovian model. A deterministic queuing model is used to compute the travel time spent by drivers given the departure time on the specific day. It is a point queue model because the waiting time derived from the model is not affected by the physical length of the queue. Considering that the model developed by Ben-Akiva et al. (1986) is applied to single origin-destination (OD) pair without overlapping routes, Vythoulkas (1990) extended their model to a general networks with multiple OD pairs. A demand model was used to calculate the time varying pattern of departure rates and the demand was adjusted to cater for the day-to-day variation by the dynamic Markovian model. The difference of his study with Ben-Akiva et al. (1986) is the traffic flow model adopted. By assuming the traffic conditions within a link is homogeneous and the speed of a vehicle within a link is constant, the time varying O-D travel times and traffic flow pattern are calculated using elementary relationships from traffic flow theory and link-

volume conservation equations. Static Bureau of Public Roads (BPR) function is adopted as the link performance function in the study. Nonetheless, this could not ensure the FIFO discipline is always maintained and the equilibrium solutions are not always guaranteed. Both studies developed the logit-based dynamic user equilibrium.

While the previous studies used the day-to-day variation to describe the stochastic process of the travel demand, Cascetta and Cantarella (1991) proposed a model that can deal with both day-to-day and within-day dynamic. The dynamic user equilibrium is formulated as a fixed point problem, in which the within-day traffic flow and travel time is computed by assuming that the demand of the day is fixed. On the other hand, the fractions or the probabilities of drivers choosing the departure time and the route are modeled based on the day-to-day stochastic process. It is assumed that drivers adjusted their choices based on their past experience, which can be described by a day-dependent Markov chain process. To compute the within-day dynamics, they divided the links in the network system into two types, namely the running arc and the queuing arc. Basically, the former compute the flow travel time while the latter compute the delay due to queuing at the bottlenecks. They then derived sets of formulas to facilitate the computation. However, their model could not guarantee the FIFO discipline and it appears to be a point model. In a later study, Cantarella and Cascetta (1995) developed a theoretical framework to show the conditions for the existence and uniqueness of the fixed point formulation and equilibrium states by making use of results from nonlinear dynamic system theory. They found that the fixed point attractor will converge to a dynamic stochastic user equilibrium state if the system dimensionality is large enough. In addition, the solution is unique if the link cost-flow functions are strictly increasing and the users' choice behavior model is non-increasing.

Ran and Boyce (1996a) categorized two types of DSUO based on the derivation of route travel time. If the actual (real) travel time is derived from the traffic flow model, the ideal (instantaneous) DSUO is obtained. They formulated both types of DSUO using variational inequality (VI) approach. For the route choice model, they proposed two types of models, namely the logit-based model and the Probit-based model. Logit-based model has the Gumbel-distributed error term in the route choice probability function while Probit-based model describe the error term as normal distributed. They also showed that Probit-based model is better than logit model since it can capture the route overlapping situation and the variation in drivers' route choice perception behavior. In addition, the link entrance and exit functions are set to simulate the traffic flow condition on the network. To tackle the limitation of the functions in fulfilling FIFO principle, Lim and Heydecker (2005) adopted a deterministic queuing model to propagate traffic and calculate the route travel time.

While the earlier studies focused on describing the route choice behavior of drivers using the day-to-day variation method, in recent studies, this has been replaced by explicitly incorporate either a logit-based or Probit-based model. In addition, most of the studies do not have an appropriate way to propagate traffic. Some of the models used are point model in nature, while some could not guarantee the FIFO principle. This has inspired researchers to look for a better traffic flow model when performing the DSUO assignment. Sun et al. (2006), tried to use CTM as the traffic flow model to compute the instantaneous logit-based DSUO in a bilevel programming optimization of the traffic signal operations. By adopting CTM as the traffic flow model, the FIFO principle can be fulfilled.

2.2.2.5 Limitations of Analytical Models

Unrealistic traffic representation is one of the limitations of analytical models. Traffic behavior is complex in nature and it is very difficult to be represented dynamically. Although many functions are proposed and presented in the literature, they suffer different kind of weaknesses. For example, exit-flow function violates FIFO constraint, exit-time functions only can represent point queue and CTM has a non-deferential behavior in the model. All these limitations pose certain level of challenges in the formulation and the solution of the analytical model. A well representation of the dynamic of traffic flow is a key issue in ensuring the reliability and feasibility of the model.

Another limitation of analytical model is the computational issue. Boyce et al. (2001) recognized that analytical model is inefficient to solve large-scale real networks. For example, complete route enumeration is needed if the formulation is route-based (Wie et al. 1995), no efficient algorithms are available to some of the formulations (Ran and Shimazaki 1989). The only study shows that the analytical formulation can be used for large-scale network is Boyce et al. (1998). They applied a VI-based model to the ADVANCE network, which contains about 23,000 links, 9700 nodes and more than 360,000 trips in a two-hour assignment period. Their algorithm consisted of two parts, namely the inner problem and the outer problem. The inner part is aimed to solve the route choice problem using the Frank-Wolfe algorithm with fixed node time intervals. At the outer problem, the node time intervals and shortest route travel times are updated based on the updated average link flows from the inner problem. However, their model is not validated.

Nevertheless, the computational limitation can be tackled if a good initial solution, an effective search direction and proper step size are adopted. Boyce et al.

(2001) recognized some potential algorithms such as gradient projection and origin-based algorithms.

2.3 Limitations of Past Studies and the Need for Research

From the review presented in the previous sub-sections, it can be observed that there are some limitations in past studies. This section highlights the limitations of the three mentioned strategies which provoke the need for the research and further investigation.

2.3.1 Contraflow Operations

The limitations of the study regarding the contraflow operations could be highlighted in three points. First, existing study on contraflow operations is little and does not provide a systematic methodology to deal with the two decision problems. For example, Theodoulou and Wolshon (2004), and Lim and Wolshon (2005) adopted the trial and error approach to determine the OCLCP. They did not show how to optimize the contraflow operations, but only proposed a few configurations and tested the effectiveness.

Second, the existing studies are not suitable to be adopted during planning stage. For example, studies done by Zhou et al. (1993) and Xue and Dong (2002) adopted the real time traffic demand approach to predict the traffic flow for OCSP, which is not appropriate during the planning stage. Sufficient information and time are required to disseminate to drivers before application of the strategy. This is crucial for the traffic safety concern especially for drivers who are unfamiliar to the network.

Third, the existing studies do not consider the route choice behavior of drivers. For example, Lim and Wolshon (2005) carried out the simulation of the traffic flow

without route choice. Zhou et al. (1993) only predicted demand without finding the traffic flow on links. Tuydes and Ziliaskouplous (2004) aimed to find the system optimal in their effort to determine the capacity expansion for contraflow operations during hurricane attack. However, the system optimal approach is inappropriate because one could not expect drivers to follow the instructions given by various ATIS means. Drivers are selfish in nature and may choose to minimize their own travel time rather than for the overall system benefit.

To address the above-mentioned limitations, bilevel programming models are used to formulate the OCLCP and OCSP problems in this research. The models consisted of two levels. The upper level problem aims to optimize the contraflow operation by developing integer programming models. The lower level problem characterizes the response of drivers to the optimal solution given by the upper level, which is modeled by the microscopic traffic simulation models. With this, a proper and systematic off-line method that can be used during planning stage of contraflow operation is developed. In addition, the drivers' selfish behavior can be captured while optimizing the operation.

2.3.2 ATIS-based Traffic Management Strategies

It could be seen that the deployment of ATIS information to alleviate traffic congestion problem is still at its infancy. There are no established strategies or algorithms that deal with the issue of adopting of descriptive information to mitigate traffic congestion problem. Some of the studies such as Shen et al. (2003), Hou and Xu (2003), Xu and Xiang (2005) assumed a perfect drivers compliance rate, which is impractical. In addition, these studies used macroscopic traffic simulation models to describe traffic flow conditions. The behavior of individual drivers, such as their

compliance rate and awareness could not be modeled and captured in detail when the macroscopic traffic simulation models are adopted. In fact, this is crucial in ATIS since ATIS deals with individual drivers' behavior. Modeling drivers as a collective stream of flow such as in macroscopic traffic simulation model could not address the problem well.

In addition, there is increasing usage of ATIS tools in megacities. Large amount of investment is made for the installation of variable message signs for information dissemination purposes, installation of video camera and detectors for data collection and monitoring. Given such situation, it urgent to have a more systematic algorithm or strategy that uses ATIS systems in order to benefit drivers as well as for the sake of the overall network system improvement. In addition, it is important to adopt a microscopic traffic simulation model as the traffic flow model during the design and evaluation process since detailed driver behavior and their response to the ATIS information need to be captured.

This research, hence, fills the gap by proposing a novel ATIS-based traffic control operation for the expressway-arterial corridor system. It shows how the ATIS tools are deployed to alleviate the traffic congestion on the corridor system. Microscopic traffic simulation is adopted as a tool to evaluate and test the control operation. It also shows the method of calculating drivers' compliance rate to the ATIS advices in the microscopic traffic simulation environment. More importantly, it is an on-line operation that can be adopted by engineers.

2.3.3 Ramp Metering Operations

The existing research studies in ramp metering are focused on two main topics, namely adopting an appropriate traffic flow model to describe traffic flow condition

under steady state and disturbances, and optimizing ramp metering operations. Researchers (Wattleworth and Berry 1965) started with a simple static model. Eventually, dynamic traffic flow model is adopted because it is deemed more appropriate since the minute-by-minute change of traffic flow condition on the network corridor and the on-ramp queue can be captured. Many different macroscopic traffic simulation models that can describe the dynamic traffic flow condition are developed, e.g. continuum model, LWR model, and second order model. Nevertheless, these point models could not model the on-ramp queue appropriately. Indeed, the ability to capture the on-ramp queue is crucial to the operation. Recently, microscopic traffic simulation model is adopted. But unfortunately, it is only adopted to test, evaluate and compare different types of practical strategies. It could not be incorporated with analytical optimization methodology due to its mathematical complexity. Hence, the choice of an appropriate traffic flow model in ramp metering study still remains as an open topic. A more recent study has tried to adopt CTM, which can model the horizontal queue effect in the ramp metering study (Gomes and Horowitz 2006).

In addition, most of the research studies mentioned focus on the expressway-ramp network system only without considering the effects of arterial streets. As a result, the drivers' route choice decision is not captured. Wu (2001) through the field study showed that there is a significant diversion of traffic from the expressway-ramp network system to the arterial streets as drivers want to avoid the long delay at the on-ramp. Hence, it is crucial to incorporate the drivers' route choice behavior in optimizing the efficiency of the ramp metering strategy.

Moreover, the ramp metering studies are more focused on optimizing the efficiency of the ramp metering operations. Various types of optimal control theories

and operational algorithms are proposed. However, researchers have neglected an important issue in ramp metering study, namely the issue of equity. Ramp metering equity issue addresses the (in)equity of delay suffered by on-ramp drivers, which is experienced after the implementation of the strategy. For example, due to drivers' dissatisfaction, ramp metering in Twin Cities, America, have been shut down for re-evaluation. Research studies carried out following the shut-down of ramp metering shows that there is a need to reinvestigate the current algorithms to address the inequity issue (Levinson and Zhang 2006). However, there are only a few studies dealing with the ramp metering equity issue. Benmohamed and Meerkov (1994) are the pioneers in investigation of the equity issue arising from the ramp metering. They used an intuitive example with three on-ramps to show that average delays experienced by drivers on the two on-ramps are dramatically different if the dynamic ramp metering rate pattern follows the feedback law suggested by Papageorgiou et al. (1990). The difference ranges from zero to two time periods of interest. To distribute the average delays incurred on the interested on-ramps more evenly, they proposed a dynamic traffic control architecture that needs to calculate the ramp metering rate based on the decentralized control law and local measurements for the benefit of individual expressway's segment. Their traffic control strategy is built on several strong traffic flow assumptions, which are actually unrealistic for traffic control engineering. This is because their approach is motivated by a data network flow control method.

Papageorgiou and Kotsialos (2001) defined an equity constraint for each on-ramp that aims to keep the queue length at the on-ramp within a desired threshold. They added these equity constraints with different pre-determined weight factors as a penalty term to the objective function of their optimization model. These weight

factors have significant impact on the optimal ramp metering rate solution, while it is a challenging issue to determine these weight factors in real life. Kotsialos and Papageorgiou (2004) proposed an optimal freeway network-wide ramp metering strategy, termed as Advanced Motorway Optimal Control (AMOC) for the ring-road of Amsterdam, The Netherlands. In the study, the coordinated ramp metering control problem is formulated as a dynamic optimal control problem by constrained control variables which can be solved numerically for the given demands and turning rates over a given time horizon. They found that by imposing a maximum queue requirement to the formulation, the burden of ramp queuing needed to reduce the total system's travel time, is distributed among the on-ramps by their proposed strategy. In addition, they mentioned that, the equity issue is achieved at the expense of further improvement of the traffic conditions, proving that equity and efficiency are two partially competing properties of a control strategy. On the other hand, the new Minnesota metering algorithm has imposed a maximum ramp delay constraint that ensures delay less than four minutes per vehicle on local ramps and less than two minutes per vehicle on freeway to freeway ramps (Cambridge Systematic 2000). All of these practical equity considerations can balance efficiency and equity to some extent. However, as Zhang and Levinson (2005) pointed out, the impact of these strategies is difficult to determine in advance and the compromising process is achieved implicitly.

Instead of using equity constraints, Zhang and Levinson (2005) transformed the total travel time according to an assumed relationship to balance the efficiency and equity of ramp metering. In order to give more priority to the ramp's users, the delay encountered by the ramp's drivers is weighted non-linearly with respect to the actual waiting time. With a longer waiting time, a larger value of the weight parameter is allocated. Their method aimed to find a dynamic ramp metering rate pattern that

ensures on-ramps within same group encounter same delay by minimizing the transformed total system travel time. However, such a ramp metering rate solution may not exist. Yin et al. (2004) also proposed a concave transformation function to capture the equity issue. While the transformation relationship is important in the ramp metering equity issue, such function or relationship is not known in practice or may not exist. It is infeasible to assume a transformation function randomly because this could not help in solving the problem completely. Although Levinson et al. (2006) attempted to quantify the transformation relationship using virtual experience stated preference (VESP) method, the study could not come to a firm conclusion nor give out any concrete mathematical model. This shows that the study of the weighting relationship is still immature at current stage. In short, it can be concluded that, there is still lack of an appropriate methodology to address both efficiency and the equity issues simultaneously.

Besides methodology, there is also lack of a performance measure or indicator to address the ramp metering equity issue adequately. Most of the study measures the on-ramp delay directly as the indicator, for example, Kotsialos and Papageorgiou (2004), Zhang and Levinson (2005). Yin et al. (2004) proposed the use of Gini coefficient on the travel time as the equity performance index while Levinson and Zhang (2004) apply a Gini coefficient to measure distributional properties of trip delays. Hence, a proper equity index is yet to be defined in the literature to address the ramp metering equity issue. This is crucial because without an appropriate indicator, the problem can not be dealt with completely even with a good methodology.

Considering these limitations, this research works on four subjects. First, it proposes the use of a traffic flow model that can capture the horizontal queue phenomenon. A modified cell transmission model, which is a variation from the

original cell transmission model (Daganzo 1994, 1995) is proposed. By using this traffic flow model, the on-ramp queuing condition can be modeled in detail. Second, it developed a single level ramp metering optimization model to optimize the efficiency of the metering rate. A dynamic stochastic user optimal (DSUO) model is developed as one of the constraints of the optimization model, in order to address the route choice decision of drivers between the expressway and the arterial roads. Third, a novel ramp metering equity index is proposed to capture the ramp metering equity issue. It could guarantee the even distribution of the on-ramp drivers' average delay for the on-ramps within the same group. It can be adopted straightforward without the need of any type of transformation. Fourth, a multi-objective optimization model is developed to simultaneously optimize both the equity and efficiency issue. This could eliminate the need to specify any kind of transformation or relationship between these issues. With this, two important issues in ramp metering equity: a proper equity index and methodology are addressed.

CHAPTER 3 MODELS AND ALGORITHMS FOR THE OPTIMAL CONTRAFLOW OPERATIONS

3.1 Introduction

This chapter presents a systematic methodology to address two decision problems arising from contraflow operations, namely Optimal Contraflow Scheduling Problem (OCSP) and Optimal Lane Configuration Problem (OCLCP). The nature of contraflow operations is a typical leader-follower or Stackelberg game in game theory. The leader is the decision-maker or engineer who decides the schedule or the lane configuration of the contraflow operations, whereas drivers are followers who are individual decision-makers in response to the change in the transportation network. From modeling point of view, the leader-follower game can be formulated as the bilevel programming approach (Shimizu et al. 1997), which has been widely applied to model the static transportation network optimization problems (Migdalas 1995; Meng et al. 2001 and 2004). To formulate the contraflow operation problems, this chapter proposes a bilevel programming model which by its name consists of two levels of problems: the upper level and lower level problems. The upper level problem is an integer programming formulation that aims to determine a feasible operation solution. It is interesting to note that for OCLCP, a binary integer programming is proposed, in which a vector of binary (0-1) decision variables is defined to represent the lane configuration. The objective of the upper level is to minimize the total travel time of a study area. In the lower level, a microscopic traffic simulation model is employed to

capture the route choice response of drivers to the feasible contraflow operation solutions of the upper level problem.

To solve the proposed model, a hybrid meta-heuristics method coupled with a microscopic traffic simulation model is proposed. The reason for microscopic traffic simulation models at the lower level problem is that ability for them to capture the route choice decision of drivers in great details. These include interactions between vehicles, lane changing, routing choice, response to incidents, and behavior at merging points. Mesoscopic traffic simulation model can also be adopted to describe the dynamic route choice behavior of the drivers. However, microscopic traffic simulation models are deemed to be more appropriate since the contraflow operation problem is formulated as a lane-based model and drivers' lane choices can only be captured appropriately by the microscopic models.

To solve the upper level problem, meta-heuristics method is adopted. This is attributed to two main reasons. First, having employed a third-party or commercialized microscopic traffic simulation model as the lower level, the objective function of the derived upper level problem does not possess an analytical expression. Therefore, the conventional branch-and-bound or branch-and-cut algorithm for integer programming problems is no longer applicable to the proposed bilevel programming model. Without the implicit expression of the objective function, the relationship between the decision variables and the objective function is unknown. In such situation, it is the bound or cut for the proposed solution cannot be specified. Second, the proposed upper level integer programming model is an NP-hard problem in which the number of feasible solution increases exponentially with the increment of the problem size (number of candidate segment-lanes). For example, in a study area with five links each consists of three lanes, there could be 2^{15} feasible solutions for OCLCP. It is difficult to enumerate

all the possible solutions manually and meta-heuristic techniques are adopted to aid in searching the optimal solutions. Meta-heuristic methods such as genetic algorithm (GA) (Holland 1975), simulated annealing (Kirkpatrick et al. 1983) and Tabu search (Glover 1986) can be employed for solving the bilevel programming model because these solution methods only require the objective function value at each iteration. Nevertheless, In view of the successful applications of GA with microscopic traffic simulation models (e.g., Ma et al. 2004 and Cheu et al. 2004), GA is chosen as the solution method for the bilevel programming model for the contraflow operations. It should be pointed out that a unique string repairing procedure is incorporated to repair the infeasible strings.

3.2 Formulation of Contraflow Operations

Let $G = (N, A)$ denotes a connected and directed transportation network, in which $N = \{1, 2, \cdots, i, \cdots\}$, the set of nodes, and $A = \{(i, j) | i, j \in N\}$, the set of links. The sets of origins and destinations are denoted by R and S, respectively, and $R, S \subseteq N$. For an O-D pair rs, $r \in R$ and $s \in S$, there is a time-dependent traffic demand denoted by $q_{rs}(\tau)$ where τ represents a time instant. Each directed link $(i, j) \in A$ composes several lanes which are numbered from the kerbside (slow lane) to the median (fast lane) along the direction from nodes i to j by integer numbers $k = 1, \cdots, l_{ij}$, where l_{ij} is the total number of lanes in the link.

Figure 3.1 illustrates an example of the lane numbering scenario for two links connecting nodes i and j. Note that Figure 3.1 shows the right-hand-drive (keep-left) convention which is used in Japan and most of the Commonwealth countries/territories. The general formulation provided in this study can also apply to the European and

North American driving convention (left-hand-drive, or keep-right) without modification, except the coding of simulation model.
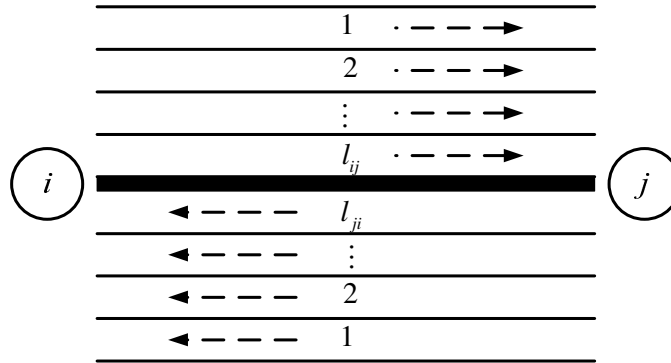
Figure 3.1 Lane numbering scenario

Let A_1 be set of links in the study area such as CBD, i.e. $A_1 \subseteq A$, A_2 be set of candidate links for contraflow implementation. Set of A_1 and A_2 are determined based on engineers professional experience and judgment. A_2 is the set of bottleneck links that have congestion problem while A_1 is the study area that is likely to be affected. Note that set A_2 may not be identical to set A_1 in practice. If a candidate link, say link (i, j), is selected, the number of lanes of its opposite link (j, i) will be augmented. Let set $\tilde{A}_2$ be the set of all the opposite links of the candidate links in set A_2, i.e.,

$$\tilde{A}_2 = \left\{ (j,i) \middle| (i,j) \in A_2 \right\} \tag{3.1}$$

To provide more flexibility on choosing a candidate link, we allow that $A_2 \cap \tilde{A}_2 \neq \varnothing$ which means that two links connecting the same node pair from different directions can be treated as candidate links.

Finally, define $[0, T]$ as the time period when the contraflow operation is considered. This could be pre-determined by traffic engineers. Assume that it is

discretized into homogeneous interval and each t represents a n-minute interval, namely, $t \in \{1, 2, \cdots, T/n\}$.

3.3 A General Bilevel Programming Framework

A general form of bilevel programming model that is adopted for the formulation of the contraflow operation problems is presented herein. The upper level of the proposed bilevel programming model for the contraflow operations

$$\min \ F(\mathbf{r}) = \sum_{n=1}^{N(\mathbf{r}, T, A_1)} \left(t_n^{out} - t_n^{in} \right) \tag{3.2}$$

subject to

Operational constraints

Definitional constraints

where $\mathbf{r}$ is the vector of decision variables for contraflow operations, which could either be the optimal schedule or the optimal lane configuration. $N(\mathbf{r}, T, A_1)$ defines the total number of vehicles travelling through the study area, A_1 within the simulation period, T. t_n^{out} and t_n^{in} are the time stamps when the nth vehicle enters and leaves the study area respectively. The objective function of the upper level is to minimize the total travel time vehicles spent in the study area and is expressed in eqn. (3.2). This includes any delay and waiting time in a queue, if any. It is possible that vehicles may originate from and/or destine to zones are within the study area (called internal zones). To include the travel times of these vehicles in eqn. (3.2), dummy zones are created outside the study area. These dummy zones are linked to the internal zones by dummy links (with zero travel time) so that vehicles can cross the boundary and be counted in the statistics. The decision variables, $\mathbf{r}$, are subjected to two constraints. The

operational constraints take into account the feasibility and the practical constraint of the operation. The lower level problem aims to describe the route choice behavior of drivers. It is represented by a microscopic traffic simulation model. The lower level problem is shown in Figure 3.2.

```
┌─────────────────────────────────────────┐
│ Input Step:                             │
│  - Network data associated with a feasible │
│    contraflow operation solution: r      │
│  - Time varying O-D demand               │
└─────────────────────────────────────────┘
                    │
                    ▼
┌─────────────────────────────────────────┐
│ Simulation Step:                        │
│  - Simulate the movement and behavior of │
│    individual vehicle in the network obtained │
│    from the input module during time [0, T] │
└─────────────────────────────────────────┘
                    │
                    ▼
┌─────────────────────────────────────────┐
│ Output Step:                            │
│  - Total travel time: F(r)              │
└─────────────────────────────────────────┘
```
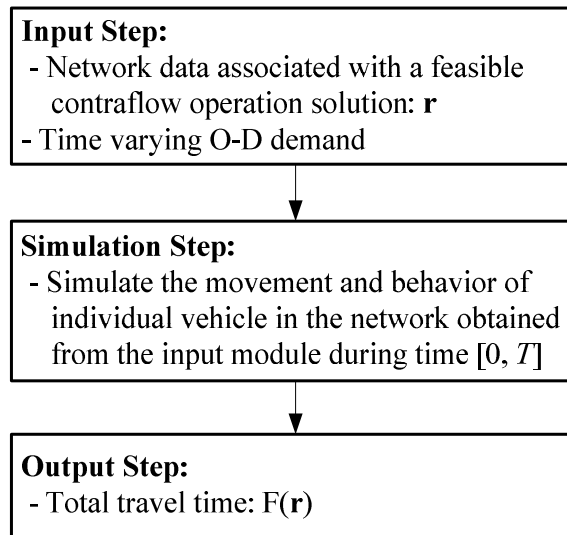
Figure 3.2 Lower level problem of the bilevel programming model

The microscopic traffic simulation model in the lower level consisted of three major steps, namely the input step, the simulation step and the output step. The input step defines the input of the model, such as the network layout and the time varying OD demand. In this step, the feasible contraflow operation solution, i.e. the schedule or the lane configuration from the upper level is captured. The network layout thus is changed according to these solutions. The simulation step simulates the movement of vehicles and drivers route choice in this amended network within the time period specified. Finally, the total travel time is calculated in the output step.

3.4 Optimal Contraflow Scheduling Problem (OCSP)

3.4.1 Bilevel programming model

The OCSP deals with the problem to determine the start time and the duration of contraflow operation. Prior to the determination of the optimal schedule, the configuration of contraflow operation, i.e. the number of lanes of each link $(i, j) \in A_2$ to be reversed is pre-determined. Moreover, it is assumed that only one shift of contraflow operation is allowed within the given time period. Thus, the integer decision variables of OCSP are defined as follows:

x_{ij} : start time (in time interval) of the contraflow operation for link $(i, j) \in A_2$

y_{ij} : end time (in time interval) of the contraflow operation for link $(i, j) \in A_2$

Note that since the time period is discretised, the decision variables take the integer value. For the sake of presentation, let **x** and **y** represent the row vectors of the start and the end times for all the links, respectively, namely

$$\mathbf{x} = \left(x_{ij} : (i, j) \in A_2 \right) \tag{3.3}$$

$$\mathbf{y} = \left(y_{ij} : (i, j) \in A_2 \right) \tag{3.4}$$

These decision variables are subjected to the following constraints:

$$\bar{T}_{ij} \le y_{ij} - x_{ij} \le \hat{T}_{ij}, \quad (i, j) \in A_2 \tag{3.5}$$

$$x_{ij}, y_{ij} \in \{1, 2, \cdots, T/n\}, \quad (i, j) \in A_2 \tag{3.6}$$

where $\bar{T}_{ij}$ and $\hat{T}_{ij}$ are the minimum and maximum durations (in terms of interval) of the contraflow operation implemented on the lanes of link $(i, j) \in A_2$. These two thresholds are set to take into account the traffic safety and operational feasibility. Note that $0 \le \bar{T}_{ij}, \hat{T}_{ij} \le \dfrac{T}{n}$. In consideration of the practical constraint in which drivers in

66

the normal direction need to be cleared before the implementation, a constant amount of time could be added to the objective function. For example, an extra of 20 minutes can be added to the objective function. This extra time is devoted as the preparation and clearance time for the traffic personnel who involve in the contraflow operations. Since this preparation and clearance time is a constant value, it has no effect to the optimization model of the upper level.

Given a feasible duration pattern, i.e. a $(\mathbf{x}, \mathbf{y})$ satisfying eqns. (3.5)-(3.6), the number of vehicles passing though the designated area A_1 during time period $[0, T]$, which is denoted by $N(\mathbf{x}, \mathbf{y}, T)$ can be obtained using the microscopic traffic simulation model shown in Figure 3.2. The total travel time experienced by those vehicles in the designated area during the time period is expressed by

$$F(\mathbf{x}, \mathbf{y}) = \sum_{n=1}^{N(\mathbf{x}, \mathbf{y}, T)} \left(t_n^{out} - t_n^{in} \right) \tag{3.7}$$

where t_n^{in} is the time of the nth vehicle entering the designated area A_1, and t_n^{out} is the time of nth vehicle leaving the area. Essentially, eqn. (3.7) is identical to eqn. (3.2).

To summarize, the optimal contraflow scheduling problem is formulated as a bilevel programming model:

$$\min F(\mathbf{x}, \mathbf{y}) \tag{3.8}$$

subject to

$$\bar{T}_{ij} \leq y_{ij} - x_{ij} \leq \hat{T}_{ij}, \quad (i, j) \in A_2 \tag{3.9}$$

$$x_{ij}, y_{ij} \in \left\{ 1, 2, \cdots, \frac{T}{n} \right\}, \quad (i, j) \in A_2 \tag{3.10}$$

where $F(\mathbf{x}, \mathbf{y})$ is evaluated by implementing a microscopic traffic simulation model shown in Figure 3.2 for any given feasible $(\mathbf{x}, \mathbf{y})$. The upper level is shown by eqn.

(3.8)-(3.10) while the lower level is shown in Figure 3.2.Comparing eqn.(3.8) to eqn. (3.2), the contraflow schedule $(\mathbf{x}, \mathbf{y})$ is equivalent to $\mathbf{r}$, and eqn. (3.8) is identical to (3.2). Eqn. (3.9) is the operational constraint and eqn. (3.10) is the definitional constraint mentioned in the general bilevel programming model in section 3.3.

3.4.2 Solution Algorithm

The hybrid GA-PARAMICS is adopted to solve the proposed bilevel programming model for OCSP. The detail of the encoding and decoding process and a step-by-step procedure are presented herein.

3.4.2.1 Encoding and Decoding of Strings

A binary coding method is adopted in OCSP to encode the decision variable, i.e. the start time and the end time of the contraflow operations. To encode, for a string, each feature represents a time interval while the value carried by the feature represents whether the contraflow is implemented or not at this time interval. The definition of the binary coding is as follows:

$$feature's\ value \begin{cases} 1 & \text{apply contraflow at this time interval} \\ 0 & \text{otherwise} \end{cases}$$

By checking the values carried by the features and marking their locations, the schedule for the candidate links could be determined. For a total operation period of T (minute) which is discretised into n-minute time interval, the length of a string (i.e. the total number of features), L_{OCSP} is expressed in eqn. (3.11) where $\tilde{a}_{ij}$ is the total number of candidate links:

$$L_{ocsp} = \left(T \middle/ n \right) \times \tilde{a}_{ij} \quad (i, j) \in A_2 \tag{3.11}$$

To elaborate the decoding procedure, a simple example is shown in Figure 3.3. Given two candidate links, a_1 and a_2 with time period of $T = 8$ and $n = 1$, the string obtained is shown as follows:
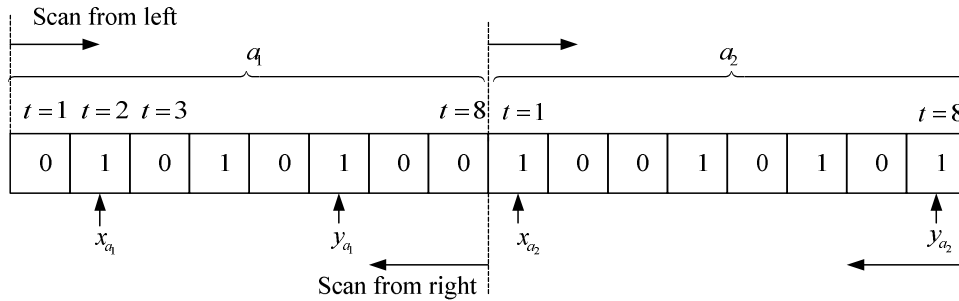
Figure 3.3 A binary string representation of contraflow schedule

Start scanning from the left of the string, the location of the feature that is the first to carry value 1 is marked as the contraflow start time of the link. To obtain the end time, scan from the right from the end of the string to find the feature that is the first to carry value 1. As shown in the figure, the start time and the end time of link a_1 and a_2 are at interval 2 and 6; 1 and 8 respectively. It is worth mentioning that the methodology proposed herein does not impose a restriction on the coding (decoding) methods used by readers. Hence, readers are free to choose any method that they are comfortable with. However, an inappropriate coding and decoding method will affect the convergence speed of the algorithm or the quality of the optimal solution value. Proper care and consideration is required when deciding the coding (decoding) of a problem (for more detail, please refer to Goldberg 1989).

3.4.2.2 String Repairing Procedure

The string repairing procedure is introduced to repair the infeasible solutions to ensure that the solutions obtained fulfill constraints (3.9)-(3.10).

Step 1. Perform the following operations for the entire string, $(\mathbf{x}, \mathbf{y})$ of links $(i, j) \in A_2$.

    Step 1.1. (First pass scan) Check each feature of the string $(\mathbf{x}, \mathbf{y})$, from left to right to identify the contraflow start time for all the candidate links, x_{ij}, $(i, j) \in A_2$.

    Step 1.2. (Second pass scan) Check each feature of the string $(\mathbf{x}, \mathbf{y})$, from right to left to identify the contraflow end time for all candidate links, y_{ij}, $(i, j) \in A_2$.

    Step 1.3. If $(y_{ij} - x_{ij}) < \bar{T}_{ij}$ or $(y_{ij} - x_{ij}) > \hat{T}_{ij}$, generate two random numbers that fulfill constraint (3.9). Assign the smaller number to x_{ij} and the larger number to y_{ij} .

    Step 1.4. Let g_{ij}^t be the feature's value (either 0 or 1) at time interval t of link (i, j). Set $g_{ij}^{x_{ij}} = 1$ and $g_{ij}^{y_{ij}} = 1$. Move to Step 2.

Step 2. Perform the following operation for the string $(\mathbf{x}, \mathbf{y})$ from Step 1.

    Step 2.1. Assign a value 0 for all the features before the start time.

$$g_{ij}^t = 0, \quad t = 1, 2 \ldots x_{ij} \text{-}1 \tag{3.12}$$

    Step 2.2. Assign a value 0 for all the features after the end time

$$g_{ij}^{y_{ij}} = 0 \quad t = y_{ij} + 1, y_{ij} + 2 \cdots T/n \tag{3.13}$$

This procedure aims to repair infeasible strings. Steps 1.1 and 1.2 are the decoding procedures that have been discussed in detail in the previous sub-section. Step 1.3 checks the feasibility of the string according to constraint (3.9). For infeasible strings, another 2 random integers are generated and are assigned to the string. Step 1.4 shows that the feature at the new start and end time is replaced with 1. In Step 2, those features before the start time and after the start time are set to 0. It is to ensure the

feasibility of the strings. This also applies to the end time. It is straightforward to verify that the string repairing procedure yield a feasible contraflow scheduling solution, i.e. a solution that fulfills constraints (3.9)-(3.10) of the bilevel programming model.

3.4.2.3 Genetic Algorithm (GA)

GA is inspired by Darwin's theory of evolution (Holland, 1975). It starts with a population of individuals represented by strings. Individuals from one population are taken and used to form a new generation of population according to their fitness – the more suitable they are, the higher chance they have to reproduce. In this study, it means that the schedule that leads to a lower total travel time have higher chances to be brought into subsequent iterations and vice versa. In the beginning of the algorithm, one needs to decide the coding method to represent the decision variable of the respective problem. There are two methods in GA, namely the binary coding and integer coding (Goldberg and Miller, 1995). Then, the string repairing procedure is necessary to repair the infeasible string obtained during any of the iterations.

A conventional GA consists of three main steps: selection, crossover and mutation. The selection step attempts to select individuals from the population to be parents for crossover. For each iteration, among the feasible strings, those strings with better objective function value will be chosen as the parents for crossover. Crossover means the exchange of some portion of the string between two selected strings. The newly produced strings after the crossover are termed as offspring. There are many ways to select a better string, for example, roulette wheel selection (Goldberg, 1989), tournament selection (Goldberg and Miller, 1995) and rank selection (Gen and Cheng, 1996). Since the binary encoding scheme of a string is chosen in this study, several crossover methods such as single point crossover or two point crossover (Gen and

Cheng, 1996) can be used. After the crossover, mutation strategy is carried out. Mutation means some features of the feasible string is changed according to some rules. It aims to manage the diversity of the solutions.

The step-by-step procedure of GA is shown as follows:

Step 0. (Initialization) Randomly generate a population of B strings, i.e. B sets of the contraflow schedule for the candidate links.

Step 1. (String repairing) For each infeasible string which violates constraints, repair the string by calling the string repairing procedure in section 3.4.2.2. The infeasible string in the population is then replaced by a feasible one.

Step 2. (Calculation of the fitness function) Given the solution, i.e. the schedule of the contraflow operation, evaluate the objective function, shown in eqn. (3.7) by implementing the microscopic traffic simulation model shown in Figure 3.2. The objective function value is assigned as the fitness function of the algorithm.

Step 3. (Generation of a new population). Repeat the following four sub-steps until the new population is completed.

Step 3.1. (Selection) According to the fitness function value evaluated in Step 2, use the rank selection method to choose two parent strings (i.e. two strings from previous iteration).

Step 3.2. (Crossover) With a crossover probability, denoted by p_c, cross over the parents to form a new offspring according to the two point crossover method (Gen and Cheng, 1996). If no crossover is performed, the offspring is the exact copy of the parents.

Step 3.3. (Mutation) With a mutation probability, denoted by p_m, mutate new offspring at selected position in the string. This means that for

certain features that have been chosen, the value of the feature is changed. If originally the feature has value 1, it is changed to 0 and vice versa.

Step 3.4. (String repairing) If the offspring cause infeasible solution to the upper level problem, they are repaired.

Step 4. (Stopping criterion). If a stopping criterion is fulfilled, terminate, and output the best solution. Otherwise, go to Step 2.

There are two termination criteria available for Step 4. GA can be terminated when it achieves a maximum number of generations specified. It can also be terminated if there is no improvement in the fitness function value for a number of generations specified.

3.4.3 An Illustrative Case Study

3.4.3.1 Network Coding

In this study, PARAMICS is adopted as the lower level problem of the bilevel programming model. By inputting the feasible contraflow schedule $(\mathbf{x}, \mathbf{y})$, the study network is simulated and the objective function value indicated by eqn. (3.7) can be evaluated. The network used in the case study covered the CBD area of Singapore. It was represented by 2046 links, 1050 nodes, and 36 zones. The skeleton of the network with main roads and origin-destination (OD) is shown in Figure 3.4. The network was calibrated by Lee et al. (2004b). The parameters calibrated are such as vehicle types, percentage of vehicles and network layout. Among these links, 18 links numbered from 1 to 18 indicated by hatched links in Figure 3.4 formed the study area, i.e. $A_1 = \{1, 2, \cdots, 18\}$ within which the GA's fitness function at the upper level problem was
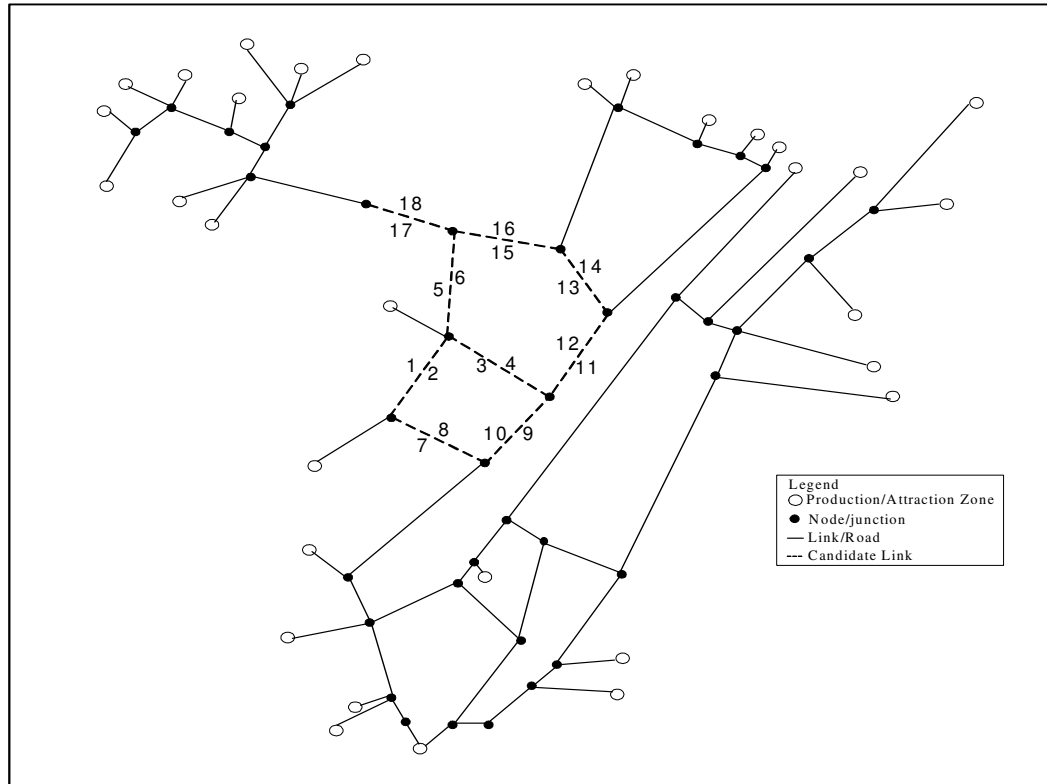
Figure 3.4 A skeleton of the study network

calculated. Among these links, only 9 links have been recognized as the candidate

contraflow links, i.e. $A_2 = \{2, 4, 6, 8, 10, 12, 14, 16, 18\}$. In the real situation, these links

can be chosen based on the traffic condition, for instance, those congested links. A set

of assumed values is assigned as the default OD demand for the network. In order to

emulate the dynamic departure time choice, the demand is loaded onto the network at

5-minute interval according to a profile specified arbitrarily. However, it is found that

the outcome of the methodology proposed is not affected by the loading pattern. Other

PARAMICS parameters are taken from Ma et al. (2004).

In this case study, the dynamic feedback traffic assignment method in

PARAMICS was chosen to reflect the dynamic behavior of drivers, includes route

choice, in response to the change of transportation network topological structure.

Under this method, PARAMICS updates the link costs that include turning cost. The

turning delays are calculated at defined feedback period, i.e. at every one minute interval. PARAMICS classifies drivers in its model into two groups, namely familiar and unfamiliar drivers. It assumes that only familiar drivers are incorporated with full knowledge of the entire network and are able to find alternative routes to avoid congestion. To do this, the route cost tables of the familiar drivers are updated at every time interval specified so that they can react to the congestion.

3.4.3.2 Contraflow Operation Modeling

PARAMICS does not support the direct emulation of contraflow operation. The traffic on the links is not reversible. To imitate the contraflow operation, shadow lanes are created in the opposite direction of the candidate links. Due to this reason, prior to the simulation, the candidate links need to be identified. A sample link, (i, j) shown in Figure 3.5 demonstrates the shadow lanes concept adopted to simulate the contraflow operation. Let the candidate lanes for contraflow be lane 1 and lane 2 of link (j, i) respectively. Two shadow lanes, i.e. lane A and lane B of link (i, j) are created to represent these candidate lanes. Lane A is the shadow lane of lane 2 while lane B is the shadow lane of lane 1. If the travel direction of lane 2 of link (j, i) is to be reversed, lane 2 will be closed to traffic and lane A of link (i, j) will be opened. The same also goes to lane 1 and lane B. Thus, lane 2 and lane A work as a pair to imitate the contraflow operation in which both of them are not allowed to be opened simultaneously. Certainly, in the real network layout, the shadow lanes do not exist.
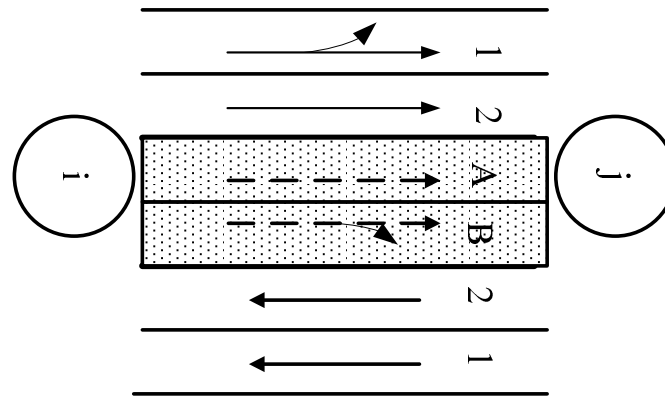
Figure 3.5 Shadow lanes and lane logic in PARAMICS

As the problem is lane-based, the lane choice rule and setting in PARAMICS is crucial especially when the shadow lanes are dealt with. In PARAMICS the lane choice logic is modeled using the signposting/hazard model. The signposting defines any hazard condition in the network and vehicles will react according to this information. Hazards in the network can be any of the following conditions: right turn or left turn at the intersection, incidents, and lane drop or lane closure. On top of this, the lane change model is prioritized into two levels, namely urgent and non-urgent. An urgent lane change is executed if the vehicle finds itself outside its target range of lanes while a non-urgent lane change is carried out to ensure that the total demand is spread over the available road space to prevent congestion building up. This understanding is crucial as how the vehicles will deal with shadow lanes should be investigated. In other words, when the shadow lanes are opened for the vehicles, they will have equal opportunity to be used by the vehicles due to the equally spread characteristic in the lane changing model. Nonetheless, extra care is taken for these lanes at the intersection. Figure 3.5 shows that originally there are two lanes for link (i, j) in which lane 1 allows straight and left turn while lane 2 allows straight and right turn. When two shadow lanes are added to this link, lane A and lane B allow straight movements, but

only lane B is allowed for right turn. This is to avoid vehicles crossing each other and behave odd in the network. Since the driver behavior during contraflow operation is vague in literature, i.e. whether the drivers will be more cautious and slow down their speed is not known in reality, it is neglected in this study. Another important issue related to the contraflow shadow lanes is the design of the lane-marking. As this could affect the overall network performance, proper methodology may be adopted for the design. Interested readers may refer Wong and Wong (2002) and Cantarella et al. (2006).

The contraflow operation modeling is implemented by using the extra API function provided. Through Dynamic Link Library (DLL) plug in, the schedule is written into PARAMICS prior to the simulation through the extending and setting functions. The objective function value is collected from PARAMICS using the getting functions. The interaction of PARAMICS and GA is smooth governed by these API functions.

3.4.3.3 Results

The parameters setting in GA are such as follows: $B = 20$, $p_c = 0.6$, $p_{mutate} = 0.03$, maximum generation= 40; $\bar{T}_{ij} = 30$ minutes and $\hat{T}_{ij} = 60$ minutes where $((i,j) \in A_2)$. These values are assumed values that chosen arbitrarily. The authority/engineers may determine according to their engineering judgment. In fact, the methodology allows any value of $\bar{T}_{ij}$ and $\hat{T}_{ij}$. The outcome of the proposed methodology is shown in Figure 3.6. It can be observed that the total travel time is 382 veh-hr at the 20th iteration and remains the same for the remaining 20 generations. This is about 38% savings in travel time if compared to the initial solution. It is unfortunate that in this case study, the optimal solution could not be proved

because the problem is an NP-hard problem. Also it is not possible to enumerate all the feasible solutions and then show the optimal one because the solutions increased exponentially with the problem size. Nevertheless, it is shown that the final solution is better than the initial solution. The corresponding schedule is shown in Table 3.1. The duration of the contraflow operation is short because the simulation period adopted in this study is only one hour and the minimum duration specified is 30 minutes. This should not impose a restriction to the methodology proposed because the methodology allows the specification of any duration with a longer simulation period.
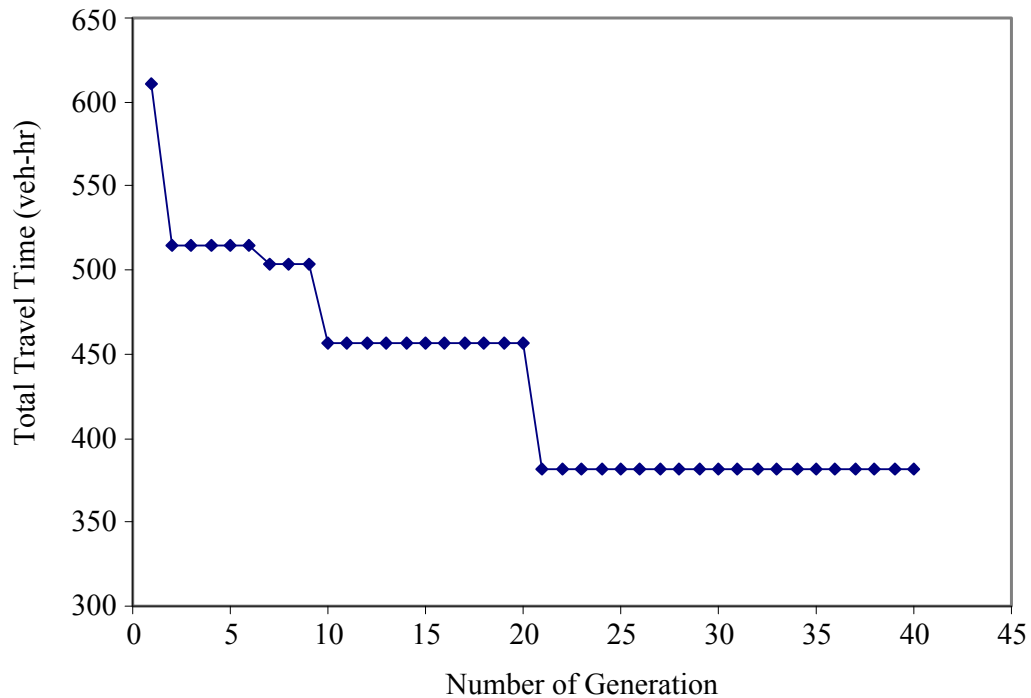
Figure 3.6 Convergent trend of GA

A benchmark case of do-nothing is carried out to facilitate the comparison. For the do-nothing case, the total travel time measured for the study area is 1127 veh-hr. Without the contraflow, it is observed that link 1,5,15,17,13 are very congested. Bottlenecks occur at the node/junction connecting these links. If both results are compared, it is observed that contraflow operation brings an improvement of 66% of

travel time savings to the drivers using the network. This is because the capacity of the roadways has been increased which reduces the waiting time of the drivers at the bottleneck locations. However, it is observed that some other roads/links which are not within the study area may become more congested with the implementation of contraflow lanes, indicating changes in drivers' route choice decision. Some of the drivers have diverted to use other roads (with the dynamic feedback ability set in PARAMICS).

Table 3.1 Contraflow schedule for individual candidate links

Link	Start Time (min)	End Time (min)	Duration (min)
2	10	57	47
4	4	40	36
6	7	37	30
8	7	37	30
10	4	40	36
12	11	45	34
14	7	58	51
16	5	39	34
18	21	51	30

3.4.3.4 Sensitivity Analysis

The sensitivity analysis is carried out to test the robustness of the proposed methodology. The sections are divided into two sub-sections, namely the network-based and algorithm-based sensitivity analysis. The network-based sensitivity analysis aims to test the effect of varying the input of the network on the outcome of the

strategy, such as OD demand, degree of driver's familiarity and minimum contraflow duration. For the algorithm-based sensitivity analysis, the values of population size, maximum generation, crossover probability and mutation probability are varied. This perhaps can shed some light on the performance of the strategy. In view of the computational time limitations, the population size of the algorithm is fixed as four for testing, and the others remain constant as per the benchmark cases. This is allowable, since in these sections, the focus is not to investigate the impact of the population size but others.

OD demand

Another 2 sets of OD demand which are 10% more and 10% lesser than the default demand are investigated respectively. Figure 3.7 shows that the result of the problem is affected by the OD demand input. It can be seen that if the OD demand is higher, the algorithm stops to a higher objective functions value and vice versa. It is observed that the stopping trend is also affected. For a lower demand value, the algorithm stops faster than the higher demand value. This means that both the optimal value and the stopping trend are affected by the OD demand. For the scheduling solution, there is no specific trend observed. It is straightforward to explain this phenomenon. With higher demand, the delay suffered by the drivers is also higher and vice versa.
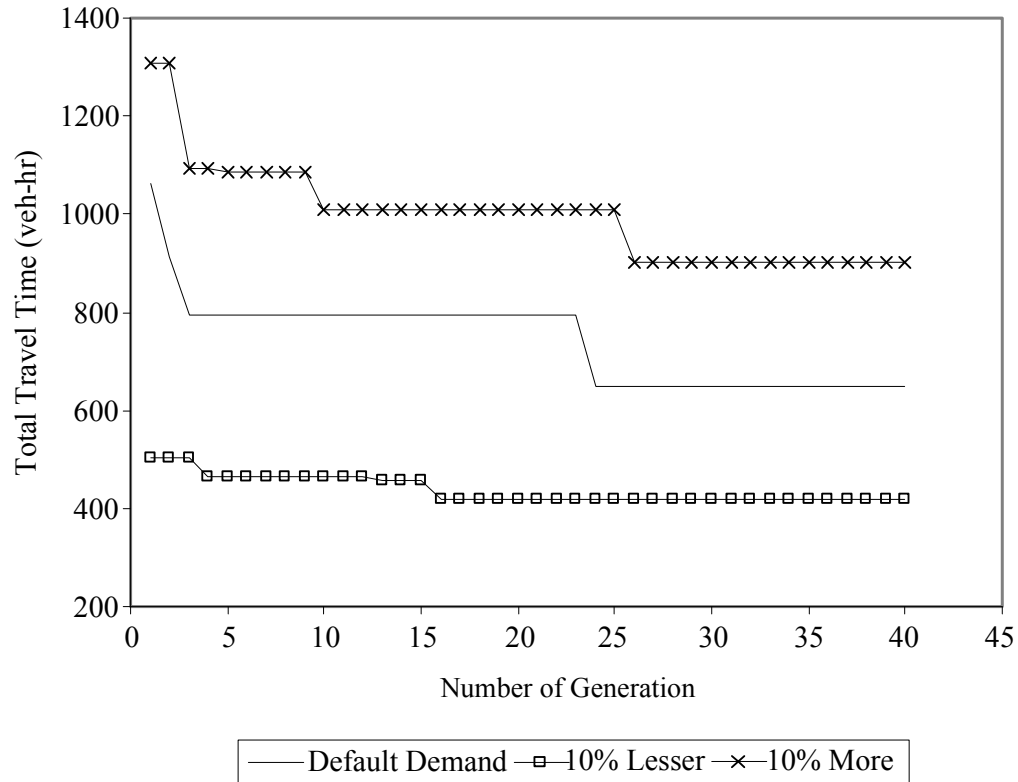
Figure 3.7 Sensitivity analysis of OD demand with population size of 4

Percentage of familiar drivers

In the result presented in previous section, the percentage of familiar drivers is set at 80%. In this section, different percentages are tested (50%, 60%, and 70%). Figure 3.8 shows the result. It can be observed that a slight change in the percentage of familiarity does not affect the results significantly. For example, the results obtained by setting 70% and 80% of drivers' familiarity in the network are similar. However, if the comparison is made between the setting of 80% and 50% familiarity, a significant different in the outcome is noted. This means that the strategy proposed is actually affected by this input. Generally, a lower percentage of familiar drivers yield a lower optimal solution. This implies that the study area assigned is more frequently used by the familiar drivers. When the percentage is low, the number of drivers using the area is fewer and contributes to a lower objective function value.

81
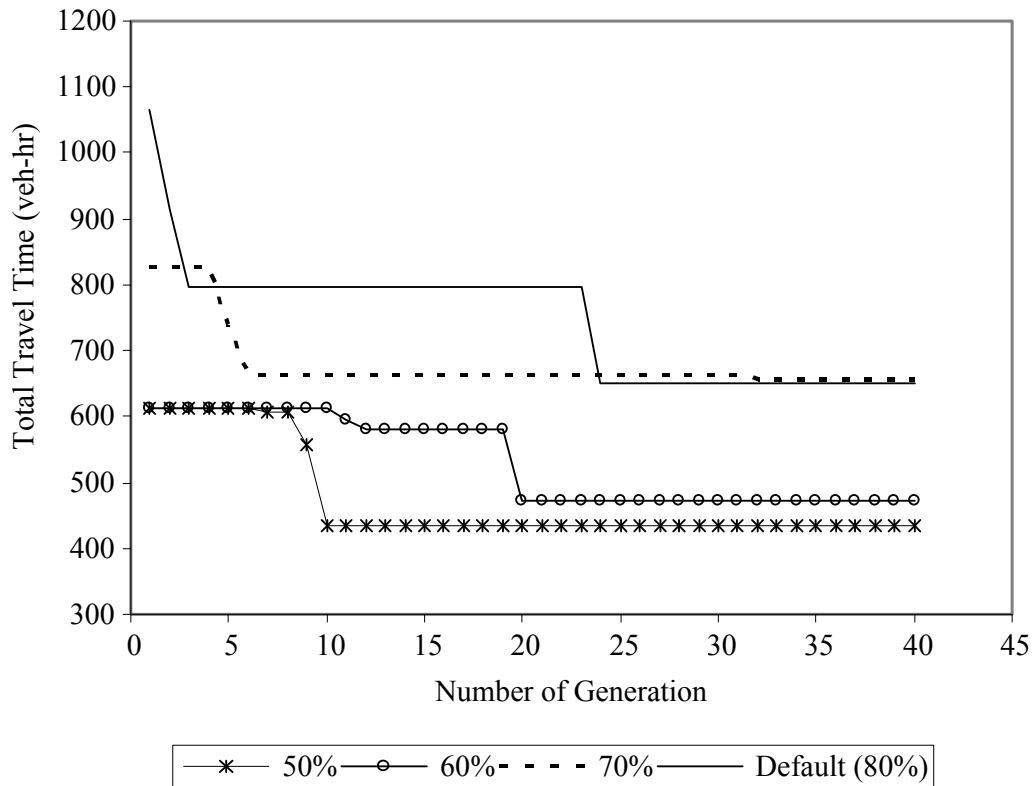
Figure 3.8 Sensitivity analysis of drivers' familiarity with population size of 4

Minimum duration of contraflow

Compared to the minimum duration set for the benchmark case, i.e. 30 minutes, a longer duration of 50 minutes is tested. The results show that the optimal objective function obtained for the latter case is 788.13 veh-hr, which is 17.5% more than the benchmark case. Drivers encounter more delay in the latter case compared to the benchmark case. By setting a longer minimum duration, it means that the constraint is imposed longer and this affects the performance of the solution algorithm.

Population size

The population size determines the number of candidate solutions during iterations. From Figure 3.9, it is observed that when the population size increases, the

solution algorithm converges to a lower value and vice versa. For a larger population size, a larger number of feasible solutions are provided in iterations. This provides higher possibility to obtain a lower optimal value. With higher population size (size of 20), the algorithm may converge faster and to a better solution if compared to a population size of 10. Although the result of population size of four has a fast convergence, it is observed that it provides the highest optimal value among all the solutions. This explains that a population size of four maybe too few to generate a good result. A population size of four is chosen herein to serve the purpose of comparison only due to the computational time limitation. However, for the results presented for benchmark case, the population size used is 20.
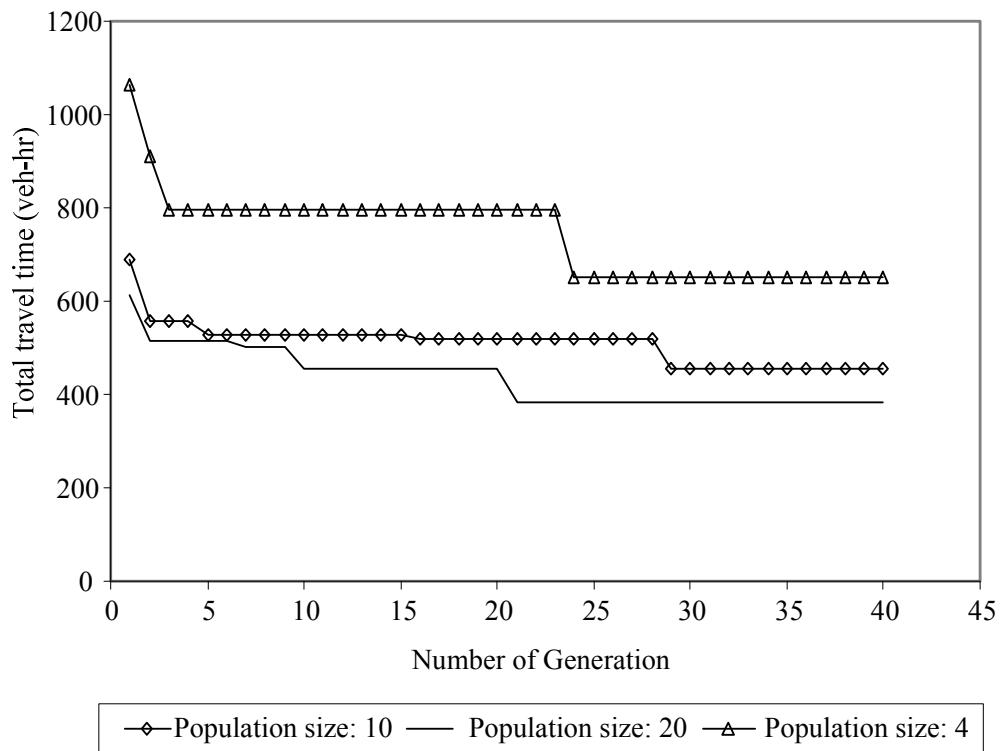
Figure 3.9 Sensitivity analysis of population size

Probability of crossover

Figure 3.10 shows the results of the sensitivity analysis for different percentage of crossover probability. It is observed that the optimal value is affected by the probability value. For a higher probability of crossover, higher chance to obtain a lower objective function value could be obtained. In addition, the algorithm converges faster for a higher probability of crossover. From the figure, crossover of 60%, 40% and 20% achieve the optimal value of 476 (veh-hr), 650veh-hr and 723 veh-hr respectively.
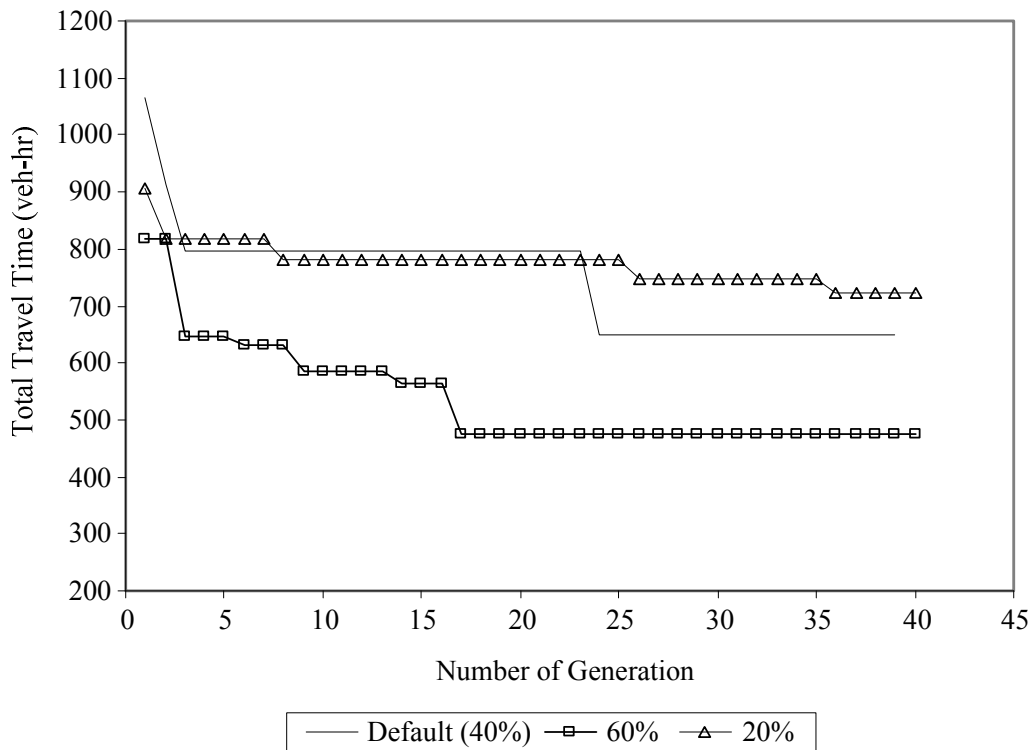
Figure 3.10 Sensitivity analysis of crossover probability with population size of 4

Probability of mutation

Figure 3.11 shows that the algorithm is sensitive to the mutation probability. Although the convergence speed is not much varying, it is observed that the objective

function value obtained is about 16% different. With a higher probability, a better result is obtained.

<u>Maximum generation</u>

A sensitivity test carried out on the maximum generation parameter reveals that there are some effects on the objective function's value for the optimal function value. If more generations are considered, a lower objective function's value can be obtained.
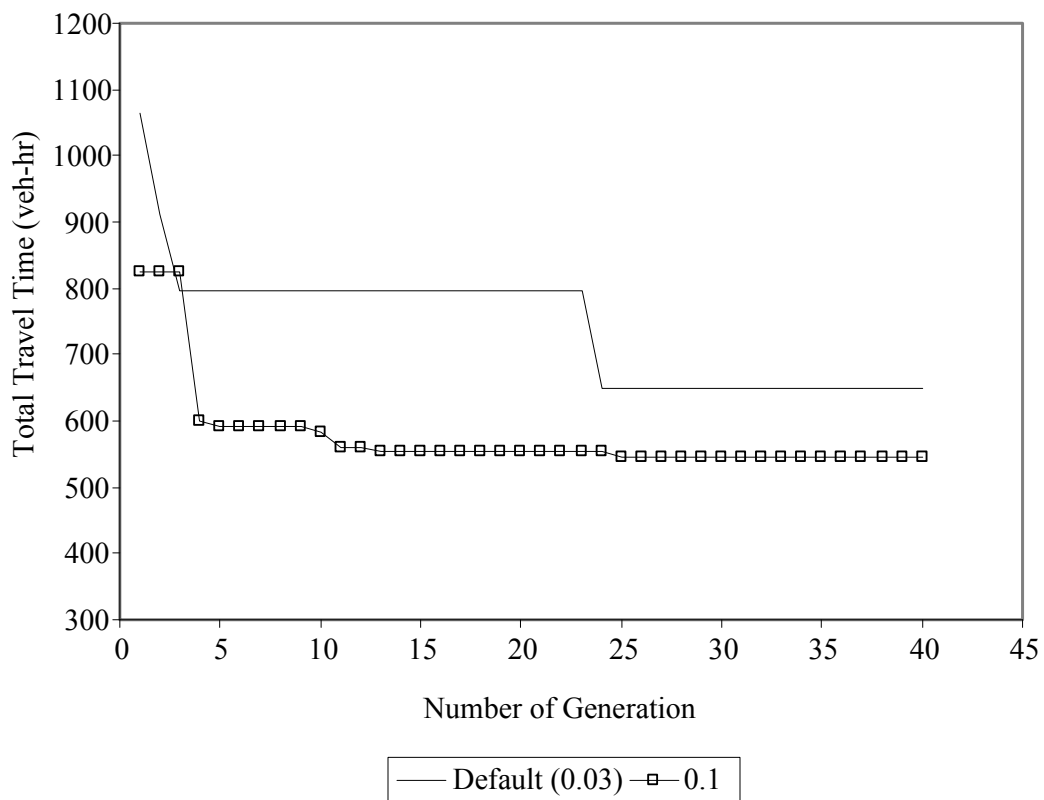
Figure 3.11 Sensitivity analysis of mutation probability with population size of 4

3.5 Optimal Lane Configuration Problem (OCLCP)

The same methodology could be applied to solve OCLCP. A bilevel programming model is proposed to solve the problem. In addition, the hybrid GA-

PARAMICS is also adopted as the solution algorithm to the proposed model. The upper level of the bilevel model for the OCLCP is different from the OCSP because it is a binary integer programming model.

3.5.1 Bilevel programming model

To find the optimal lane configuration solution, the binary (0-1) variable is defined as:

$$c_{ij}^k = \begin{cases} 1, \text{ if lane } k \text{ of link } (i,j) \text{ is chosen to reverse} \\ \quad \text{travel direction} \\ 0, \text{otherwise} \end{cases}, (i,j) \in A_2 \text{ and } k = 1,2,\cdots,l_{ij} \quad (3.14)$$

For the sake of presentation, let $\mathbf{c}$ be the row vector of all the binary decision variables, namely, $\mathbf{c} = (\cdots, c_{ij}^k, \cdots)$ where $k = 1,2,\cdots,l_{ij}$ and link $(i,j) \in A_2$. It is assumed that the contraflow operation is applied for the entire simulation period and started from the beginning of the simulation. In consideration for traffic safety and operational feasibility, the sequence of reversal in travel direction in a candidate link must be consecutively from the median lane to the kerbside lane. This constraint can be mathematically described by

$$c_{ij}^1 \le c_{ij}^2 \le \cdots \le c_{ij}^{l_{ij}}, (i,j) \in A_2 \qquad (3.15)$$

In the situation that both links (with opposing travel directions) connected by a same node pair are candidate links, only one of them will be allowed to reverse its travel direction. Such a constraint can be expressed as the constraint shown in eqn. (3.16).

$$c_{ij}^{l_{ij}-k} + c_{ji}^{l_{ji}-k} \le 1, (i,j) \in \left(A_2 \cap \tilde{A}_2\right), k = 0,\cdots,\min\left(l_{ij},l_{ji}\right)-1 \qquad (3.16)$$

Any $\mathbf{c}$ satisfying constraints (3.15)-(3.16) is referred as a feasible contraflow lane configuration solution. Given a feasible contraflow lane configuration solution $\mathbf{c}$, the

network's topological structure has changed. Thus, the total travel time experienced by drivers in the study area is expressed as follows:

$$F(\mathbf{c}) = \sum_{n=1}^{N(\mathbf{c},T,A_1)} \left(t_n^{out} - t_n^{in} \right) \tag{3.17}$$

where $N(\mathbf{c},T,A_1)$ is the total number of vehicles passing through the study made up by links A_1 during time $[0,T]$; t_n^{out} is the time when the nth vehicle leaving the area and t_n^{in} is the time when the nth vehicle entering the area. The difference $(t_n^{out} - t_n^{in})$ is the travel time of nth vehicle in the study area. Note that eqn. (3.17) is identical to eqn. (3.2) in which $\mathbf{r}$ in this problem is expressed by $\mathbf{c}$.

The full formulation of the bilevel programming model for the OCLCP is therefore:

$$\min\ F(\mathbf{c}) = \sum_{n=1}^{N(\mathbf{c},T,A_1)} \left(t_n^{out} - t_n^{in} \right) \tag{3.18}$$

subject to

$$c_{ij}^1 \le c_{ij}^2 \le \cdots \le c_{ij}^{l_{ij}}, (i,j) \in A_2 \tag{3.19}$$

$$c_{ij}^{l_{ij}-k} + c_{ji}^{l_{ji}-k} \le 1, (i,j) \in \left(A_2 \cap \tilde{A}_2 \right), k = 0,\cdots,\min\left(l_{ij}, l_{ji} \right) + 1 \tag{3.20}$$

$$c_{ij}^k = 0 \text{ or } 1, (i,j) \in A_2, k = 1,\cdots,l_{ij} \tag{3.21}$$

where value of the objective function in eqn. (3.18) is obtained by executing a microscopic traffic simulation model schematically illustrated in Figure 3.2 for any given $\mathbf{c}$. In terms of terminology in the subject of bilevel programming, the binary integer programming problem expressed by eqns. (3.18)-(3.21) is referred to as the upper level problem, and the microscopic traffic simulation model depicted in Figure 3.2 is called the lower level problem. Constraints (3.19)-(3.20) is operational constraint

mentioned in the general bilevel model presented in section 3.2, and constraint (3.21) is definitional constraint.

3.5.2 Solution Algorithm

3.5.2.1 Encoding and Decoding of Strings

A string is encoded by the row vector of all the binary decision variables, namely, $\mathbf{c} = \left(\cdots, c_{ij}^k, \cdots \right)$, where $(i, j) \in A_2$ and $k = 1, 2, \cdots, l_{ij}$. Note that each gene of the string, c_{ij}^k, only takes values of 0 or 1. The length of a string is equal to the total number of lanes in all the candidate links, i.e.,

$$L_{oclcp} = \sum_{(i,j) \in A_2} l_{ij} \qquad (3.22)$$

The objective function in eqn. (3.18) is the fitness function in GA. Each location of the gene represents a lane of the candidate link in the network. Hence, in the decoding process, the associated outcome, whether reverse or not, can be obtained directly from the string.

3.5.2.2 String Repairing Procedure

This procedure is embedded in the GA-PARAMICS framework in order to repair the infeasible solutions that violate the constraints as indicated by (3.19) and (3.20).

Step 1: (The first pass scan) Check each gene of the string $\mathbf{c}$, a string of $\cdots, c_{ij}^k, \cdots$ with length L_{oclcp}, starting from its left hand side to identify those genes that represent all the lanes of a candidate link, say (i, j), which forms a substring,

$c_{ij}^1 c_{ij}^2 \cdots c_{ij}^{l_{ij}}$, of the string. For the substring, $c_{ij}^1 c_{ij}^2 \cdots c_{ij}^{l_{ij}}$, let $\hat{l}_{ij}$ be the number of genes with the value of one, namely:

$$\hat{l}_{ij} = \sum_{k=1}^{l_{ij}} c_{ij}^k \tag{3.23}$$

The following operation is implemented for the string, with k numbered from the kerbside lane to the median lane:

$$c_{ij}^k = \begin{cases} 0, & k = 1, 2, \cdots, l_{ij} - \hat{l}_{ij} \\ 1, & k = l_{ij} - \hat{l}_{ij} + 1, l_{ij} - \hat{l}_{ij} + 2, \cdots, l_{ij} \end{cases} \tag{3.24}$$

Step 2: (The second pass scan) Let $\hat{\mathbf{c}} = \left(\cdots, \hat{c}_{ij}^k, \cdots \right)$ be the string generated from Step 1.

For each link $(i, j) \in A_2 \cap \overline{A}_2$, perform manipulation on this string as follows.

Case 1: The substring, $\hat{c}_{ij}^1 \hat{c}_{ij}^2 \cdots \hat{c}_{ij}^{l_{ij}}$, associated with link (i, j), precedes the substring $\hat{c}_{ji}^1 \hat{c}_{ji}^2 \cdots \hat{c}_{ji}^{l_{ji}}$, related to link (j, i), in the string $\cdots, \hat{c}_{ij}^k, \cdots$. If there is at least one gene with value of 1 in the substring $\hat{c}_{ij}^1 \hat{c}_{ij}^2 \cdots \hat{c}_{ij}^{l_{ij}}$, the value of all genes in another substring $\hat{c}_{ji}^1 \hat{c}_{ji}^2 \cdots \hat{c}_{ji}^{l_{ji}}$ is set to zero, i.e.,

$$\hat{c}_{ji}^k = 0, k = 1, 2, \cdots, l_{ji} \ \ \text{if} \ \sum_{k=1}^{l_{ij}} \hat{c}_{ij}^k > 0 \tag{3.25}$$

Case 2: The substring, $\hat{c}_{ij}^1 \hat{c}_{ij}^2 \cdots \hat{c}_{ij}^{l_{ij}}$, associated with link (i, j) follows the substring $, \hat{c}_{ji}^1 \hat{c}_{ji}^2 \cdots \hat{c}_{ji}^{l_{ji}}$, related to link (j, i) in the string, $\cdots, \hat{c}_{ij}^k, \cdots$. If there is at least one gene with value of 1 in the substring $\hat{c}_{ji}^1 \hat{c}_{ji}^2 \cdots \hat{c}_{ji}^{l_{ji}}$, the value of all genes in substring $\hat{c}_{ij}^1 \hat{c}_{ij}^2 \cdots \hat{c}_{ij}^{l_{ij}}$ is set to zero, i.e.,

$$\hat{c}_{ij}^k = 0, k = 1, 2, \cdots, l_{ij} \ \ \text{if} \ \sum_{k=1}^{l_{ji}} \hat{c}_{ji}^k > 0 \tag{3.26}$$

89

The above Case 1 and Case 2 are mutually exclusive. If both links (i,j) and (j,i) has at least a contraflow lane (i.e., $\hat{c}_{ij}^k > 0$ and $\hat{c}_{ji}^k > 0$ for any k), the choice of Case 1 or Case 2 is done randomly. Let λ be an indicator to determine the choice of the case, where $0 \leq \lambda \leq 1$. If $\lambda \leq 0.5$, Case 1 is chosen; otherwise Case 2 is chosen. It is easy to verify that the string repairing procedure yields a feasible contraflow lane configuration solution, i.e., a solution fulfills constraints (3.19)-(3.20) of the bilevel programming model.

The idea behind the string repairing procedure is demonstrated using an example comprising four links and three nodes, as shown in Figure 3.12.
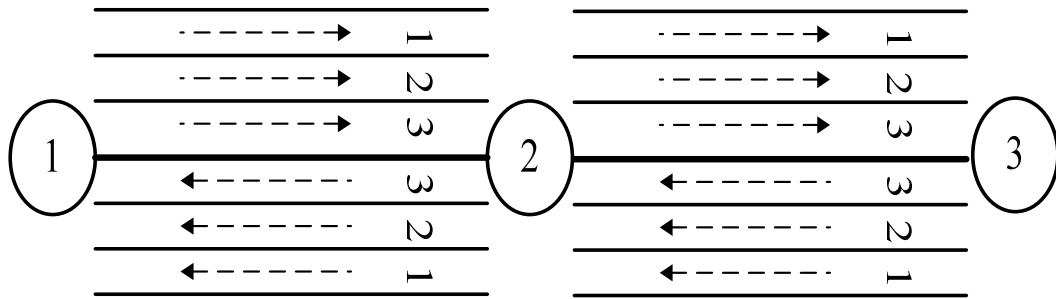
Figure 3.12 An example for the string repairing procedure

Assume that links, (1,2), (2,1) and (2,3), are the candidates links, i.e., $A_2 = \{(1,2),(2,1),(2,3)\}$, and each link has three lanes. The string is coded by

$$\overbrace{c_{12}^1 c_{12}^2 c_{12}^3}^{\text{link (1,2)}} \overbrace{c_{23}^1 c_{23}^2 c_{23}^3}^{\text{link (2,3)}} \overbrace{c_{21}^1 c_{21}^2 c_{21}^3}^{\text{link (2,1)}} \tag{3.27}$$

Let the binary string, $\overbrace{101}^{\text{link (1,2)}} \overbrace{100}^{\text{link (2,3)}} \overbrace{001}^{\text{link (2,1)}}$, be a string generated by the GA. It can be seen that substrings 101 and 100 associated with links (1,2) and (2,3) violate constraint (3.19), and that the string does not fulfill constraint (3.20) in light of the relevant substrings associated with links (1,2) and (2,1). Therefore, the string is an infeasible solution of the bilevel programming model and it must be repaired. The first pass of

the above string repairing procedure yields a new string, 011001001. After the second pass for the new string, the string repairing procedure outputs the final string, 011001000, which is a feasible solution of the bilevel programming model.

Given a string that leads to a feasible contraflow lane configuration solution, the relevant network topological data can be obtained. After inputting the network data, the microscopic traffic simulation model shown in Figure 3.2 can yield the fitness function value associated with the string. The GA incorporating both the microscopic traffic simulation model and the string repairing procedure is presented follow. This is exactly the same as section 3.4.2.3 but with the schedule being replaced by lane configuration. It is presented herein for the sake of completeness.

3.5.2.3 A Step-By-Step GA

Step 0. (Initialization) Randomly generate a population of B strings.

Step 1. (String repairing) For each infeasible string in the population generated in Step 0, apply the string repairing procedure. The infeasible string in the population is then replaced by a feasible one

Step 2. (Calculation of the fitness function) For each feasible string $\mathbf{c}$ in the population, the value of fitness function $F(\mathbf{c})$ defined is evaluated by implementing the microscopic traffic simulation model shown in Figure 3.2.

Step 3. (Generation of a new population) Repeat the following four sub-steps until the new population is completed.

Step 3.1. (Selection) According to the fitness functions value evaluated in Step 2, use the rank selection method to choose two parent strings from the population.

Step 3.2. (Crossover) With a crossover probability, denoted by p_c, apply the one point cross over method to form a new offspring. If no crossover is performed, the offspring is the exact copy of the parents.

Step 3.3. (Mutation) With a mutation probability, denoted by p_m, mutate the new offspring at selected positions in the string.

Step 3.4. (Repairing) If the offspring results in an infeasible solution of the upper level problem, repair it by executing the string repairing procedure.

Step 4. (Stopping criterion). If a stopping criterion is fulfilled, then terminate, and output the best solution from the population. Otherwise, go to Step 2.

3.5.3 Numerical Results

The network adopted to test the proposed methodology is shown in Figure 3.4. The network coding is similar to those illustrated in section 3.4.3.1, except the demand loading is different. In addition, the contraflow operation is emulated according to the method described in section 3.4.3.2. The parameters are such as the population size *B*, the crossover probability p_c and the mutation probability p_m in GA. In this case study, values of these parameters were set as follows: $B = 12$, $p_c = 0.6$ and $p_m = 0.03$.

Figure 3.13 plots the change of the total travel time over the study area with respect to the generation number (taken from the best string) when applying the proposed GA for the case study. The average total travel time of these runs is calculated and taken as the objective function $F(\mathbf{c})$ of GA. Figure 3.13 shows that GA reduces the initial solution from 168 vehicle-hrs of total travel time to 120.3 vehicle-hrs, which is about 28.5% reduction. Every generation has objective function value that is at least equal or better than the previous generations. Finally, the objective

function converges after eight generations. The associated contraflow lane configuration solution is shown in Table 3.2. According to this table, the travel direction of thirteen lanes in eight candidate links should be reversed during peak hour to cater for the high traffic demand in the opposite directions and to reduce the total travel time in the study area. This is similar to the scheduling problem where the optimal solution cannot be ascertained. It is impossible to enumerate manually each and every possible solution in order to prove that the solution given by GA is optimal. Note that the OD demand trip value is set differently with those presented in the OCSP. Hence, the total travel time obtained is not comparable.
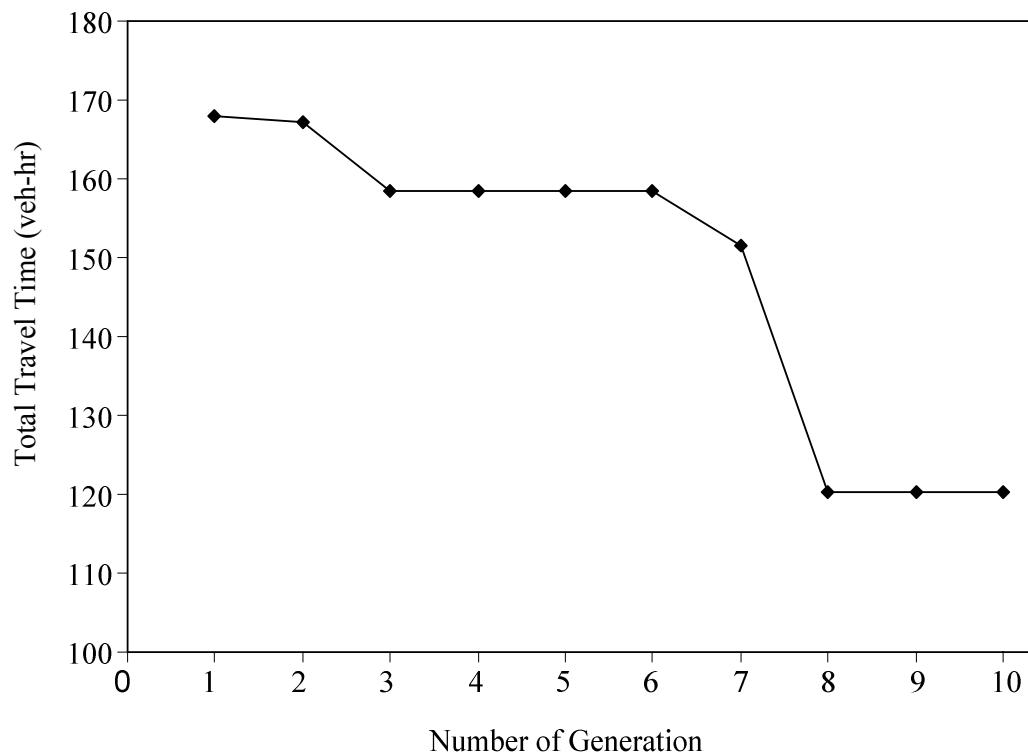
Figure 3.13 Convergent trend of the GA for OCLCP

Table 3.2 Optimal contraflow lane configuration solution

Link Number	Lanes to be reversed
2	Lane 3
4	Lane 2, Lane 3
6	Lane 3
8	NA
10	Lane 3
12	Lane 2, Lane 3
14	Lane 2, Lane 3
16	Lane 2, Lane 3
18	Lane 2, Lane 3

3.5.3.1 Sensitivity Analysis

Sensitivity analysis test on the parameters setting in GA, namely population size B, crossover probability p_c and mutation probability p_m is adopted to test the robustness of the proposed methodology. Figure 3.14 illustrates the performance of GA with different population sizes. It shows that the optimal solution may be affected by the choice of population size. If a small population size is used, the GA may end up with a higher objective function value. However, unlike OCSP, tests on the crossover probability, mutation probability and total number of generation have shown that the objective function value is not sensitive to these parameters values. The convergence trends and the objective function values are almost similar if the values of these parameters are varied.
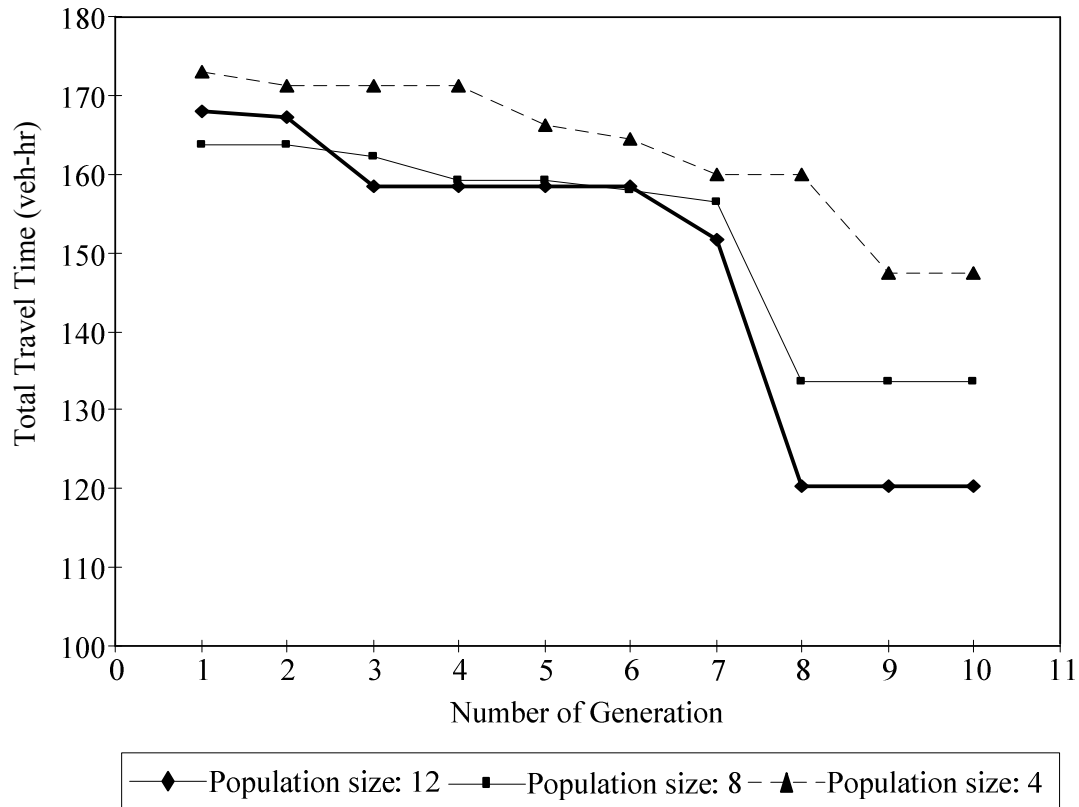
Figure 3.14 Performance of the Genetic Algorithm with different population size

3.6 Some Implementation Issues

3.6.1 Computational Limitations

A drawback in the proposed solution algorithm is the long computational time. This is mainly due to the extensive number of replications required by the methodology. For example, a large number of population size and maximum number of generation are required to produce reliable result. Using an Intel Core 2 Processor 2.4 GHz with RAM speed 3 GHz, each simulation run with PARAMICS takes about 2 minutes. To obtain one set of data with 20 strings and 40 maximum generations, the computational time is about 12 hours. However, since the proposed methodology is meant for offline decision making process during planning stage, the long

computational time should not be an obstacle to the application of the proposed methodology. Alternatively, parallel or distributed computing technology can be employed (see, Lu et al. 2005b).

3.6.2 Practical Implementation Issues

The practical implementation of contraflow lanes may cause safety and operational problem. Changing the travel direction of the roadways can cause confusion to drivers. Hence, there is a need for careful planning prior to the implementation. In the presented case study, the homogeneous issue is crucial to avoid drivers' confusion. For a paired link, for example candidate links 2 and 6 shown in Figure 3.4, which is physically connected, must have same schedule and lane configuration. Abrupt change of the schedule and lane configuration for a continuous road segment may confuse the drivers that would lead to the occurrence of accidents. To do this, only little amendment on the bilevel model coding needs to be carried out. It can be done in two ways, either adding extra constraints on the model or to code these paired link as one.

This study assumes that any necessary preparation is accomplished before the start of the contraflow operation. Movable barrier or delineator post can be used to separate the contraflow lane from the normal lane in the case where median exists. Electronic devices such as variable message signs (VMS) and traffic lights can also be adopted to inform and advise drivers about the implementation. The choice on the types of suitable methods to be used depends on the network layout. For example, zipper lane is more suitable for expressway, variable message sign is more suitable for arterial road and traffic light is more suitable for tunnel. Whatever method chosen, drivers need to be informed and warned before they enter the roadway with contraflow

lanes and this can be done by disseminating this type of information through radio, roadside board or VMS. If the manual methods, i.e. the zipper lane and moveable barrier are chosen for the contraflow operation, it is suggested that the schedule of the implementation should be the same for all the roads in the study area. This could reduce the work load and the inconvenience of the workers who prepare and clean the site. This preparation time issue can be easily reflected in the objective function value by adding an additional term or constant to the objective function. In fact, a more straightforward way to model this in microscopic traffic simulation models is to set two signals on both entrances of the reversible lanes, with which the contraflow operations are simply manned by the signal settings to imitate the clearance times.

Traffic safety is another important issue in contraflow operation. Perhaps, signboards should be placed at the starting road segment to alert drivers about the operation. Information about the operation should also be disseminated widely to drivers through VMS and radio. Delineator posts should be placed along the roadway to separate the contraflow lanes with the normal lanes. Policemen or traffic personnel could be placed at the start point and end point of contraflow lane to direct traffic. All these precaution measures could enhance the safety when the operation is applied.

3.7 Summary

This chapter systematically initiated a novel study on the contraflow operations. Two important issues arising from the operation during the offline planning stage have been addressed in detail, namely OCSP and OCLCP. A bilevel programming model is proposed to solve the problems. The upper level problem is an integer programming model, which minimizes the total travel time during peak hour over a study area, with the schedule or lane configuration of the contraflow operations as the decision

variables. The lower level problem is a microscopic traffic simulation model which reflects the traffic flow conditions resulting from drivers responding to a feasible contraflow operation solutions. The proposed formulation is then solved by a hybrid GA-PARAMICS framework. The results from the illustrative case study show that the proposed methodology can be used to solve the contraflow operation problems. A sensitivity analysis test on the proposed methodology for OCSP is performed. It is shown that the methodology is sensitive to the setting of the parameters used in the case study. Some important issues are also highlighted especially related to the computational efficiency and practical implementation issues.

CHAPTER 4 ATIS-BASED EXPRESSWAY-ARTERIAL CORRIDOR SYSTEM TRAFFIC CONTROL OPERATIONS

4.1 Introduction

This chapter proposes a novel online dynamic traffic control strategy on the expressway-arterial corridor system which does not have ramp meter. This traffic management strategy employs ATIS tools such as VMS and radio to disseminate traffic flow information to drivers to affect their route choice decisions. It is a descriptive measure, in which a detail route guidance system is not provided. It comprises three components, namely the Expressway Mainline Control (EMC) mechanism, the Off-ramp Control (OffC) mechanism and the On-ramp Control (OnC) mechanism. The EMC mechanism is the major component in which traffic condition in the expressway is dynamically monitored. It assumes that, the cumulative number of vehicles that arrive and leave the bottleneck section can be obtained through the detectors installed on the expressway. The EMAS system (LTA 2006) in Singapore provides a means to obtain the traffic counts. If the average difference of these cumulative numbers is greater than a pre-determined threshold, the control strategy will be triggered and diversion advice will be disseminated to the drivers. The OffC mechanism will be activated if the vicinity of an off-ramp being chosen as a diversion point reaches its capacity. If this happens, the EMC mechanism will be terminated. On the other hand, the vicinity of an on-ramp will be monitored by the OnC mechanism. If

the capacity of the on-ramp is reached, drivers will be advised not to travel into the expressway through that particular ramp.

The proposed online dynamic traffic control operation is not only applicable to a single expressway section, but can also be used to coordinate and integrate a number of expressway sections and the on- and off-ramps in order to obtain a better system performance. It can be seen that the strategy is an integrated expressway and urban corridor traffic control approach that aims to reduce total travel time. The strategy can be adopted to mitigate recurrent and non-recurrent congestion as long as the congestion is caused by a traffic bottleneck. Moreover, the strategy decides when to disseminate the diversion advice to the drivers down to any time interval specified by the traffic engineers, for example at every one minute interval.

One significant property of this operation is that the drivers' compliance rate to the diversion advice is implicitly integrated in the operation. It should be emphasized that only drivers themselves can decide whether to follow the advice or not. Thus, a microscopic traffic simulator is necessary to evaluate the effectiveness of the online dynamic traffic control operation. This study adopts PARAMICS, which can simulate dynamic behavior of drivers in lane changing, gap acceptance, car following and route choice. PARAMICS incorporates unique drivers' behavior such as awareness and familiarity of the network for each individual driver. This can be employed to simulate the drivers' compliance rate to the congestion information.

4.2 Urban Expressway-Arterial Corridor

Figure 4.1 schematically illustrates an urban expressway-arterial corridor without any ramp meter. It is assumed that drivers are able to obtain the traffic information before or when they are approaching to the bifurcation point (indicated by

triangle) of the network by various ATIS means such as VMS or radio. In addition, there are traffic counting devices such as loop detectors along the expressway and ramps, which dynamically provide the cumulative number of vehicles entering and leaving the expressway. In fact, other devices or techniques can also be used as long as the traffic condition on the network can be obtained dynamically.
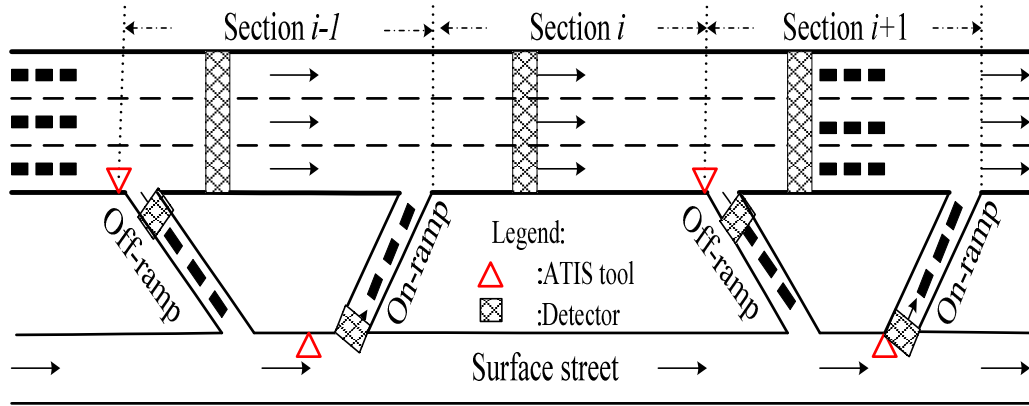
Figure 4.1 A schematic illustration of an urban expressway-arterial corridor

4.3 The Traffic Control Strategy

Assume that traffic counting devices mounted on the expressway-arterial corridor can provide the real-time cumulative arrivals and departures of vehicles at each lane in the expressway sections and the on- and off-ramps. The average number of vehicles per lane in an expressway section $i : a_i(t)$, in the off-ramp $j : a_j^{off}(t)$ and in the on-ramp $j : a_j^{on}(t)$ at time t can be calculated by:

$$a_i(t) = \frac{\sum_{k=1}^{l_i} \left[N_i^k(t) - O_i^k(t) \right]}{l_i} \tag{4.1}$$

$$a_j^{off}(t) = \frac{\sum_{k=1}^{l_j^{off}} \left[N_{j,k}^{off}(t) - O_{j,k}^{off}(t) \right]}{l_j^{off}} \tag{4.2}$$

$$a_j^{on}(t) = \frac{\sum_{k=1}^{l_j^{on}} \left[N_{j,k}^{on}(t) - O_{j,k}^{on}(t) \right]}{l_j^{on}} \tag{4.3}$$

where $N_i^k(t)$ denote the cumulative arrival of vehicles at lane k of section i at time t, $N_{j,k}^{off}(t)$ and $N_{j,k}^{on}(t)$ indicate the cumulative arrival of vehicles at lane k on the off-ramp j and on-ramp j at time t respectively. $O_i^k(t)$ indicate the cumulative departure of vehicles from lane k at section i at time t, $O_{j,k}^{off}(t)$ and $O_{i,k}^{on}(t)$ indicate the cumulative departure of vehicles from lane k of off- ramp j and on-ramp j at time t respectively. l_i, l_j^{on}, and l_j^{off} denote the total number of lanes at expressway mainline, on-ramp and off-ramp respectively. Generally speaking, the computation of the average vehicle on the corridor system is obtained by dividing the total number of vehicles existed in the network by the total number of lanes.

4.3.1 Expressway Mainline Control Mechanism (EMC)

This is the main component of the strategy. It utilizes the number of vehicles (which reflects the level of service) as a threshold to control congestion level on an expressway section with traffic bottleneck. When the propagation of the congestion due to overflow of vehicles in the expressway section is detected, i.e. the traffic density level at the expressway segment exceeds the pre-determined level denoted by Π, such information will be disseminated to the drivers at least one off-ramp upstream of the scene, so that they will have the opportunity to choose alternative routes to the destinations. The decision for diversion is up to each driver. Therefore, there is a

diversion probability associated with each driver. At the same time, OnC and OffC mechanisms will be triggered for the on- and off- ramps at upstream of the traffic bottleneck. The pre-determined level denoted by Π could be determined based on the level of service in Highway Capacity Manual (HCM 2000) or historical data of the roadway. The flowchart of the proposed EMC mechanism is depicted in Figure 4.2. According to this figure, two types of 0-1 indicator variables, $\Lambda_j^{\text{on-ramp}}(t), j \in J_i^{\text{on}}$ and $\Lambda_j^{\text{off-ramp}}(t), j \in J_i^{\text{off}}$, are used to activate the on-ramp and off-ramp control mechanisms, respectively. Note that J_i^{on} and J_i^{off} denote the set of on-ramps upstream of the traffic bottleneck section i, and the set of off-ramps upstream of the traffic bottleneck section i respectively.
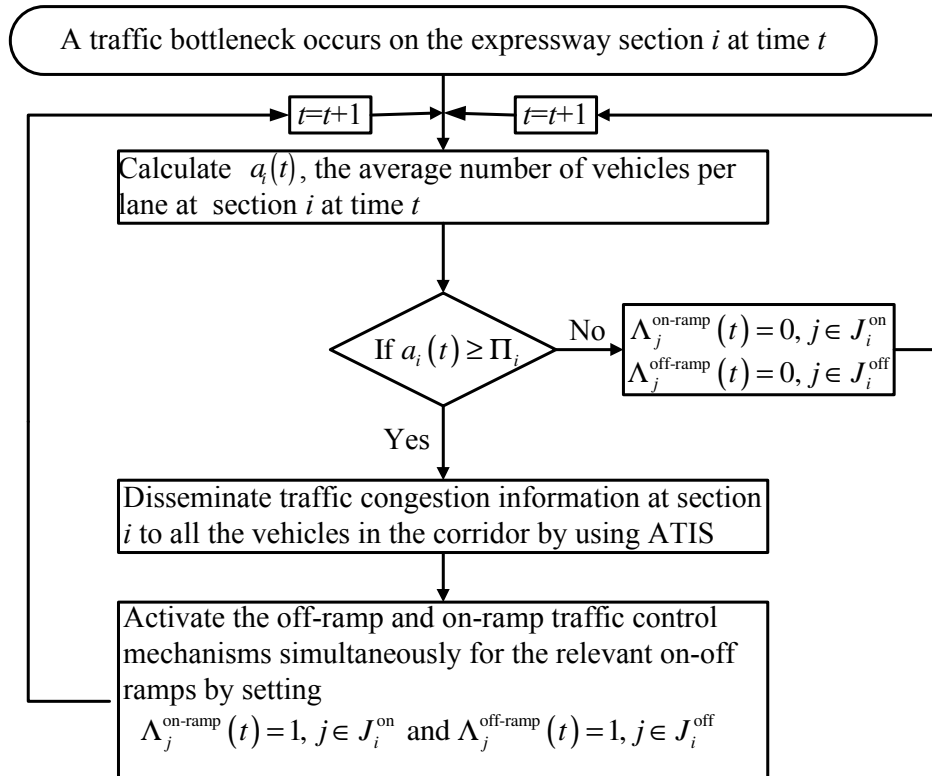
Figure 4.2 Expressway mainline control mechanism (EMC)

4.3.2 Off-Ramp Control Strategy Mechanism (OffC)

To guarantee an improvement in the corridor system performance, the traffic condition at the off-ramp immediately upstream of the traffic bottleneck needs to be monitored closely. This is because diverting vehicles from the expressway will increase the volume, and may deteriorate the traffic condition on the surface street. A strategy is required to ensure that the off-ramp and surface street are able to accommodate the diverted flow while ensuring that vehicle queue at the off-ramp does not spill back to the expressway. Hence, the OffC mechanism is incorporated to terminate the EMC mechanism if the ramp is congested. When the OffC strategy is triggered, the diversion advice will not be disseminated to the drivers. Figure 4.3 gives the flowchart of the OffC mechanism for a off-ramp $j \in J_i^{\text{off}}$. The parameter, Π_j^{off} indicated in Figure 4.3 could be determined from the level of service using the Highway Capacity Manual (HCM 2000).
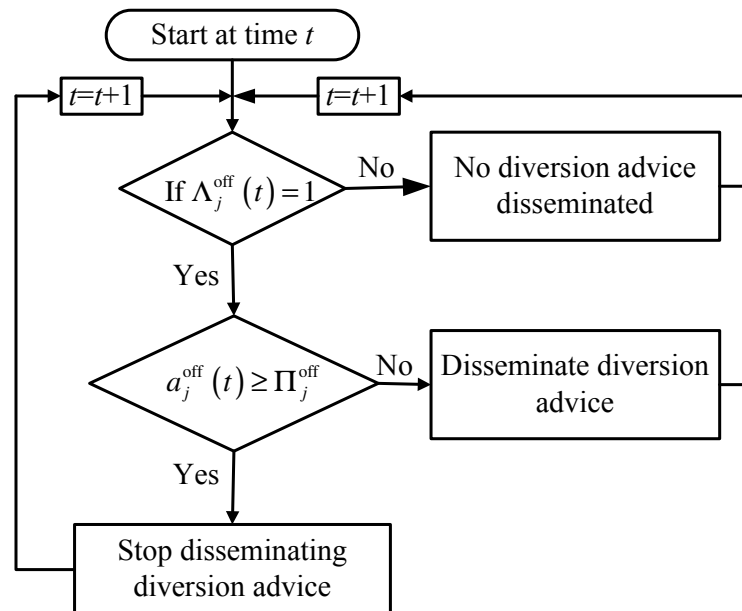
Figure 4.3 Off-ramp control mechanism (OffC)

4.3.3 On-ramp Control Mechanism (OnC)

This strategy is to ensure that drivers will be informed so that some of them will choose alternative routes when the expressway mainline and the on-ramp are congested. It is different from ramp metering operation because it does not regulate the entry of the vehicles. This strategy provides the congestion information at the bifurcation node on the network so that drivers may not choose to enter the expressway via the on-ramp and further congest the expressway. This is crucial because once a driver has entered the on-ramp, he has to travel through the congested bottleneck location on the expressway mainline segments before choosing to leave the expressway through downstream off-ramps. Again, there is a probability associated with the decision of each driver. Figure 4.4 shows the flowchart of the OnC mechanism. The parameter Π_j^{on} could be determined based on the level of service of Highway Capacity Manual (HCM 2000).
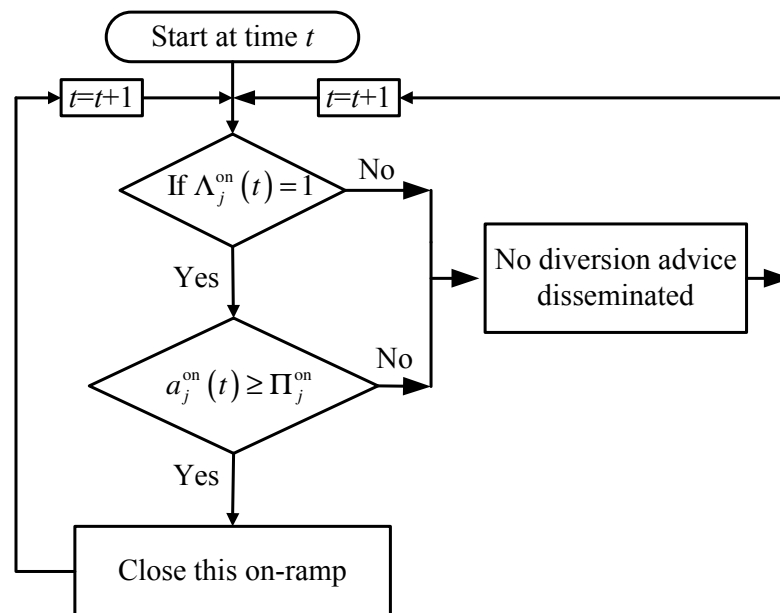
Figure 4.4 On-ramp traffic control mechanism

4.4 Evaluation Method

The proposed strategy is evaluated using a microscopic traffic simulation model, namely PARAMICS. The assignment method adopted to simulate the drivers' route choice decision is explained. Besides, a novel method on how to simulate the drivers' compliance rate is presented.

4.4.1 Mixed Dynamic Traffic Assignment in PARAMICS

Having considered the effect of ATIS in the online dynamic traffic control operation, it is reasonable to assume that there are two types of drivers: familiar and unfamiliar (see, Quadstone 2005) on the corridor. The route choice simulations for the familiar and unfamiliar driver are realized by the deterministic and stochastic traffic assignments with dynamic feedback, respectively.

PARAMICS simulates three types of route choice, namely all-or-nothing assignment, stochastic assignment and the dynamic traffic assignment. The route cost calculation is based on the generalized cost using the formula (Quadstone 2005):

$$\text{cost} = aT + bL + cP \tag{4.4}$$

where a, b and c are the parameters that can be altered by the users while T, L and P are the free-flow travel time, length of the link and the price of the toll (monetary cost) respectively. If the deterministic assignment is only based on eqn. (4.4), the assignment method is all-or-nothing assignment in which the choice of the next link to travel is purely calculated from the free flow travel time and distance (in T and L respectively) without taking into consideration the traffic conditions. Stochastic traffic assignment introduces some random noise into the total perceived cost of each route. This can be easily done by setting some percentage variation in the *Perturbation* tab. Dynamic traffic assignment takes place if the cost recalculation option is activated. If a

dynamic feedback period is set, PARAMICS uses the mean travel times simulated in the last feedback period for T in eqn. (4.4) rather than the free-flow travel time. Two routing tables are constructed for the familiar and unfamiliar drivers respectively. In the dynamic traffic assignment, only the routing table of the familiar driver will be updated at the end of every feedback period. This means that only familiar drivers will react to avoid the congested bottleneck in the network while the unfamiliar drivers will not change their routes.

The extra functions are embedded in PARAMICS through the Application Programming Interface (API). Through API, the route choice of a driver can be overridden. The control strategy proposed in this paper is coded in the PARAMICS through the Dynamic Link Library (DLL) which emulates the response of drivers to the strategy. The strategy proposed will influence unfamiliar drivers as well as the familiar drivers. This means that when the strategy is triggered, the DLL plug-in will take over the original route choice model in PARAMICS and control the route choice of the familiar drivers. However, the original route choice model will resume for the remaining journey after the drivers have left the study area.

4.4.2 Determination of Drivers Complying ATIS Information

PARAMICS incorporates detailed driver behavior in the model. Besides familiarity, each driver is incorporated with a value to measure the drivers' awareness behavior, indicated as *Aw* value in this study. The drivers' awareness behavior is appropriate in this study as a performance measure for the drivers' compliance rate to the ATIS information. This is straight forward since drivers need to be aware of the ATIS information prior to make decision whether to comply or not to the information or advice given. The *Aw* value in PARAMICS ranges from a scale of 1 to 8; in which

a Aw value of 1 has the least awareness while a value of 8 has the most awareness. The Aw value varies according to the distribution specified. When a driver is released from the origin zone, a Aw value is assigned to it. The Aw value is used in the car following and gap acceptance model in PARAMICS. A driver with a high Aw value will be more aware of the ATIS information released at the roadside or through radio on-board and decide whether to comply or not. The default distribution for Aw values in PARAMICS is the normal distribution. By assuming the mean value, $\overline{Aw} = 4$ and standard deviation, $\sigma = 2.5$, the compliance decision of each individual driver to the ATIS information could be determined. Let γ be the drivers' compliance rate to the ATIS information, which can be obtained from survey. Thus, a driver follows the ATIS information if the driver's Aw value satisfies the following condition:

$$\text{Aw value} \geq \overline{Aw} + \sigma\Phi^{-1}(1-\gamma) \tag{4.5}$$

where $\Phi(x)$ is the cumulative distribution function of the standard normal distribution.

The derivation of eqn. (4.5) is shown as follows. Let $Aw_1, Aw_2, Aw_3......Aw_n$ be the Aw values for driver $1, 2, 3.....n$. According to the default setting of Aw value in PARAMICS, these Aw values are normal distributed with mean, $\overline{Aw}$ and standard deviation, σ such as shown follows:

$$\{Aw_1, Aw_2, Aw_3....Aw_n\} \sim N(\overline{Aw}, \sigma) \tag{4.6}$$

Assume that the Aw value associate with a driver's compliance rate γ is Aw', and any driver with Aw value greater than Aw' will follow the ATIS advice. With this,

$$P(Aw \geq Aw') \geq \gamma \tag{4.7a}$$

$$P(\frac{Aw - \overline{Aw}}{\sigma} \leq Aw') \geq (1-\gamma) \tag{4.7b}$$

$$\Phi(\frac{Aw - \overline{Aw}}{\sigma}) \geq (1 - \gamma) \tag{4.7c}$$

Eqn. (4.7a) indicates that the probability of a driver complied with the ATIS advice is greater than or equal to the compliance rate if it has Aw value greater than or equal to the cutting point Aw'. Eqn. (4.7b) transforms the Aw value into a standard normal random variable. This leads to eqn. (4.7c). By doing some simple mathematical manipulation on eqn. (4.7c), eqn. (4.5) could be obtained. When the control strategy is activated, both familiar and unfamiliar drivers are affected. This is because the control strategy will override the route choice models embedded in the microscopic traffic simulation model.

4.4.3 Simulation Replications Using Statistics Analysis

PARAMICS is a stochastic simulation model which deals extensively with random numbers. For example, it uses random numbers to simulate vehicle releases, aggressiveness, awareness, route choice and others. Hence, the result of one simulation run may not be representative of the typical outcome. The results may change substantially with different random number seeds. Multiple runs are needed to produce reliable results. The number of simulation replications is determined using statistical analysis. This is because collecting simulation results is similar to collecting samples from a population. We need to specify α, the confidence level required and β, the desired relative error between sample mean $\overline{X}$ and the true population mean to determine the sufficient sample size. The formula used to compute this is:

$$B = \left(t_{\alpha/2} \times \frac{\delta}{\overline{X}\beta} \right)^2 \tag{4.8}$$

where δ is sample standard deviation; $t_{\alpha/2}$ is the critical value of the t-distribution at significance level α. The t-distribution is adopted in eqn. (4.8) because the population variance is unknown and the initial number of replication is small. In this case, an initial replication of 5 is set. Then, $\overline{X}$ and δ are calculated. Using α =0.05 and $\beta = 5\%$, the required number of replication B is calculated using eqn. (4.8). If the sample size is not adequate, one additional simulation run will be performed and the whole process is repeated until the sample size is sufficient.

4.5 Case Study

4.5.1 Network Coding and Setting

The network selected for the simulation covers the entire length of 26.5km of Ayer Rajah Expressway (AYE) in Singapore. It stretches from the Second Link to Malaysia at the west end of Singapore to the container port terminals at the edge of the Centre Business District (CBD). Included in the network are the parallel arterials forming possible diversion routes. The network is about 177km^2, which is about one quarter of Singapore. The network is coded in Version 5.1 PARAMICS Modeler. It consisted of 56 original-destination zones, 2950 links and 1162 nodes. The modeling parameters have been calibrated by Qi (2002). From the expressway, a length of about 5km segment is designated as the study area under the control of the proposed online strategy. The network and the study area are shown in Figure 4.5. The study area is highlighted in a circle.

The bottleneck on the expressway is simulated by creating an incident that blocks some lanes on the expressway at the location denoted in the figure. The entire study area is chosen because it consisted of three pairs of on- and off-ramps at the

upstream of the incident location. This enables the vehicles to divert through the off-ramps in response to the control algorithm. There are also many on-ramps downstream of the incident location. These on-ramps allow traffic to return to the expressway. Between the off- and on-ramps, there are surface streets available for diversion. The expressway section where the bottleneck occurs has a length of 1.23km. The bottleneck location is simulated at location 1.1km from the upstream off-ramp, which is near the end of the section. The warm up period of the simulation takes 20 minutes while an extra 20 minutes is set after the incident has ended. The incident occurs fifteen minutes after the end of the warm up period. The total simulation period is one hour and forty minutes.
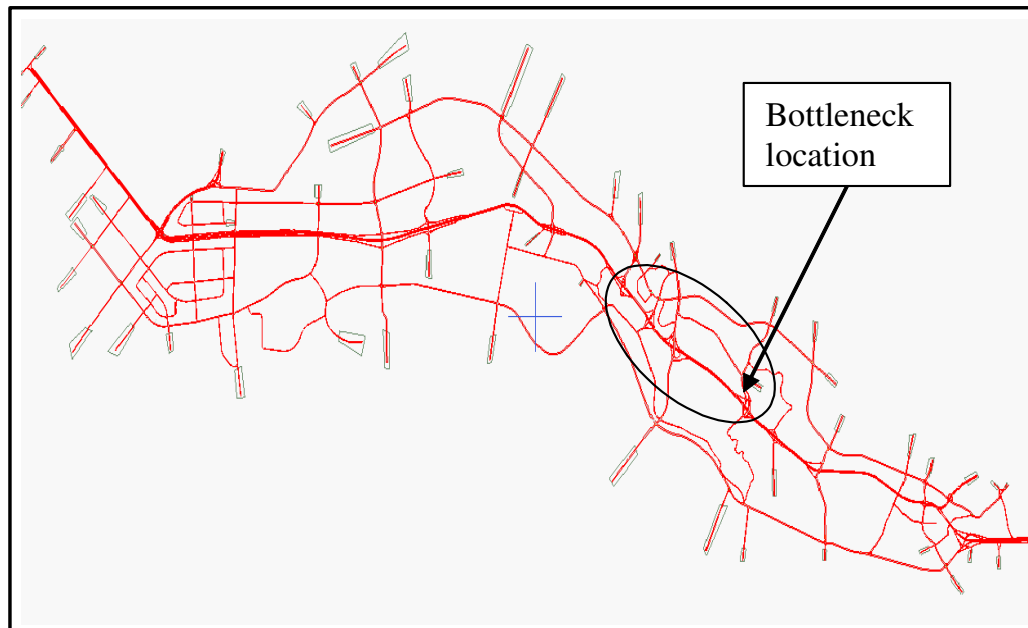
Figure 4.5 Study network and study area

The route choice override cannot be done directly in PARAMICS. This is because PARAMICS determines the turning decision of the vehicles (either familiar or unfamiliar driver) two links prior to the junction. In order to override the route choice

decision, two short dummy links, each less than 4m, need to be created before the bifurcation point of the ramp and expressway. These dummy links also prevent unusual driving behavior such as last-minute lane change at the bifurcation node.

4.5.2 Simulation Scenarios

4.5.2.1 Benchmark Scenarios

A base case scenario is developed as a benchmark for comparison and for sensitivity analysis. For the base case scenario, the origin-destination (OD) demand is assumed to generate traffic flow of about 1500veh/hr/lane at the study area. The dynamic feedback period is set as 3 minutes while the percentage of familiar driver is 30%. For the ATIS-based online dynamic traffic control strategy, the average allowable number of vehicles remained on the link a_i is set as 25 veh/km/lane according to the Level of Service D in the Highway Capacity Manual (HCM 2000), i.e. $\Pi = 25$. The length of the queue at ramps is limited by the ramps' physical capacity. A driver is not allowed to turn into an on- or off-ramp when the ramp is at its capacity. The calculation of the ramp queue size for each lane is such that:

$$S_{j,\max} = h_j \Big/ E \tag{4.9}$$

where h_j is length of ramp j and E is the effective length of private cars that include spacing between them. This means that both Π_j^{off} and Π_j^{on} are determined using eqn. (4.9).

Eqn. (4.9) is an approximation of the allowable queue size on ramps. There are four vehicle types in the simulation model, namely ordinary goods vehicle type 1 (10%), type 2 (10%), long goods vehicle (5%) and private cars (75%). Because private cars have the largest proportion in the vehicle population, the length of this vehicle

class is used in calculation. In the model, the length of private cars is set at 4m. A strict control on the ramps is not imposed because the proposed strategy is not aimed to control ramps.

The percentage of familiarity and the driver's compliance rate is assumed to be 30% and 50% respectively since their true values are unknown in literature. Sensitivity analysis will be carried out to evaluate the impact of these values in the system performance. Finally, to simulate the traffic bottleneck, it is assumed that one lane blockage with duration 45 minutes, long enough to let the congestion propagate to the upstream ramps.

4.5.2.2 Network-based Sensitivity Test Scenarios

Four scenarios were created to test the effect of network setting on the performance of the proposed strategies. Scenarios 1 and 2 simulate the different level of traffic demand in the study area. Scenario 1 has light demand which is about 1000veh/hr/lane while Scenario 2 has heavy demand that is 2000veh/hr/lane. Scenarios 3 and 4 manipulate the percentage of driver familiarity at 20% and 40% respectively. For these scenarios, all other parameters are kept to be the same with the benchmark scenario.

4.5.2.3 Strategy-based Sensitivity Test Scenarios

The strategy proposed in this paper has a few parameters that need to be calibrated. Sensitivity tests were performed by varying these parameters. Scenarios 5 and 6 have different driver compliance rates compared to the benchmark scenario. Scenario 5 has a compliance rate of 30% while Scenario 6 has a compliance rate of 70%. Scenarios 7 and 8 investigate the ability of the strategy to control the traffic condition at the study area by setting different target of allowable vehicles. Scenario 7

has a target of 16 veh/km (LOS B) while Scenario 8 has a target of 11 veh/hr (LOS C). The last parameter to vary is incident severity. Incident severity is governed by two parameters: lane blockage and the incident duration. In this paper, only the incident duration is adopted to justify the severity of the incident. Scenario 9 simulates an incident duration of 30 minutes (short incident) while Scenario 10 simulates an incident duration of 45 minutes (long incident). Incident duration is chosen as the parameter instead of the number of lane blocked. This is because in reality, traffic engineers have some degree of control on the incident duration. For example, a fast response to the incident may reduce the incident duration. Hence, Scenarios 9 and 10 are actually simulating the incident management efficiency indirectly.

4.5.3 Performance Measure

The performance indicator used in the analysis is the Total System Travel Time (TSTT). The computation of total travel time encountered by all the drivers is based on the equation:

$$TSTT = \sum_{n=1}^{N(T)} \left(\hat{t}_n^{out} - \hat{t}_n^{in} \right) \tag{4.10}$$

where $N(T)$ is the total number of vehicles traversing through the study area during time period $[0,T]$, and n is the specific index of these $N(T)$ vehicles. $\hat{t}_n^{in}$ and $\hat{t}_n^{out}$ are the time when vehicle n entered the study area either through expressway mainline or on-ramp, and time when vehicle n left the study area respectively. The difference between $\hat{t}_n^{in}$ and $\hat{t}_n^{out}$ is the total travel time vehicle n spent in the system.

For each of the scenario, the TSTT for the "no-control" case and "with-control" case is recorded. The effectiveness of the strategy is computed using the formula:

$$Effectiveness = \frac{TSTT_{no-control} - TSTT_{with-control}}{TSTT_{no-control}} \times 100\% \qquad (4.11)$$

4.5.4 Results and Discussions

4.5.4.1 Efficiency of the Control Strategy

Comparison between no-control and with-control is done to evaluate the effectiveness of the algorithm. Figure 4.6 shows the results of the average number of vehicles remained in the expressway section for the benchmark case.

For the first 20 minutes of simulation time, there are no difference between with- control and no-control. This is because during the warm up period, the control strategy is yet to be triggered. The incident that creates bottleneck started at time=35 minutes. There are some drivers that created congestion in the study area between 15 to 35 minutes. This is the transient congestion that vanished with time and may not warrant any diversion strategy. The transient congestion may be due to the unstable traffic condition at the expressway weaving area. It is observed that after the incidents have occurred, there is an immediate impact on the expressway section because the number of vehicles in the system increases substantially for the no-control case. However, for the with-control case, the level of congestion does not deteriorate as much compared to the no-control condition. Nevertheless, it is observed that the number of vehicles that remained in the study area fluctuated around the target level.

The fluctuating condition can be explained by two reasons. The first reason is the drivers' compliance rate. There is a probability that a driver or a series of successive drivers refuse to divert from the expressway section and thus they cause in the number of vehicles in the study area to be larger than the desired value. On the other hand, if more drivers accept the diversion advice, the strategy keeps the number

of vehicles below the desired level. The second reason is the OnC and OffC mechanisms. Because the setting of the surface street network (for example, the signal timing) remains the same with and without diversion control, the surface road may not have enough capacity to handle the sudden increase of traffic flow due to diversion. When a queue spillback was propagated to the ramp, OnC and OffC mechanism will override and terminate the EMC mechanism. Since the EMC is terminated earlier, driver on expressway will not receive any diverting advice and continue traveling on the expressway. This causes the number of remaining vehicles to be greater than the desired level. The incident ended at simulation time= 80 minutes and a sharp decrease in the average number of vehicles on the site is observed for the subsequent period of time. At this point, the EMC is terminated because the average queue size is smaller than the desired level. It can be seen that for the with-control case, the traffic condition on the highway is more manageable and is able to recover to a better level of service in shorter time. In addition, the efficiency of the strategy is a saving of 11.7% of TSTT.
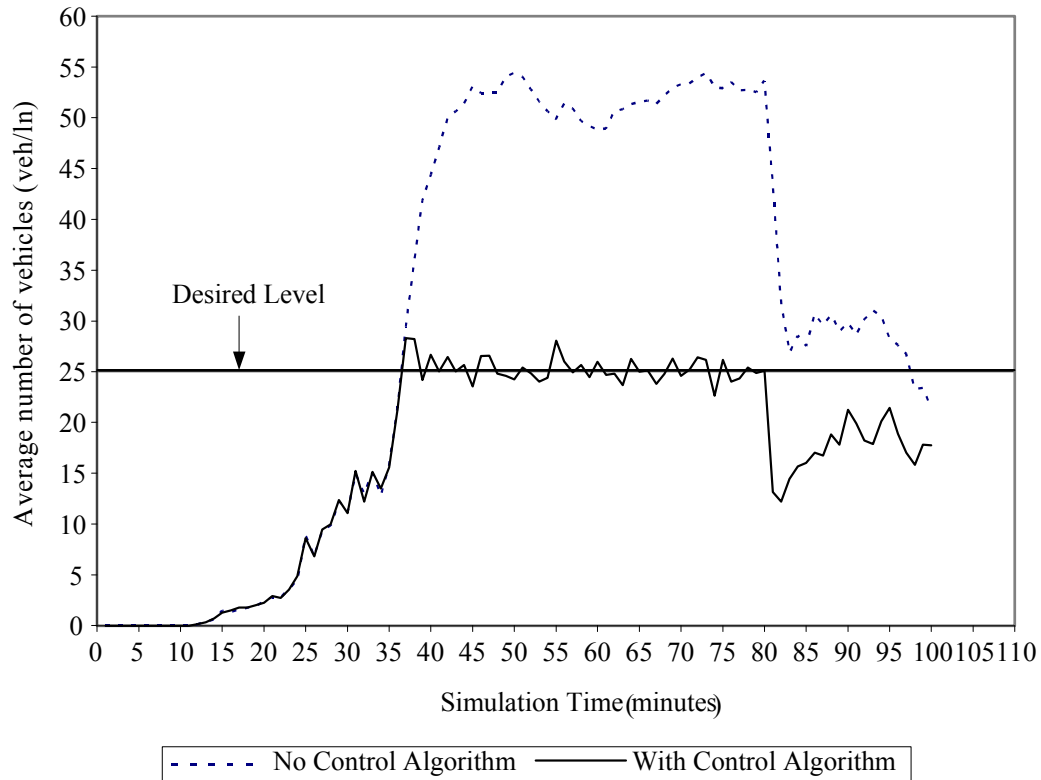
Figure 4.6 Average queue size for no control and control algorithm

4.5.4.2 Sensitivity Analysis Test Results

Table 4.1 shows the results of the sensitivity tests. It is observed that the effectiveness of the proposed strategy is sensitive to all the parameters. The results of Scenarios 1 and 2 indicate that the strategy has a better performance for the medium demand level, i.e. the benchmark case. The results of Scenarios 3 and 4 show that the strategy is more effective if there are fewer familiar drivers in the network. Our strategy is quite insensitive to the percentage of familiar drivers. Both types of drivers are affected by the strategy. However, for the no-control case, it is sensitive to the percentage of familiar drivers. A higher total travel time (greater delay) is observed when there are lesser familiar drivers. This is because drivers may not choose alternative road during congestion. Hence, when the no-control case and control case is compared, our strategy is more effective for lesser percentage of familiar drivers.

Table 4.1 Result of the sensitivity analysis test

Scenario	Parameter	Effectiveness (%)
1	Light demand	7.7
2	Heavy demand	2.3
3	Familiarity: 20%	14.3
4	Familiarity: 40%	7.6
5	Compliance rate: 30%	9.8
6	Compliance rate: 70%	12.5
7	Target level: LOS C	14
8	Target level: LOS B	19
9	Incident duration: 30 minutes	6.2
10	Incident duration: 60 minutes	19.8

The results of Scenarios 5 and 6 show that the strategy is more effective if more drivers comply with the diversion advice. On the other hand, the results of Scenarios 7 and 8 shows that the strategy is still efficient at both target levels. Finally, the results of Scenarios 9 and 10 show that the strategy is more effective for more severe incidents. This means that the strategy can be applied to the expressway-arterial corridor that does not incorporate any effective incident management strategy. In such a case, the proposed strategy plays an important role in managing traffic on the network corridor while waiting for the agency to respond to the incident.

4.6 Summary

This chapter consists of two main parts. The first part of the chapter proposes a novel real-time dynamic traffic control operation for the expressway-arterial corridor systems using ATIS. The proposed strategy comprises three components which aim to minimize the congestion level in the corridor system. The model assumed that real-time traffic measurements can be obtained from the detectors on the expressway and the congestion information is disseminated to the expressway drivers by ATIS means. The second portion of the chapter is devoted to develop a simulation-based methodology to evaluate the proposed traffic control operation. A case study has been described to show the applicability of the proposed control operation in alleviating the congestion level on the expressway-arterial corridor system. It was shown that the proposed operation is effective in reducing the total system travel time of the expressway-arterial corridor. A sensitivity analysis was conducted to test the effects of using different values for the parameters. It was found that the control operation is sensitive to the network and parameters tested.

CHAPTER 5 MODIFIED CELL TRANSMISSION MODEL FOR RAMP METERING OPERATIONS

5.1 Introduction

The design, testing and evaluation of any dynamic traffic control and management strategies at the operational level require a dynamic traffic flow model that can simulate the minute-by-minute change in traffic flow condition. This is important especially in capturing the queue propagation phenomenon due to bottlenecks. One of the major aspects in the study of ramp metering operations is the selection of an appropriate traffic flow model to describe traffic flow, and the merging and diverging situation in the expressway-ramp-arterial network. Most of the models adopted are point models where the horizontal queue propagation cannot be captured.

The Cell Transmission Model (CTM) introduced by Daganzo (1994, 1995) could overcome this shortcoming by capturing the horizontal queue formation and dissipation. This is crucial in the ramp metering context because queue at the on-ramps due to the metering rate needs to be captured and modeled appropriately. By doing so, the performance of the metering strategy could be analyzed more accurately. In addition to this, CTM could capture shockwaves; maintain first-in-first-out (FIFO) principle and model dynamic traffic interactions across multiple links. Considering these advantages, it is used as the traffic flow model for ramp metering operations analysis presented in the chapter. However, the original CTM needs to be modified. This chapter thus presents a custom CTM for the ramp metering operation. It discusses how the rules at the merging links/cells can be modified to imitate ramp metering

operations. The modified version of CTM is termed as modified cell transmission model, abbreviated as MCTM.

The MCTM is employed as the traffic flow model for ramp metering operations presented in the subsequent chapters. It is employed to simulate traffic flow propagation over time and space on an expressway-ramp-arterial network. In addition, the actual route travel time can be computed from MCTM in performing the dynamic stochastic user optimal assignment. A detailed formulation of the model and the modified rules are discussed in this chapter. Note that, most part of the chapter reviews the CTM introduced by Daganzo (1994,1995). The purpose of having this chapter is to connect the later chapters since they use the same traffic flow model. This could avoid unnecessary repetition. The model presented herein could be applied directly in Chapter 7, while it could also be used in Chapter 6, with a little modification to incorporate the route choice of drivers.

5.2 Cell-based Network Coding

The expressway-ramp-arterial network system is represented by a directed graph $G = (N, A)$, where N is the set of nodes denoting heads and tails of on-ramps or off-ramps; A is the set of directed arcs denoting the on-ramps, the off-ramps, the expressway stretches, and the arterial roads, connecting two consecutive points meeting an on-ramp or an off-ramp, namely $A \subset N \times N$. More specifically, let set $A = A_{on}^+ \bigcup A_{on}^- \bigcup A_{off} \bigcup A_E \bigcup A_{art}$, where A_{on}^+, A_{on}^-, A_{off}, A_E and A_{art}, are the sets of metered on-ramps, un-metered on-ramps, off-ramps, expressway stretches, and arterial segments. Note that any two of these five subsets are disjoint. Essentially, the network system defined here is similar to that defined in Chapter 4. However, in this chapter, it

is assumed that ramp meters are installed at the on-ramps to regulate traffic flow, while there is no ramp meter considered in the corridor system in Chapter 4. To differentiate these two systems, the network system explained herein is termed as expressway-ramp-arterial network system in current and subsequent chapters, while the previous chapter has the term of expressway-arterial corridor system.

Figure 5.1 schematically clarifies the convention of the symbols used. It shows an expressway-ramp-arterial network system with four on-ramps, four off-ramps, ten expressway stretches, and six arterial segments. Hence, according to the nomenclature,

$N = \{1, 2, \cdots, 16\}$, $A = \{1, 2, \cdots, 24\}$, $A_{on}^+ = \{1, 2, 3\}$, $A_{on}^- = \{4\}$, $A_{off} = \{15, 16, 17, 18\}$,

$A_E = \{5, \cdots, 14\}$, and $A_{art} = \{19, 20, 21, 22, 23, 24\}$.
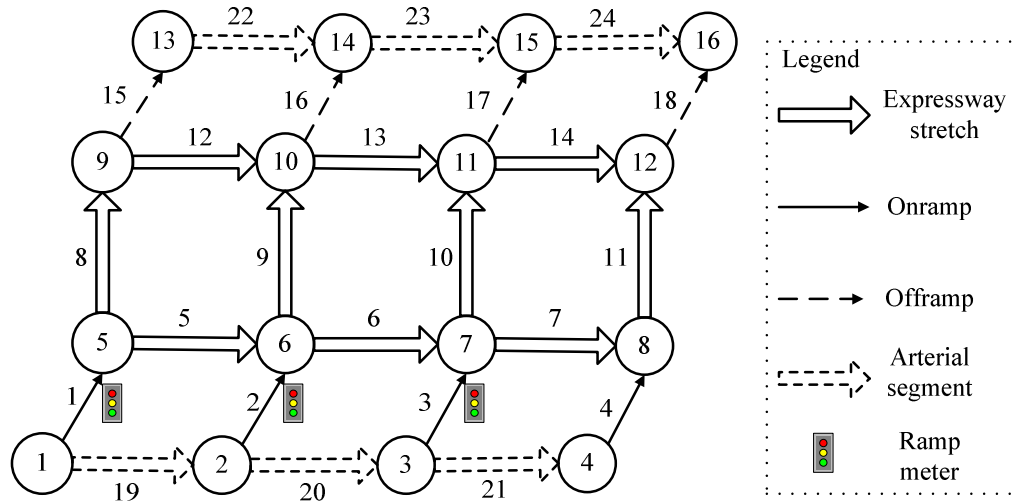
Figure 5.1 A schematic directed graph representation of an expressway-ramp-arterial network system

Let R and S denote the sets of origins and destinations, respectively, and $R, S \subseteq N$. Let also K_{rs} denote the set of all routes from an origin $r \in R$ to a destination $s \in S$, namely OD pair rs and $r \neq s$. Assume that the demand for an OD pair rs at time t is denoted by $q_{rs}(t)$ and is loaded at the origin node r . This demand

will travel to the destinations via K_{rs} routes. At each node, the stream of traffic chooses to move to the next node according to a pre-determined route or the best route in terms of travel time. The traffic is allowed to enter the expressway through any on-ramp and similarly, leave the expressway using any off-ramp between origin and destinations. Alternatively, traffic may choose to use arterial roads to travel.

In addition to this, a sub-graph defining an expressway-ramp network system that does not contain the arterial segments could be defined. Essentially, the expressway-ramp network system is represented by $G' = (N, A')$, in which $A' = A \setminus A_{art}$ and $G' \subseteq G$. By eliminating the arterial roads from the network system, the drivers' route choice decision, i.e. the choice to travel on the alternative roads is neglected. It is assumed that for the expressway-ramp network system, traffic demand is generated at the tail of each on-ramp without specifying their destinations, and that traffic leaves the system through the head of an off-ramp with a known split ratio to the total traffic flow of the expressway stretch. The assumption of given split ratios is quite common in the ramp metering studies (e.g. Gomes and Horowitz, 2006, Zhang and Levinson, 2004) but is not perfect, since these are actually functions of the control variable. However, this could help eliminate the intractability problem of ramp metering (Erera et al., 1999). If only the expressway-ramp network system is considered with split ratios pre-determined, the route choice decision of drivers could not be captured in the system.

An interested time period interval is discretized into T equal time intervals of length Δt, namely, $\left[0, \Delta t, 2\Delta t, 3\Delta t, \cdots, (T-1)\Delta t \right]$. Without loss of generality, it is assumed that $\Delta t = 1$ for easy of exposition. The entrance flow of vehicles is controlled by the on-ramps ramp meters, which is defined by the ramp metering rates. Let $Z_a(t)$

be the metering rate from clock tick t to clock tick $t+1$, where $t \in [0,1,2,\cdots,T-1]$, for on-ramp $a \in A_{on}^+$, and **Z** be a dynamic ramp metering rate solution denoted by a row vector of all on-ramp time-dependent metering rates, namely:

$$\mathbf{Z} = \left(Z_a(t), a \in A_{on}^+, t = 0,1,\cdots,T-1 \right) \tag{5.1}$$

After discretizing each arc of a network into homogenous cells and the interested time period into equal time intervals such that the cell length is equal to the distance traveled by free-flowing traffic in one time interval, a cell-based network is obtained. A link which connects two consecutive cells in the cell-based network merely is used only to represent the flow of vehicles moving from one cell to another one more conveniently. For the numerical calculation purposes, it is prescribed that within the cell-based network, the maximum number of links entering and/or leaving a cell (degree of the cell) is three. Thus, links can be classified into three types: ordinary if it is connected to one entry link and one exit link, denoted by $\Omega_{ordinary}$, diverge if it is connected to one entry link and two exit links, denoted by $\Omega_{diverge}$, merge if it is connected to two entry links and one exit link, denoted by Ω_{merge}. It can be further assumed that each arc $a \in A$ is subdivided into l_a cells, where l_a is a limited positive integer. These cells are consecutively numbered starting with integer one from the upstream direction of traffic flow and follow this notation: cell (a,i) represents the ith cell in arc a.

Figure 5.2 shows these fundamental blocks. The maximum connection for a link is three. In the case of a network with some nodes whose degree is greater than three, Daganzo (1995) demonstrated that the cell-based network can be constructed by shortening the discretizing time interval. Besides, this could also be tackled by building dummy cells for connection. Figure 5.3 shows the cell-based network for the

expressway-ramp-arterial network shown in Figure 5.1. In order to cope with nodes with degree more than 3 - node 6, 7, 10, 11 – shown in Figure 5.1, 8 cells with small length marked by a dash rectangle is added in Figure 5.3.

(i) Ordinary link

(ii) A merge based three-legged junction

(iii) A merge based three-legged junction
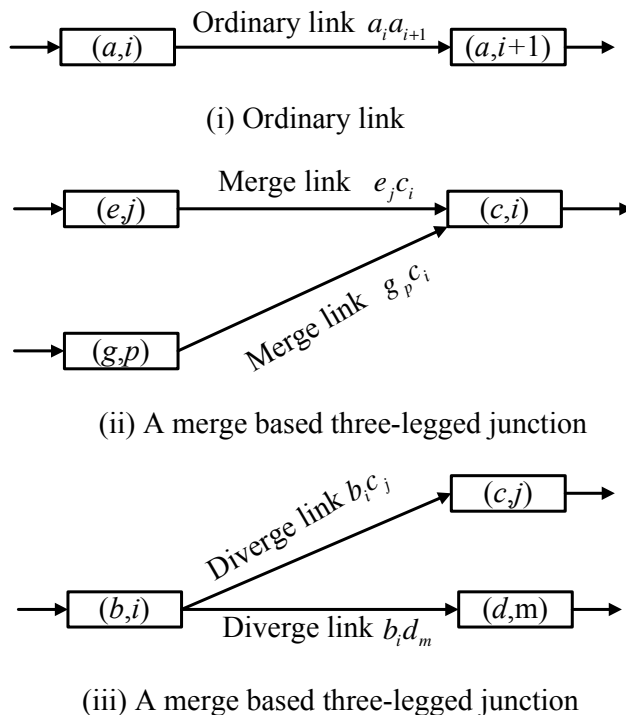
Figure 5.2 Three fundamental blocks constructing the cell-based network
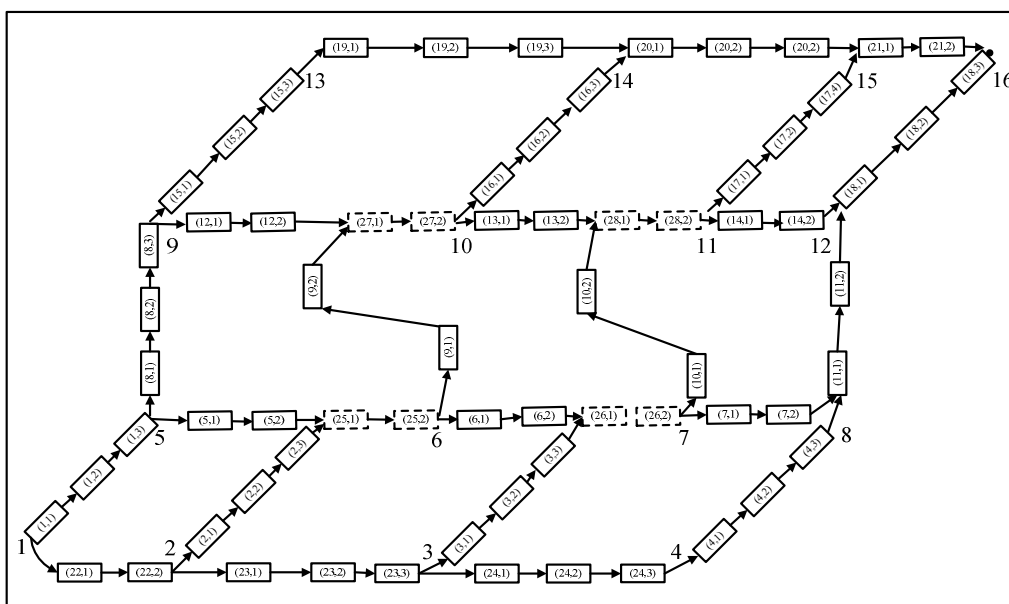
Figure 5.3 The cell-based expressway-ramp-arterial network

125

5.3 Two MCTM Updating Procedures

The MCTM with the ramp metering consists of two computational interrelated procedures –Modified Procedure 1: calculation of the flow of vehicle on a link in time interval from t to $t+1$, Procedure 2: update of the number of vehicles presented in a cell at time $t+1$. Modified Procedure 1 determines the maximum flow of vehicles that can move from one cell to the next cell when the time advances from t to $t+1$, namely dynamic traffic flow of a link connecting these two cells in the cell-based network. With the dynamic link flow pattern obtained in Procedure 1, the purpose of Procedure 2 is to update the occupancies of each cell at time t following the fundamental flow conservation law.

To carry out the calculation according to the mentioned procedures, the sending and receiving capacities of a cell need to be defined. Given a cell state variable $n_{a_i}(t)$, the vehicle volume presented in cell (a,i) at time t, two cell reserve capacities for sending and receiving vehicles in the time interval between t and $t+1$: $S_{a_i}(t)$ and $R_{a_i}(t)$ need to be identified. For cells in the expressway segments, un-metered on-ramps, off-ramps, and arterial segments, Daganzo's formulae (Daganzo, 1995) could be adopted to calculate the effective capacities of cells, i.e.

$$S_{a_i}(t) = \min\left\{Q_{a_i}(t), n_{a_i}(t)\right\}, \quad i = 1, 2, \cdots, l_a, a \in A_{on}^- \bigcup A_{off} \bigcup A_E \bigcup A_{art} \tag{5.2}$$

$$R_{a_i}(t) = \min\left\{Q_{a_i}(t), w/v\left[N_{a_i}(t) - n_{a_i}(t)\right]\right\}, \quad i = 1, 2, \cdots, l_a, a \in A_{on}^- \bigcup A_{off} \bigcup A_E \bigcup A_{art} \tag{5.3}$$

where ω and v are backward shock wave speed and free flow speed respectively.

For a metered on-ramp, the maximum possible vehicle volume through the last cell at the on-ramp is governed by the metering rate. Note that the metering rate does not directly regulate vehicles presented in the other cells of the metered on-ramp.

Accordingly, the two reserve sending and receiving reserve capacities for the metered on-ramp can be determined by:

$$S_{a_i}(t) = \begin{cases} \min\{Q_{a_i}(t), n_{a_i}(t)\}, & i = 1, 2, \cdots, l_a - 1 \\ \min\{Z_a(t), n_{a_i}(t)\}, & i = l_a \end{cases}, a \in A_{on}^+ \qquad (5.4)$$

$$R_{a_i}(t) = \left\{ \min\{Q_{a_i}(t), w/v[N_{a_i}(t) - n_{a_i}(t)]\} \right\}, \quad i = 1, 2, \cdots, l_a, a \in A_{on}^+ \qquad (5.5)$$

After determination of the sending and receiving capacities according to eqn. (5.2)-(5.5) for the respective cell types, the above-mentioned two procedures are carried out to update the cell contents at every clock tick of time.

5.3.1 Modified Procedure 1

The computation of the flow of vehicle across links/cells from time t to $t+1$ follows different set of rules or formulae for different types of the links.

5.3.1.1 Ordinary Links

For ordinary links that only have one input and one output link, the determination of flow across cells is only dependent on the sending and receiving capabilities of the cells. Daganzo's method [Equations (4), (9a) and (9b) of Daganzo (1995)], as shown in eqn. (5.6), can be used for computation, namely:

$$y_{a_i a_{i+1}}(t) = \min\{S_{a_i}(t), R_{a_{i+1}}(t)\}, \ a_i a_{i+1} \in \Omega_{ordinary} \qquad (5.6)$$

where ordinary link $a_i a_{i+1}$ connect ordinary cells (a_i, t) to (a_{i+1}, t). The amount of vehicles that can flow from the sending cell to the receiving cell is governed by the occupancy of the sending cell as well as the ability of the receiving cell to accept additional vehicle given its current occupancy. Hence, the flow should be the minimum of these two terms as shown in eqn. (5.6).

5.3.1.2 Merging Links

Daganzo (1994, 1995) did not mention about how to deal with the merging links of the ramp metering. This chapter filled in the gap with regards to this topic. There are two situations occurs in merging links. The first case is when the receiving links have the full capacity to receive all the vehicles while the second case is that it cannot receive all the vehicles. Two sets of equations are derived to describe these different situations.

For the first case, all vehicles that can be sent by both start cells of the paired merging links $\left(e_j c_i, g_p c_i \right)$ - (e,j) of arc $e \in A$ and (g,p) of arc - can reach the merging cell (c,i) if $R_{c_i}(t) \geq S_{e_j}(t) + S_{g_p}(t)$, namely:

$$y_{e_j c_i}(t) = S_{e_j}(t) \quad \text{and} \quad y_{g_p c_i}(t) = S_{g_p}(t), \text{ if } R_{c_i}(t) \geq S_{e_j}(t) + S_{g_p}(t) \qquad (5.7)$$

For the second case, if $R_{c_i}(t) < S_{e_j}(t) + S_{g_p}(t)$, i.e. the merge cell cannot accommodate all vehicles from these two start cells that of the paired merge links, the fractions of vehicles from these two start cells will go to the merge in time interval $(t, t+1)$ is identified under the FIFO rule. Let $\lambda_{c_i}(t)$ denote the minimum wait of the vehicles presented in two cells (e,j) and (g,p) at time t that will enter merge (c,i) in time interval $(t, t+1)$. With FIFO rule, $\lambda_{c_i}(t)$ readily indicates which vehicles are allowed to flow. Accordingly, we assume that the traffic flows disaggregated by waiting time τ on the paired merge links are given by the following function of $\lambda_{c_i}(t)$:

$$y_{e_j c_i \tau}(t) = \begin{cases} n_{e_j \tau}(t), & \text{if } \tau > \left| \lambda_{c_i}(t) \right|^+ \\ \left[\tau - \lambda_{c_i}(t) \right] n_{e_j \tau}(t), & \text{if } \tau = \left| \lambda_{c_i}(t) \right|^+ \\ 0, & \text{if } \tau < \left| \lambda_{c_i}(t) \right|^+ \end{cases} \qquad (5.8)$$

$$y_{g_p c_i \tau}(t) = \begin{cases} n_{g_p \tau}(t), & \text{if } \tau > \left|\lambda_{c_i}(t)\right|^+ \\ \left[\tau - \lambda_{c_i}(t)\right] n_{g_p \tau}(t), & \text{if } \tau = \left|\lambda_{c_i}(t)\right|^+ \\ 0, & \text{if } \tau < \left|\lambda_{c_i}(t)\right|^+ \end{cases} \tag{5.9}$$

where $\left|\lambda_{c_i}(t)\right|^+$ denotes the smallest integer equal to or greater than $\lambda_{c_i}(t)$.

Logically, eqns. (5.8)-(5.9) indicates that all the occupancies in the two start cells of a paired link that have entered to these cells at time interval prior to $t - \lambda_{c_i}(t)$ can move to the merge cell. Conversely, none of those vehicles with $\tau < \left|\lambda_{c_i}(t)\right|$ can proceed. Of those vehicles having entered the two start cells in the time interval containing time $t - \lambda_{c_i}(t)$, only a fraction equal to the proportion of the interval that precedes $\lambda_{c_i}(t)$ is advanced. In addition, the aggregate traffic flow on these two merge links can be expressed as:

$$y_{e_j c_i}\left(t, \lambda_{c_i}(t)\right) = \sum_{\tau=1}^{\tau_{e_j}(t)} y_{e_j c_i \tau}\left(t, \lambda_{c_i}(t)\right) \tag{5.10}$$

$$y_{g_p c_i}\left(t, \lambda_{c_i}(t)\right) = \sum_{\tau=1}^{\tau_{g_p}(t)} y_{g_p c_i \tau}\left(t, \lambda_{c_i}(t)\right) \tag{5.11}$$

where $\tau_{e_j}(t)$ and $\tau_{g_p}(t)$ denote the maximum waits present in cell (e, j) and (g, p), respectively. Let $Y_{e_j g_p c_i}\left(t, \lambda_{c_i}(t)\right)$ be the aggregate traffic flows on the paired merge links $\left(e_j c_i, g_p c_i\right)$, namely:

$$Y_{e_j g_p c_i}\left(t, \lambda_{c_i}(t)\right) = y_{e_j c_i}\left(t, \lambda_{c_i}(t)\right) + y_{g_p c_i}\left(t, \lambda_{c_i}(t)\right) \tag{5.12}$$

Eqns. (5.8) and (5.9) define nonincreasing, piecewise-linear functions of the minimum wait $\lambda_{c_i}(t)$, which change in slop at the integers (see Section 4.2, Daganzo, 1995). Because these functions are nonincreasing, function $Y_{e_j g_p c_i}\left(t, \lambda_{c_i}(t)\right)$ is also a

nonincreasing and piecewise-linear function with respect to the minimum wait. Hence, it is possible to define a nonincreasing inverse function of function $Y_{e_j g_p c_i}\left(t, \lambda_{c_i}(t)\right)$, denoted by $\Phi_{e_j g_p c_i}$, which gives the minimum wait $\lambda_{c_i}(t)$ at the two start cells of the paired merge links, for a given value of the aggregate flow $Y_{e_j g_p c_i}$:

$$\lambda_{c_i}(t) = \Phi_{e_j g_p c_i}\left(t, Y_{e_j g_p c_i}\right) \tag{5.13}$$

Given $R_{c_i}(t)$, the corresponding minimum wait can be calculated by

$$\lambda_{c_i}(t) = \Phi_{e_j g_p c_i}\left(t, R_{c_i}(t)\right), \quad \text{if } R_{c_i}(t) < S_{e_j}(t) + S_{g_p}(t) \tag{5.14}$$

Thus, the dynamic traffic flow on the paired merge links can be calculated below.

$$y_{e_j c_i}(t) = \sum_{\tau=1}^{\tau_{e_j}(t)} y_{e_j c_i \tau}\left(t, \Phi_{e_j g_p c_i}\left(t, R_{c_i}(t)\right)\right) \tag{5.15}$$

$$y_{g_p c_i}(t) = \sum_{\tau=1}^{\tau_{g_p}(t)} y_{g_p c_i \tau}\left(t, \Phi_{e_j g_p c_i}\left(t, R_{c_i}(t)\right)\right) \tag{5.16}$$

Let show a simple example to demonstrate the calculation of the minimum wait. Suppose that at time $t = 15$ the receiving capacity of merge c_i is 15 vehicles, namely, $R_{c_i}(15) = 15$, and that there are two groups of vehicles with different wait times at the start cell of the merge links: $n_{e_j,14}(15) = 5$, $n_{e_j,10}(15) = 8$, $n_{e_j,9}(15) = 5$; $n_{g_p,12}(15) = 4$, $n_{g_p,10}(15) = 4$, $n_{g_p,8}(15) = 3$. The minimum wait is computed by:

$$\left[10 - \lambda_{c_i}(15)\right] \times \left[n_{e_j,10}(15) + n_{g_p,10}(15)\right] = \left[R_{c_i}(15) - n_{e_j,14}(15) - n_{g_p,12}(15)\right] \tag{5.17}$$

Solving the above equation yields that $\lambda_{c_i}(15) = 9.5$. According to eqns. (5.15)-(5.16), it follows that

$$y_{e_j c_i}(15) = 5 + (10 - 9.5) \times 8 = 9 \tag{5.18}$$

$$y_{g_p c_i}(15) = 4 + (10 - 9.5) \times 4 = 6 \tag{5.19}$$

It should be pointed out that the above method is motivated by Section 4 of Daganzo (1995).

5.3.1.3 Diverging Links

The determination of vehicle flow for diverging links can be classified into two categories depends on whether the origin-destination (OD) trip information is specified or not. If the OD trip is not specified, this means that the route choice of the drivers is not considered. The fraction of the advancing vehicles at the diverging cell (b, i) to the both end cells, (c, j) and (d, m) is determined by specifying the turn proportion to these cells, $\beta_{b_i c_j}(t)$ and $\beta_{b_i d_m}(t)$ respectively. Daganzo (1995) showed that the flow on each of the diverging links are indicated by eqn. (5.20).

$$\begin{cases} y_{b_i c_j}(t) = \beta_{b_i c_j}(t) \times \min\left\{ S_{b_i}(t), R_{c_j}(t)/\beta_{b_i c_j}(t), R_{d_m}(t)/\beta_{b_i d_m}(t) \right\} \\ y_{b_i d_m}(t) = \beta_{b_i d_m}(t) \times \min\left\{ S_{b_i}(t), R_{c_j}(t)/\beta_{b_i c_j}(t), R_{d_m}(t)/\beta_{b_i d_m}(t) \right\} \end{cases}, \quad (b_i c_j, b_i d_m) \in \Omega_{diverge} \quad (5.20)$$

Note that: $\beta_{b_i c_j}(t) + \beta_{b_i d_m}(t) = 1$. For each direction, the flow on each link is the minimum of the three terms indicated in eqn. (5.20), the sending cell occupancy, the expressway receiving cell and off-ramp receiving cell according to the proportion specified. It is clear that, the sending cell cannot send more than its occupancy while the receiving cell cannot accept flow that is greater than its receiving ability according to the proportion specified.

If the OD trips of the vehicles are defined, the proportions of the vehicles going to both end cells of the diverging cell could be derived. It is assumed that, the best-route information takes the form of route choice constants defined for the two end cells of the diverge, denote respectively by, $\beta_{c_j dest}(t)$ and $\beta_{d_m dest}(t)$. These parameters give the proportion of vehicles with destination *dest* that would travel from the diverging

cell into each of the end cells. In addition, to ensure the FIFO principle, the vehicles in the diverging cell needs to be disaggregated according to their waiting time. Daganzo (1995) derived sets of equations to compute the diverging flow and the minimum link wait time. The concept is similar to those adopted in section 5.3.1.2 in determining the vehicle flow for merging cells, with the exception that for the diverging cell, the destination of the vehicles need to be captured. A brief description of the methodology is presented here. For details, readers may refer to Section 4.2, Daganzo (1995).

The flow of vehicle from the diverging cell (b,i) to cell (c,j) and (d,m) respectively is disaggregated by the vehicles waiting time τ at the diverging cell. Assume that $\gamma_{b_i}(t)$ denote the minimum wait of the vehicles at the diverging cell (b,i), the flow on the diverging links with destination *dest* are shown as follows:

$$y_{b_i c_j \tau}(t) = \begin{cases} n_{b_i \tau}(t)\beta_{b_i c_j dest}, & \text{if } \tau > \left|\gamma_{b_i}(t)\right|^+ \\ \left[\tau - \gamma_{b_i}(t)\right]n_{b_i \tau}(t)\beta_{b_i c_j dest}, & \text{if } \tau = \left|\gamma_{b_i}(t)\right|^+ \\ 0, & \text{if } \tau < \left|\gamma_{b_i}(t)\right|^+ \end{cases} \qquad (5.21)$$

$$y_{b_i d_m \tau}(t) = \begin{cases} n_{b_i \tau}(t)\beta_{b_i d_m dest}, & \text{if } \tau > \left|\gamma_{b_i}(t)\right|^+ \\ \left[\tau - \gamma_{b_i}(t)\right]n_{b_i \tau}(t)\beta_{b_i d_m dest}, & \text{if } \tau = \left|\gamma_{b_i}(t)\right|^+ \\ 0, & \text{if } \tau < \left|\gamma_{b_i}(t)\right|^+ \end{cases} \qquad (5.22)$$

where $\left|\gamma_{b_i}(t)\right|^+$ denotes the smallest integer equal to or greater than $\gamma_{b_i}(t)$. Essentially, eqns. (5.21)-(5.22) shows that those vehicles that have waited for $\gamma_{b_i}(t)$ can proceed, and vice versa. Only a portion of the vehicles is allowed to flow if their waiting time is exactly the same to the minimum waiting time. In addition, Daganzo (1995) also demonstrated how to determine the minimum wait time from the aggregate flow of

vehicles. He showed that the minimum wait time for diverging cell is the maximum of the three parameters as shown below:

$$\gamma_{b_i} = \max\left\{\varphi_{b_i c_j}\left(t, R_{b_i c_j}\right), \varphi_{b_i d_m}\left(t, R_{b_i d_m}\right), \varphi_{b_i}\right\} \tag{5.23}$$

where $\varphi_{b_i}\left(t, y_{b_i}\right)$ is the non-increasing inverse function of the flow on cell b_i. Having determined the minimum wait time, the diverging flow of the vehicles can be computed.

5.3.2 Procedure 2

According to the flow conservation law, occupancy of a cell at time $t+1$ equals its occupancy at time t, plus the inflow and minus the outflow, i.e.,

$$n_{a_i}\left(t+1\right) = n_{a_i}\left(t\right) + \sum_{h_m \in \Gamma^+\left(a_i\right)} y_{h_m a_i}\left(t\right) - \sum_{h_m \in \Gamma^-\left(a_i\right)} y_{a_i h_m}\left(t\right), i = 1, 2, \cdots, l_a; a \in A \tag{5.24}$$

where $\Gamma^+\left(a_i\right)$ is the set of all start cells of those links that enter cell $\left(a, i\right)$, and $\Gamma^-\left(a_i\right)$ is the set of all end cells of those links that leave cell $\left(a, i\right)$.

5.4 Summary

This chapter presents a custom dynamic traffic flow model for the expressway-ramp-arterial network system. The modified version of cell transmission model (MCTM) is incorporated with a FIFO principle in determining the merging flow of vehicles at the merging links. It is carried out by disaggregating vehicle flow according to the waiting time of vehicle in the cell. On the other hand, the procedures for the other two types of cells, namely the ordinary cell and the diverging cell used the rules introduced by Daganzo (1995). The advantages of MCTM are its ability to capture shock wave, model horizontal queue propagation and dissipation, and fulfill the FIFO

principle. This is crucial especially in modeling traffic conditions in an expressway-ramp-arterial network, where the queuing condition at the on-ramps needs to be adequately addressed. The model developed in this chapter will be used as the dynamic traffic flow model for ramp metering analysis in the subsequent chapters.

CHAPTER 6 OPTIMAL RAMP METERING OPERATIONS WITH PROBIT-BASED IDEAL STOCHASTIC DYNAMIC USER OPTIMAL CONSTRAINTS

6.1 Introduction

Ramp metering operations have been proved to bring improvement to the expressway-ramp network system by increasing the system's throughput and reducing the total travel time. Research studies about ramp metering operations aim to find the best ramp metering rate that optimizes the performance of the operation. Given any ramp metering rate, drivers will respond by choosing routes that minimize their own travel time. This phenomenon can be captured by Stackelberg game in game theory (Shimuzu et al. 1997). The problem can be formulated as a bilevel programming model. For the ramp metering operation, the upper level problem determines the optimal ramp metering rate that can minimize the total system travel time and is formulated as a nonlinear programming model. The lower level problem describes the drivers' decision of selecting routes that minimize their own travel time. This could be represented by a route choice decision model. The lower level problem is important in the optimization of ramp metering studies. Wu (2001) has shown that a significant number of drivers rerouted due to the installation of ramp meters. According to his field study on USH 45 section located in Milwaukee, and 2 parallel streets at the State Highway 100 and 124th Street in US, he observed that there is about 10% of drivers

divert due to ramp metering operation. However, the driver route choice decision is neglected in most of the existing studies about ramp metering operations. By proposing a bilevel programming model for ramp metering operations, this study eliminates this limitation. In order to capture the route choice behavior, the expressway-ramp-arterial network needs to be considered.

There are many established models, describing drivers' route choice behavior, that are suitable to be adopted in the lower level problem. This study adopts the dynamic version of route choice model due to the dynamic nature of ramp metering operations. Generally speaking, there are two broad categories of the dynamic route choice models, namely the dynamic deterministic user optimal (DDUO) and the dynamic stochastic user optimal (DSUO). Both models do not optimize the departure time of drivers. In this study, DSUO is adopted. DSUO assumes that drivers are not identical in making choice decision and do not have full knowledge of the network. It allows variation in drives' behavior be addressed properly. Furthermore, DSUO aims to minimize the drivers' perceived travel time. A Probit-based DSUO model is chosen that assumes the drivers' variation in travel time is normal distributed. In this study, the DSUO is formulated as a fixed point problem. The above-mentioned bilevel programming model is thus can be converted into a single level optimization problem with the DSUO flow condition as constraints. This means that any ramp metering rate acquired from the optimization model of the ramp metering operations need to fulfill the DSUO flow constraint.

In this study, the method of successive averages (MSA) is employed to solve the Probit-based DSUO. It is an approximation solution. This is because only a subset of all the possible routes existed between an origin and destination is considered. This is reasonable since some of the routes may be less attractive and has less probability to

be chosen by the drivers. As a Probit-based route choice model is adopted, the covariance matrix is necessary to address the covariance of overlapped routes. In addition, to facilitate the computation of the actual route travel time, the MCTM developed in the earlier chapter is adopted. MCTM is chosen considering its ability to capture the propagation of traffic, model horizontal queue phenomena by shock wave, and address the first-in-first-out (FIFO) principle. However, Lo and Szeto (2002c) found that under certain traffic flow condition such as congestion, the actual route travel time function with respect to the flow is non-differentiable. This complicates the problem because such characteristics of the route travel time function affects the feasibility of solving the Probit-based DSUO with exact solution. This explains why MSA is employed to solve the model in this study.

When the Probit-based DSUO is considered as the lower level problem and formulated as a fixed point problem, the bilevel programming model for the ramp metering operations can be simplified into a single level optimization model with DSUO flow as constraints. Since there are no exact solutions, the single level optimization model is solved using meta-heuristic algorithms. This study chooses genetic algorithm in view of its popularity of the application in traffic engineering world (Gen and Cheng 1996).

6.2 Problem Statement

Consider an expressway-ramp-arterial network system, $G = (N, A)$ where N is the set of nodes; A is the set of directed arcs denoting the on-ramps, the off-ramps, the expressway stretches, and the arterial roads, connecting two consecutive points meeting an onramp or an off-ramp, namely, $A \subset N \times N$. The network system

comprises expressway segments, A_E, a group of on-ramps with ramp meters, A_{on}^+ or without ramp meters, A_{on}^-, the off-ramps, A_{off} and arterial road segments, A_{art} that stretches parallel to the expressway is linked with it through these on-off-ramps. More specifically, $A = A_{on}^+ \bigcup A_{on}^- \bigcup A_{off} \bigcup A_E \bigcup A_{art}$. Assume R and S denote the sets of origins and destinations, respectively, where $R, S \subseteq N$. Let K_{rs} denote the set of all routes from an origin $r \in R$ to a destination $s \in S$, namely OD pair *rs*. The notation of the network system defined herein is exactly similar to those mentioned in section 5.2, but is repeated for the sake of completeness.

Let $[0, T]$ denote the time horizon for the dynamic ramp metering operations, which is long enough to allow drivers departing during the time horizon to complete their trips. The time horizon is discretized into T equal time intervals of δ, namely, $[0, \delta, 2\delta, 3\delta, \cdots, (T-1)\delta]$. Let $q_{rs}(t)$ be the fixed traffic demand between OD pair *rs* departing from the origin r during time interval $[t, t+\delta]$, $t \in \{0, 1, 2, \cdots, T-1\}$; $f_k^{rs}(t)$ be route flow on route $k \in K_{rs}$ between O-D *rs* departing from origin r during time interval $[t, t+\delta]$, $t \in \{0, 1, 2, \cdots, T-1\}$. Without loss of generality, it is assumed that $\delta = 1$ hereafter. According to the flow conservation law, it follows that:

$$\sum_{k \in K_{rs}} f_k^{rs}(t) = q_{rs}(t), r \in R, s \in S, t = 0, 1, 2, \cdots, T-1 \tag{6.1}$$

For the sake of presentation, two row vectors are introduced associated with route flows:

$$\mathbf{f}(t) = \left(f_k^{rs}(t) : k \in K_{rs}, r \in R, s \in S \right) \tag{6.2}$$

$$\mathbf{f} = \left(\mathbf{f}(t) : t = 0, 1, 2, \cdots, T-1 \right) \tag{6.3}$$

Eqn. (6.2) indicates that the flow on all the used routes k for OD pair rs in the expressway-ramp-arterial network system at departure time t. On the other hand, eqn. (6.3) denotes the route flow on all the departure time during the time period considered.

Given a route flow solution, $\mathbf{f}$ satisfying the flow conservation eqn. (6.1), referred to as a feasible route flow solution, the actual route travel time on route, $k \in K_{rs}$ for OD pair rs for flows departing at time t should be a function of the departure time, t and the feasible route flow solution, denoted by $\eta_k^{rs}(\mathbf{f}, t)$. The objective of the ramp metering operations optimization model is to minimize the total travel time as shown in the following equation by deciding the optimal ramp metering rate:

$$F(\mathbf{Z}) = \sum_{\forall rs} \sum_{\forall k} \sum_{\forall t} \eta_k^{rs}(\mathbf{f}, t) f_k^{rs}(t, \mathbf{Z}) \tag{6.4}$$

where $\mathbf{Z} = \left(Z_a(t), a \in A_{on}^+, t = 0, 1, \cdots, T-1 \right)$ is a row vector that denote the dynamic ramp metering rate solution of all on-ramps time dependent metering rates.

6.3 Probit-based Ideal DSUO

In this study, the Probit-based ideal DSUO is adopted to describe the route choice decision of drivers. Since each driver perceives actual route travel time differently, the actual route travel time can be assumed as a random variable that follows some distribution. Probit-based model assumes a normal distribution, while the logit-based model considers Gumbel distribution. There are some limitations for logit-based model. First, the overlapping of routes is not considered. Second, the model assumes that the random error term is independently and identically distributed.

Third, it assumes equal variances among alternative routes provided. However, these limitations could be eliminated if the Probit-based model is adopted.

6.3.1 Fixed Point Formulation

Because of the variations in perception and exogenous factors, actual route travel times are perceived differently by individual driver. It is more realistic to assume that each driver's perceived actual route travel time is a random variable being equal to the actual route travel time plus a random error term, which random error term has mean of zero, namely:

$$C_k^{rs}(t,\mathbf{f}) = \eta_k^{rs}(t,\mathbf{f}) + \xi_k^{rs}(t), k \in K_{rs}, r \in R, s \in S, t = 0,1,2,\cdots,T \tag{6.5}$$

where $E\left[\xi_k^{rs}(t)\right] = 0$.

It is further assumed that all the error terms associated with departure time t for OD pair rs follows a multivariate normal distribution, namely:

$$\left(\xi_k^{rs}(t) : k \in K_{rs}\right) \succ N\left(\mathbf{0}, \Sigma^{rs}(t)\right), t = 0,1,\cdots,T \tag{6.6}$$

where $\Sigma^{rs}(t)$ is the $\left(|K_{rs}| \times |K_{rs}|\right)$ covariance matrix, $|K_{rs}|$ represents the cardinality of set K_{rs} . The probability of a driver perceiving route $k \in K_{rs}$ as the shortest one among routes of OD pair rs can be expressed by:

$$p_k^{rs}\left(\mathbf{C}^{rs}(t,\mathbf{f})\right) = \text{Prob}\left[C_k^{rs}(t,\mathbf{f}) \le C_l^{rs}(t,\mathbf{f}), \forall l \in K_{rs} \text{ and } l \ne k\right] \tag{6.7}$$

where $\mathbf{C}^{rs}(t,\mathbf{f})$ is the vector of all perceived route travel times between OD pair rs with the departure time t , namely:

$$\mathbf{C}^{rs}(t,\mathbf{f}) = \left(C_k^{rs}(t,\mathbf{f}), k \in K_{rs}\right), r \in R, s \in S, t = 0,1,2,\cdots,T-1 \tag{6.8}$$

Eqn. (6.7) denotes that route k being chosen by a driver who is traveling on OD pair rs if the driver perceived that the actual route travel time of route k is the minimum compared to the route travel time of all available alternative routes K_{rs} .

According to the Probit-based DSUO principle (Ran and Boyce 1996a), a route flow pattern, $\mathbf{f}^* \geq 0$ is the DSUO solution if and only if:

$$f_k^{*rs}(t) = p_k^{rs}\left(\mathbf{C}^{rs}(t,\mathbf{f}^*)\right)q_{rs}(t), k \in K_{rs}, r \in R, s \in S, t = 0,1,2,\cdots,T-1 \qquad (6.9)$$

It should be pointed out that a nonnegative route flow vector satisfying eqn. (6.9) implicitly fulfils the flow conservation constraints shown in eqn. (6.1). Define a bounded set and a map as follows:

$$\Omega = \left\{\mathbf{f}\left|f_k^{rs}(t) \leq q_{rs}(t), k \in K_{rs}, r \in R, s \in S, t = 0,1,2,\cdots,T\right.\right\} \qquad (6.10)$$

$$\mathbf{H}(\mathbf{f}):\Omega \to \Omega \qquad (6.11)$$

where $\mathbf{H}(\mathbf{f})$ is the vector function expressed by

$$\mathbf{H}(\mathbf{f}) = \left(p_k^{rs}\left(\mathbf{C}^{rs}(t,\mathbf{f})\right)q_{rs}(t)\left|k \in K_{rs}, r \in R, s \in S, t = 0,1,2,\cdots,T-1\right.\right) \qquad (6.12)$$

Therefore, eqn. (6.9) can be rewritten by the fixed-point formulation:

Find a vector $\mathbf{f}^* \in \Omega$ such that

$$\mathbf{H}(\mathbf{f}^*) = \mathbf{f}^* \qquad (6.13)$$

According to the Brouwer's fixed-point theorem (Nagurney 1993), if all the actual route travel time functions, $\left\{\eta_k^{rs}(t,\mathbf{f}), k \in K_{rs}, r \in R, s \in S, t = 0,1,\cdots,T-1\right\}$, are continuous with respect to the route flow pattern $\mathbf{f}$, then the fixed-point formulation (6.13) has at least one solution. In other words, the Probit-based DSUO traffic assignment problem has a solution. The uniqueness of the Probit-based DSUO route flow solution will be determined by the monotone property of these actual route travel time functions.

Szeto (2003) showed that under certain travel demand when there is horizontal queue blocking the junction, the actual route travel time function may not be continuous w.r.t. flow. This implies no guarantee on the monotonic property of the

actual route travel time function. Hence, the uniqueness of the solutions still remains an open question. In addition, since the approximated solution method is used to solve the DSUO model, it could be expected that the solutions may not unique.

6.3.2 An Approximation Solution Method

One of the main concerns when solving the DSUO is how to define the route choice set. For a small network, all the available routes between each OD pair can be enumerated. However, if the network size is large, such enumeration exercise may take a lot of time and resource. In such cases, one could reduce the route choice set to include only some of the prevailing routes that will considered by the drivers. Specifically, those routes that are long and less attractive are removed because they are less likely to be chosen by drivers. Since not all the routes are considered, the solutions obtained are approximated solutions.

6.3.2.1 Determination of sets of routes $\hat{K}_{rs}, r \in R,\ s \in S$

There are some existing studies on the determination of routes considered by drivers in the network loading process. Bovy and Catalano (2007) proposed a doubly stochastic choice set generation method to find the route choice set required by DSUO. By carrying out a repeated shortest route search, the method guarantees that attractive alternatives will be in the choice set with high probability while unattractive ones will have negligible probabilities. Besides, in an earlier research, Cascetta et al. (1997), Lo and Chen (2000) proposed that a reasonable subset of all the available routes for an OD pair that needs to be considered for a medium-sized network is 4-7 routes. They showed that by considering this number of routes, it is sufficient and representative.

If the number of routes considered is fixed, $\hat{K}_{rs}$-shortest route algorithm could be applied to find the $\hat{K}_{rs}$-shortest route for each OD pair rs. Lim and Kim (2005) proposed an efficient algorithm that can minimize the degree of similarity when finding the $\hat{K}_{rs}$-shortest route. This is guaranteed by defining three parameters in the algorithm, namely overlapped ratio, maximum degree of overlapped and link penalty. However, the $\hat{K}_{rs}$-shortest route algorithm adopted in this study is modified from the original one. Instead of using the length of the overlapped route to calculate the overlapped ratio, the number of links that is overlapped is counted. In such a case, the overlapped ratio is more stable and well defined compared to the original version.

$\hat{K}_{rs}$-Shortest Route Algorithm With Route Overlapped

Let $\tau_{k_a}^{rs}$ is the given cost of link a, $a \in A$ on route k, $k \in \hat{K}_{rs}$ of OD pair rs, O_p is the maximum degree of overlap between links of routes allowable which is predefined. O_z is the link penalty, in which $O_z = \left[\dfrac{1}{O_p} \right]^\alpha$ where α is a positive parameter. Let $ol_{n/k}^{rs}$ is the number of overlapping links between route k and n, where $k, n \in \hat{K}_{rs}$ of OD pair rs, where $n = k-1, k-2, \ldots 1$, and $Op_{n/k}^{rs} = \dfrac{ol_{n/k}^{rs}}{l_k^{rs}}$ is the overlapping ratio between route k and n with l_k^{rs} is the number of links that route k of OD pair rs has. When the overlapped ratio between current found route and the entire previous route that has been found is calculated, $O_{max}^{rs/k}$ is taken as the maximum overlapped degree for the current k route of OD pair rs. In other words, $O_{max}^{rs/k} = \max \left\{ Op_{k-1/k}^{rs}, Op_{k-2/k}^{rs}, \ldots Op_{1/k}^{rs} \right\}$. Define also E_k^{rs} is the link set for the k-th route of OD pair rs which consisted of J number of links, i.e.

$E_k^{rs} = \left\{ a_{k_1}^{rs}, a_{k_2}^{rs}, \ldots, a_{k_j}^{rs}, \ldots a_{k_J}^{rs} \right\}$, $a_{k_j}^{rs}$ indicates the link j for route k of OD pair rs,

where $a_{k_j}^{rs} \in A$ and P^{rs} is the route set for OD pair.

The step-by-step procedure is illustrated as follows:

Step 0: (Initialization) Set a value for O_p, and set $rs = 1$, which indicate the first OD pair.

Step 1: (Shortest route algorithm) With any of the existing shortest route algorithm, find the first shortest route for OD pair rs, set $k = 1$. Add the links that form the route to E_k^{rs}, and add the route set $P^{rs} = \left\{ p_k^{rs} \right\}$.

Step 2: (Link cost update) Examine the links in E_k^{rs}. Update the link cost by applying the penalty on the link. Essentially, $\tau_{k_I}^{rs} = \tau_{k_I}^{rs} \times O_z$, where $\mathbf{I} = \{1, 2, \ldots, J\}$.

Step 3: (Route search and overlapping ratio calculation)

Step 3a: (Proceed to next route) Set $k = k + 1$.

Step 3b: (Find shortest route) With link label setting algorithm, find k-shortest route. Add the links that form the route to E_k^{rs}, and add the route set $P^{rs} = \left\{ p_k^{rs} \right\}$.

Step 3c: Calculate degree of overlap, $Op_{n/k}^{rs} = \dfrac{ol_{n/k}^{rs}}{l_k^{rs}}$, where $n = k-1, k-2, \ldots 1$.

Step 3d: Find the maximum overlapped degree from these routes, $O_{max}^{rs/k}$.

Step 4: (Route convergence test) If $O_{max}^{rs/k} > O_p$, proceed to Step 5. Otherwise, proceed to Step 2.

Step 5: (OD pair convergence test) If $rs = RS$, where all the OD pairs have found the routes, terminate the algorithm. Otherwise, set $rs = rs + 1$, proceed to Step 1.

In Step 1, any algorithms can be adopted to find the shortest route at each iteration, such as the link label setting algorithm and Dijkstra algorithm (Dijkstra 1959). In step 2, it could be seen that the link cost of the previous shortest route is increased by applying the penalty charges. This will deter these links from being chosen as members in the subsequent shortest route search. From Step 3c, the degree of overlap has the value of $[0,1]$, where 0 means no overlapped. If the overlapped degree is 1, means that the route found is exactly the same as the previous one. Hence, O_p is defined such a way to control the number of $\hat{K}_{rs}$-shortest route found. For instance, if only one route is needed, $O_p = 1$. Proper tuning of the parameter is thus very important.

6.3.2.2 Derivation of Covariance Matrix $\hat{\Sigma}^{rs}(t)$, $r \in R$, $s \in S$, $t = 0,1,2,\cdots,T$

Sheffi (1985) has shown the derivation of the covariance matrix for the Probit-based model used when performing the static version of the stochastic user equilibrium model. The concept however could be generalized into the dynamic case shown herein. Considering two routes, k and l, where $k,l \in \hat{K}_{rs}$ of OD pair rs available for drivers to choose from during departure time t, the covariance matrix to describe the overlapped of the routes is presented in the following equation:

$$Cov\left(\xi_k^{rs}(t),\xi_l^{rs}(t)\right) = \theta_{kl}^{rs}(t)\sum_{a \in A} t_a^0 \delta_{a,k}^{rs}\delta_{a,l}^{rs}, \ k,l \in \hat{K}_{rs} \tag{6.14}$$

where $\delta_{a,k}^{rs}$ is a binary parameter which takes value of 1 if route k of OD pair rs comprise link a, and it takes value of 0 if it is not; t_a^0 denotes the free flow travel time of link a. By doing so, eqn. (6.14) actually indicates that the covariance between the perceived travel times of two routes is the summation of the free flow travel time of

the common links that shared by two routes. The $\theta_{kl}^{rs}(t)$ in eqn. (6.14) is interpreted as

the drivers' variance of the perceived travel time over a road segment for both routes

k and l of OD pair rs at departure time t. This parameter may change according to

the different departure time t. It is clear that this study used the link free flow travel

time to compute the covariance matrix for the overlapped route. Historical link travel

time can be used if they are available.

6.3.2.3 Actual Route Travel Time: Determination and Properties

To obtain the actual route travel time on any used route of any OD pair, the

development of a dynamic traffic flow model that can characterize traffic flow

condition is utmost importance. This chapter adopted the Modified Cell Transmission

Model (MCTM) that was elaborated in the previous chapter to imitate the traffic flow

condition. Given the OD demand for the expressway-ramp-arterial network system,

the MCTM is called to calculate the actual route travel time, $\boldsymbol{\eta}$, where

$$\boldsymbol{\eta} = \left(\eta_k^{rs}(t, \mathbf{f}) : k \in K_{rs}, r \in R, s \in S, t = 0, 1, 2, \cdots, T-1 \right)$$ by performing the DSUO

assignment. By adopting the method introduced by Lo and Szeto (2002), the actual

route travel time $\eta_k^{rs}(t, \mathbf{f})$ of flow $f_k^{rs}(t)$ can be computed through cumulative counts

at the origin cell r and destination cell s. Figure 6.1 shows the cumulative count

curves, $\chi_k^r(t)$ and $\chi_k^s(t)$ on route k between OD pair rs. ω denotes the arrival time

of the packet of traffic. The actual route travel time experienced by traffic departing at

time t $\left(f_k^{rs}(t) \right)$, is the horizontal distance between the two cumulative curves as

shown in the figure. Since time is discretized, there is no guarantee that the entire

packet of flow will arrive at the destination in the same clock tick of time. Hence, the

travel time averaging scheme is introduced so that the entire departing traffic has one

uniquely determined average actual route travel time.

Eqn. (6.15) calculates the average actual route travel time:

$$\eta_k^{rs}(t, \mathbf{f}) = \frac{\int_{\chi_k^r(t-1)}^{\chi_k^r(t)} \left[\chi_k^{s^{-1}}(v) - \chi_k^{r^{-1}}(v) \right] dv}{\chi_k^r(t) - \chi_k^r(t-1)} \tag{6.15}$$

where

$$f_k^{rs}(t) = \chi_k^r(t) - \chi_k^r(t-1) \tag{6.16}$$

Eqn. (6.16) defines the departure flow at time t.
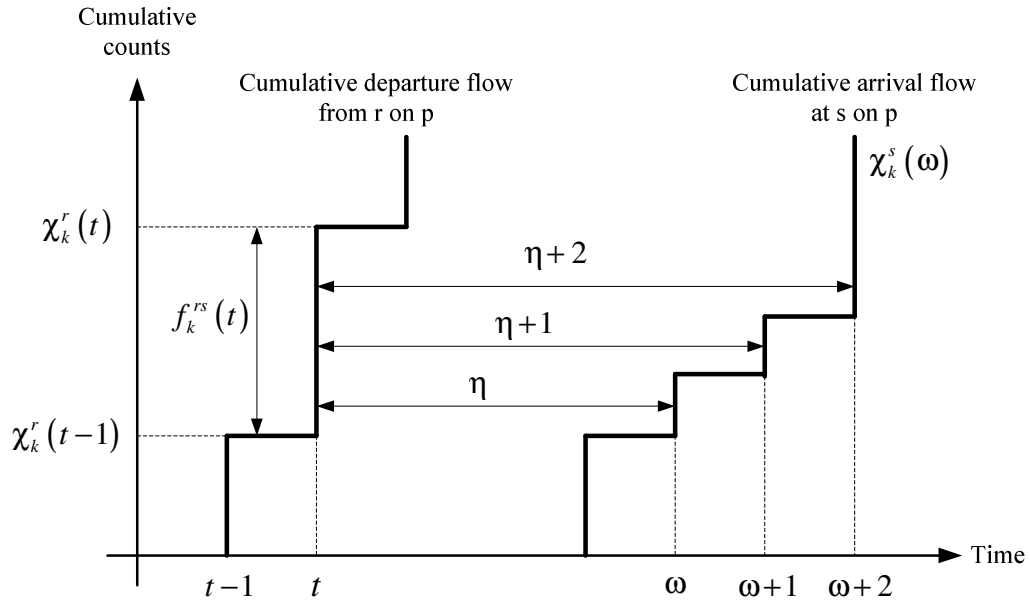
Figure 6.1 The traffic cumulative curve (Source: Lo and Szeto (2002c))

Lo and Szeto (2002c) showed that the actual route travel time function, $\mathbf{\eta}$

obtained using eqn. (6.15) is continuous w.r.t. to the flow, $\mathbf{f}$. However, it is also

shown that the actual route travel time function may not be continuously differentiable

with respect to $\mathbf{f}$. This may cause inexistence of solution for the Probit-based DSUO

assignment. In addition, it is indicated that the average actual route travel time is not

monotonic with respect to the route flow by using CTM. Thus, the solutions (if exist)

may not be unique. Hence, an approximated solution or a tolerance-based solution

could only be found due to this limitation. Szeto and Lo (2005) developed a tolerance-

based deterministic dynamic user optimal (DDUO) principle, which allows the

declaration of the user optimal flow even if the travel times of the used routes between

the same origin-destination pair are not exactly the same. This could be regarded as a

special case to the traditional DDUO.

This tolerance-based concept is applied here for the fixed point problem

indicated in eqn. (6.9). The exact equality of the left hand side and the right hand side

of the eqn. (6.9) could not be obtained in some cases. The solution is declared found if

the difference between both sides is small enough, which can be pre-determined as a

theoretical gap. It is shown in the following equation:

$$\phi = \max\left\{ f_k^{rs}(t) - p_k^{rs}\left(\mathbf{C}^{rs}(t,\mathbf{f})\right) q_{rs}(t) \right\} \tag{6.17}$$

where ϕ is the theoretical gap.

6.3.2.4 Method of Successive Averages (MSA)

There are many methods available for solving the fixed point problem. To

name a few: Ishikawa algorithm, Mann algorithm, Krasnoselskij iterations and so on

(Berinde, 2007). MSA is one of these methods that also can be adopted to solve the

fixed point problem. It is adopted in this study to solve the fixed point formulation of

the DSUO model with a predefined partial routes, $\hat{K}_{rs}$ and the relevant $\left|\hat{K}_{rs}\right| \times \left|\hat{K}_{rs}\right|$

covariance matrix of $\hat{\Sigma}^{rs}(t)$. In addition, Monte Carlo simulation method is adopted

to do the Probit-based ideal DSUO loading, in which the probability of drivers

choosing each route $k \in \hat{K}_{rs}$ is computed.

The step-by-step procedure of MSA is shown as follows:

Step 1: (Initialization)

Step 1a: Define $\hat{K}_{rs}$-shortest route choice set for each OD pair rs based on the link free-flow travel time using the method present in Section 6.3.2.1.

Step 1b: Initialize the departure flow, $f_k^{rs(n)}(t)$ for each route k in the partial route choice set, $\hat{K}_{rs}$ for each OD pair rs at each departure time t randomly. The flow assigned must fulfill the following equation:

$$q_{rs}(t) = \sum_{k=1}^{\hat{K}_{rs}} f_k^{rs(n)}(t) \qquad (6.18)$$

Set $n = 1$.

Step 2: (Calculate actual route travel time) Load the departure flow $f_k^{rs(n)}(t)$ to the MCTM. Run the MCTM to propagate traffic across the cells until all the vehicles arrived at the destinations. Compute the actual route travel time according to eqn. (6.15) for each traffic flow depart at time t for route k of OD pair rs.

Step 3: (Compute probability of route) Using Monte Carlo sampling method, sample the drivers' perceived actual route travel time according to eqn. (6.5), compute the probability of each route of OD pair rs being used by drivers, namely, $p_k^{rs(n)}\left(\mathbf{C}^{rs}(t,\mathbf{f})\right)$.

Step 4: (Compute auxiliary flow, $\tilde{f}_k^{rs(n)}(t)$) Calculate the auxiliary flow based on the probability using the following equation:

$$\tilde{f}_k^{rs(n)}(t) = p_k^{rs(n)}\left(\mathbf{C}^{rs}(t,\mathbf{f})\right)q_{rs}(t) \quad \forall k, rs, t \qquad (6.19)$$

Step 5: (Calculate the flow for next iteration) The flow for next iteration is computed by the following equation:

$$f_k^{rs(n+1)}(t) = f_k^{rs(n)}(t) + \frac{1}{n}\left[\tilde{f}_k^{rs(n)}(t) - f_k^{rs(n)}(t)\right] \quad \forall k, rs, t \qquad (6.20)$$

Step 6: (Convergence checking) If convergence criteria are satisfied, stop. Otherwise,

set $n = n+1$ and proceed to Step 2.

The stopping criteria of the algorithm in Step 6 are based on the calculation of

the average absolute error shown in eqn. (6.21). There are two stopping rules that can

be adopted based on this, one is to stop the algorithm when the ∇ in eqn. (6.21) is

smaller than ϕ in eqn. (6.17), or to stop the algorithm when the ∇ in eqn. (6.21) is

stable over N iterations. In this study, the latter rule is adopted.

$$\nabla = \frac{\sum_{\forall k}\sum_{\forall rs}\sum_{\forall t}\left| f_k^{rs(n)}(t) - p_k^{rs(n)}\left(C^{rs}(t,\mathbf{f})\right)q_{rs}(t)\right|}{\sum_{\forall rs}\hat{K}_{rs}T} \qquad (6.21)$$

A solution exists if the actual route travel time used in eqn. (6.10) is continuous w.r.t.

flow. The solution however is not unique because the actual route travel time is not

monotone w.r.t. route flow. MSA could solve the problem and produce a tolerance-

based solution. Hence, it is stable.

6.4 Optimization Model

The nature of the ramp metering operation allows it to be formulated as a

bilevel programming model, in which the upper level problem aims to find an optimal

metering rate that minimizes the total travel time while the lower level problem

addresses the drivers' route choice decision behavior. Due to the fact that DSUO is

chosen to formulate the lower level problem and it is formulated as a fixed point

problem, the bilevel programming model of the ramp metering operation problem

could be converted into a single level optimization model by adopting the DSUO flow

as a constraint. Hence, given any feasible metering rate solution, $\mathbf{Z}$, the DSUO flow

obtained from the network system is actually subjected to the metering rate used. In

other words, the DSUO flow is a function of the metering rate given. Mathematically, eqn. (6.2) and eqn. (6.3) could be rewritten into:

$$\mathbf{f}(t,\mathbf{Z}) = \left(f_k^{rs}(t,\mathbf{Z}) : k \in \hat{K}_{rs}, r \in R, \ s \in S \right) \tag{6.22}$$

$$\mathbf{f} = \left(\mathbf{f}(t,\mathbf{Z}) : t = 0,1,2,\cdots,T-1 \right) \tag{6.23}$$

The objective of this single level ramp metering optimization model is to find an optimal solution, $\mathbf{Z}$ which can minimize the total travel time of the study area. Since the actual route travel time can be obtained using eqn. (6.15) for each route of each OD pairs, the total travel time for all the drivers in the study area can be computed. Essentially, the total travel time of the drivers in the study area is the summation of all the routes used for all OD pairs during the time period. Mathematically, it is expressed as:

$$F(\mathbf{Z}) = \sum_{\forall rs} \sum_{\forall k} \sum_{\forall t} \eta_k^{rs}(t,\mathbf{f}) f_k^{rs}(t,\mathbf{Z}) \tag{6.24}$$

To ensure stability of traffic flow at a metered on-ramp, the ramp metering rate may vary after multiple consecutive time intervals, but not at every time interval. These multiple consecutive periods form a ramp metering period. More specifically, the discretized time horizon $\left[0,\delta,2\delta,\cdots,(T-1)\delta \right]$ is completely partitioned into N disjoint ramp metering periods, where N is a positive integer, named by $\left(\Delta_{l-1},\Delta_l \right]$, $l = 0,1,2,\cdots,N$, in which $\Delta_0 = 0$ and $\Delta_N = (T-1)\delta$. Having defined these ramps metering periods, the period-dependent ramp metering rate pattern is obtained.

To imitate the traffic responsive type of ramp metering operation, the ramp metering rate should be adjusted according to the average queuing length of the on-ramp in the previous metering period, expressed by,

$$Z_a(t) = \rho_a \times H_a(\Delta_{l-1}), \quad \forall t \in \left(\Delta_{l-1},\Delta_l \right], l = 1,2,\cdots,N, a \in A_{on}^+ \tag{6.25}$$

where $0 \leq \rho_a \leq 1$, referred to as the ramp metering ratio, is a decision variable to be determined, and $H_a(\Delta_{l-1})$ is the average volume of vehicle at the on-ramp in the previous metering period $(\Delta_{l-2}, \Delta_{l-1}]$ that can be calculated by

$$H_a(\Delta_{l-1}) = \frac{\sum_{t \in (\Delta_{l-1}, \Delta_l]} \sum_{i=1}^{l_a} \left[n_{a_i}(t) - \sum_{h_m \in \Gamma^-(a_i)} y_{a_i h_m}(t) \right]}{|\Delta_{l-1} - \Delta_l|} \qquad (6.26)$$

where $n_{a_i}(t)$ is the vehicle volume in cell (a,i) of arc $a \in A$ at time t and $y_{a_i h_m}(t)$ is the vehicle flow on link (a_i, h_m) of the cell-based network, which connects cell (a,i) of arc $a \in A$ to cell (b,i) of arc $b \in A$, from clock tick t to $t+1$. Readers are referred to Chapter 5 to familiar with the notation presented herein. To some extent, eqn. (6.25) gives a dynamic feedback control strategy of traffic flow.

It is now ready to present the single level optimization model for the ramp metering operation:

$$\min F(\mathbf{Z}) = \sum_{\forall rs} \sum_{\forall k} \sum_{\forall t} \eta_k^{rs}(t, \mathbf{f}) f_k^{rs}(t, \mathbf{Z}) \qquad (6.27)$$

subject to

$$Z_a(t) = \rho_a \times H_a(\Delta_{l-1}), \quad \forall t \in (\Delta_{l-1}, \Delta_l], l = 1, 2, \cdots, N, a \in A_{on}^+ \qquad (6.28)$$

$$f_k^{rs}(t, \mathbf{Z}) = \phi\left\{ p_k^{rs}(\mathbf{C}^{rs}(t, \mathbf{f})) q_{rs}(t) \right\}, k \in \hat{K}_{rs}, r \in R, s \in S, t = 0, 1, 2, \cdots, T-1 \quad (6.29)$$

As a brief explanation, eqn. (6.27) denotes the objective function of the optimization model is to minimize the total travel time of the study area with the ramp metering rate as the decision variables. Eqn. (6.28) expresses the ramp metering rate constraint, which can reflect the dynamic feedback control strategy of the traffic responsive type ramp metering rate modeled in this study. Eqn. (6.29) determines the tolerance-based DSUO flow condition.

6.5 Solution Algorithm

A meta-heuristic method is adopted to solve the proposed single level optimization model of the ramp metering operations. Solving the DSUO flow using MSA yields approximated solutions. In view of this, meta-heuristics algorithms could be adopted to solve the optimization model since these algorithms only require the objective function values to judge the solution, without the necessary exact relationship or solution. Genetic algorithm (GA) is chosen as the solution method in light of its wide application in the engineering related problems (Gen and Cheng, 1996).

A dynamic ramp metering rate solution $\mathbf{Z}$ is encoded into a binary string, of which a specified portion represents a metering rate for a particular on-ramp at time t. More specifically, the metering ratio, $\{\rho_a, a \in A_{on}^+\}$ for the period-dependent ramp metering rate schemes shown in eqn. (6.25) is encoded as the binary strings. The strings then are decoded to find the corresponding dynamic ramp metering rate solution by means of eqn. (6.25). This is then embedded into MCTM to compute the corresponding actual route travel time.

The step-by-step procedure is shown as follows:

Step 0. (Initialization) Randomly generate a population of B strings.

Step 1. (Decoding) For each string, decode the string to get the dynamic time dependent ramp metering rate for the on-ramps of the expressway-arterial network, $\mathbf{Z}$.

Step 2. (Call MSA) Decode each string b in the population and perform the DSUO assignment by implementing the procedures shown in Section 6.3.2.4.

Step 3. (Fitness function calculation) Compute the value of fitness function defined by eqn. (6.24) for each string.

Step 4. (Generation of a new population) Repeat the following four sub-steps until the new population is completed.

Step 4.1. (Selection) According to the fitness functions value evaluated in Step 2, use the rank selection method to choose two parent strings from the population.

Step 4.2. (Crossover) With a crossover probability, denoted by p_c, cross over the parents to form a new offspring according to the one point crossover method . If no crossover is performed, offspring is the exact copy of the parents.

Step 4.3. (Mutation) With a mutation probability, denoted by p_m, mutate new offspring at selected position in string.

Step 5. (Stopping criterion) If a stopping criterion is fulfilled, terminate the algorithm. Output the best solution from the population. Otherwise, go to Step 2.

There are two termination criteria that can be adopted in Step 4 of the GA. The GA can be terminated when it achieves maximum number of generations specified or if there is no improvement in the fitness function value for more than the number of generations specified. Note that the performance of the GA may rely on the size of population and crossover and mutation probabilities.

6.6 Numerical Example

An illustrative expressway-ramp-arterial network system shown in Figure 6.2 is used to test the applicability of the proposed methodology. It is a hypothetical network based on the I210W expressway-ramp network in Pasadena, California (Munoz et al. 2004). It is 22-km long and consisted of 21 on-ramps and 18 off-ramps.

154

From the figure, it is observed that the expressway mainline is connected to the arterial roads through the on-ramps and off-ramps. The arterial roads serve as an alternative road to the expressway-ramp system. The origins and destinations nodes are labeled in the figure. In addition, dummy links are created to load the demand from these origins into the network. The cell-based network created for MCTM is shown in Figure 6.3. It is shown that the origin cells serve as the big parking lots that store all the vehicles that will be loaded into the network system. The vehicles are loaded into the network through dummy links/cells. Note that the length of the arterial roads in Figure 6.3 is not to the scale.

The input parameters for the MCTM is such as follows - the free flow speed for expressway mainline section, ramps, and arterial roads are 100km/hr, 60km/hr, and 60km/hr respectively. The backward shock wave speed is 28km/hr, and the jam densities are 17 veh/km/lane for the expressways, ramps, and arterial roads. The size of a cell on the expressway mainline segments, ramps, and arterial roads are 0.28km, 0.17km, and 0.17km respectively. The time interval δ is 10 seconds and the time dependent ramp metering period is 5 minutes. The capacity of the expressway is set to be 2200 veh/hr/lane while the arterial roads capacity is 2000 veh/hr/lane. The ramp capacity is 2000 veh/hr for one-lane ramp with speed of 60km/hr and 2400 veh/hr for two-lane ramp. This is obtained by referring to the Highway Capacity Manual (HCM 2000).
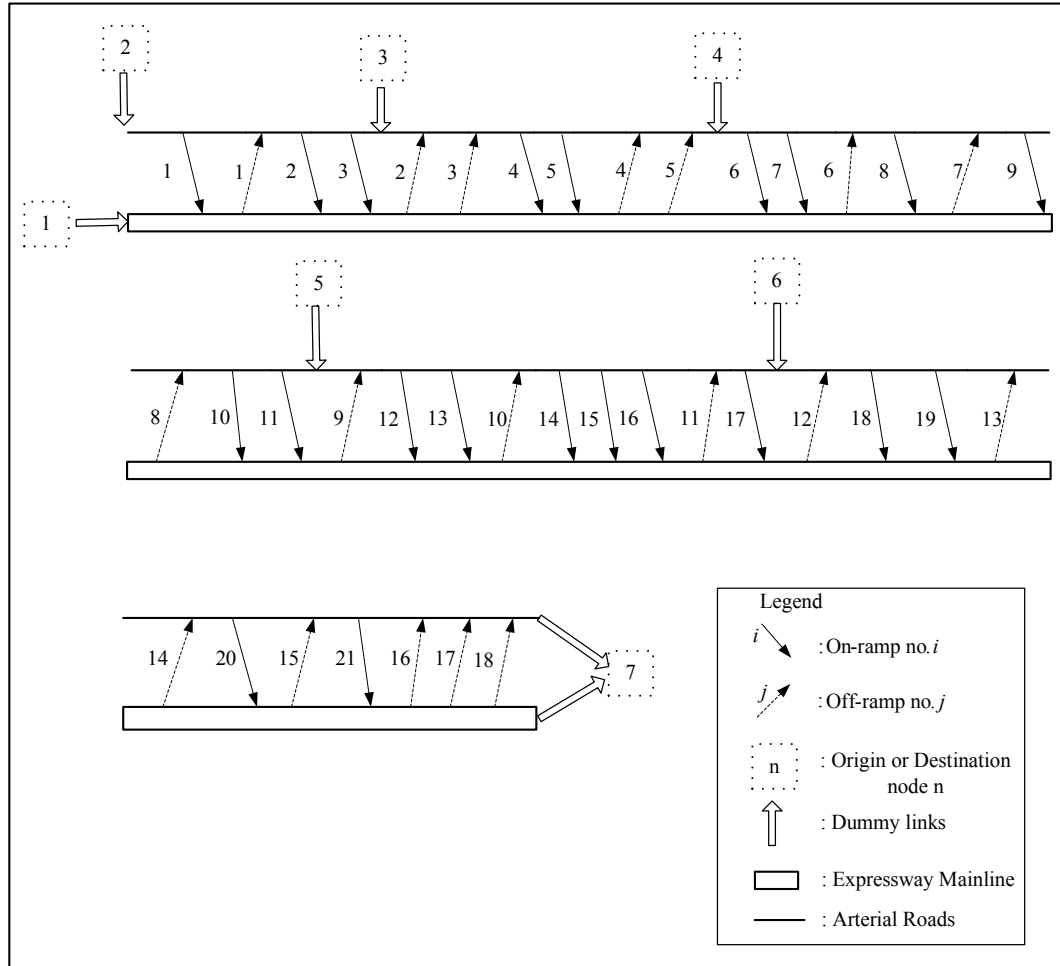
Figure 6.2 The hypothetical expressway-arterial network system
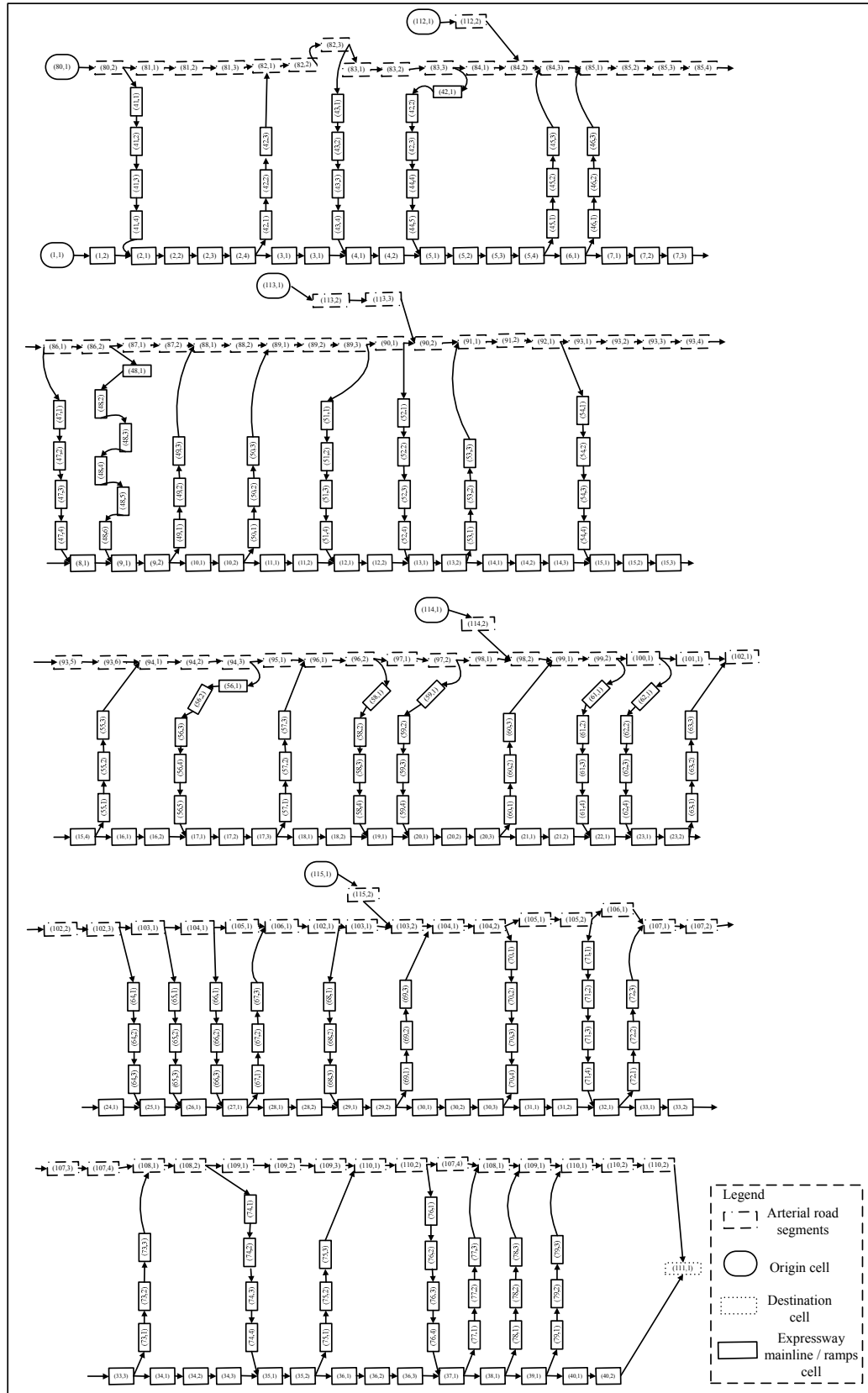
Figure 6.3 The cell-based network for Figure 6.2

Assume that there are 6 pairs of origin-destination (OD) trips in the network system with assumed demand value. These OD pairs and their demand are presented in Table 6.1. The departure time of the drivers is fixed and is carried out during the first 10 time step of the assignment period. The simulation is run until all the vehicles arrived at the destinations, which takes about 40 minutes of the simulation time. The k-shortest route between each OD pair is found using the algorithm proposed in section 6.6.1. The shortest route size is shown in Table 6.1 using $\hat{K}_{rs}$-shortest route algorithm presented in section 6.3.2.1. The value for the parameters is set such as follows: $O_p = 0.5$ and $\alpha = 2$.

For the genetic algorithm setting, each ramp metering ratio in eqn.(6.25), $0 \leq \rho_a \leq 1, a \in A_{on}^+$, is encoded by a 7-bit binary chromosome. With a total of 21 on-ramps, the length of a chromosome is 147 genes in the GA embedding with the MCTM. In addition, the population size is 50 and the algorithm is terminated when the objective function value does not change for 3 generations, and the crossover probability and the mutation probability are 0.7 and 0.03, respectively.

For the Monte Carlo sampling, the value for $\theta_{kl}^{rs}(t)$ shown in eqn. (6.14) is taken as 10 and is identical for all overlapped route of all OD pairs at all departure time. The sampling process is run for 200 iterations before the probability is computed.

Table 6.1 The OD pair and the route number

OD Pair	Origin Node	Destination Node	Route Number	Demand (veh/hr)
1	1	7	20	750
2	2	7	20	500
3	3	7	12	200
4	4	7	14	200
5	5	7	10	155
6	6	7	10	195

6.6.1 Results

Using the proposed strategy, the result for the hypothetical network is shown in Figure 6.4. It can be seen that GA terminates at the 22nd generation when the objective function values does not change for 3 subsequent generations. The final objective function value is 61.96 veh-hr. Compared to the initial solution of 62.78 veh-hr, this is a 1.32% saving in total travel time. The improvement is small because a relatively light demand (2000 veh/hr) is used. The resulted ramp metering rate obtained is shown in Figure 6.5. Due to the small improvement observed, the algorithm is rerun with a heavier demand (3000 veh/hr) loaded onto the network. It is found that there is about 11.25% of improvement in the total travel time if compared to the initial solution. Figure 6.6 shows the result. These results show that the proposed methodology is feasible and can find a better solution compared to the initial solution.
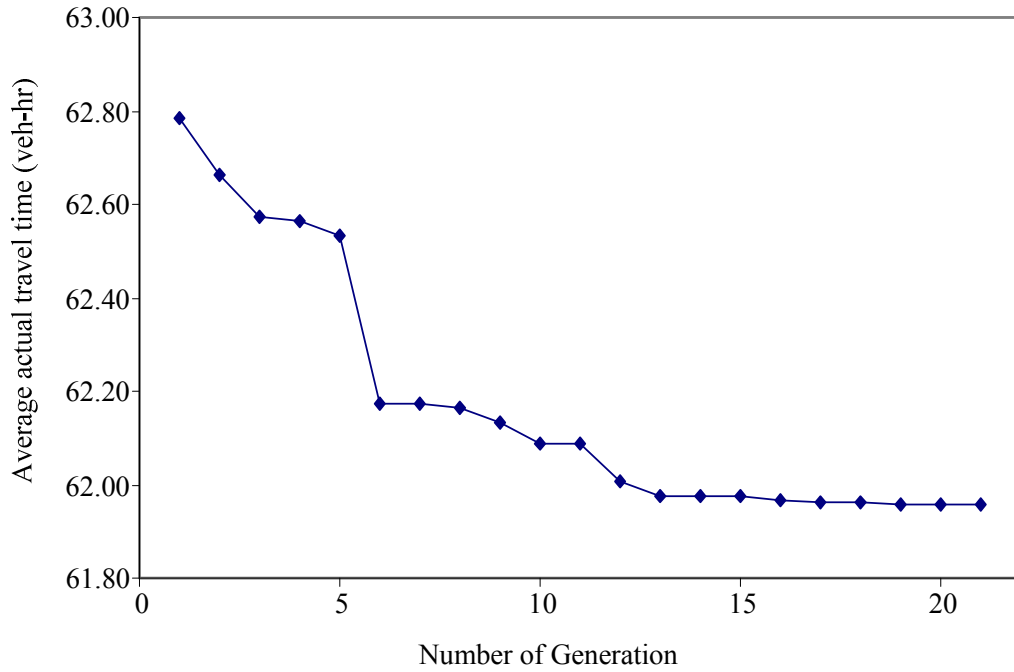
Figure 6.4 The convergence trend of the GA-DSUO
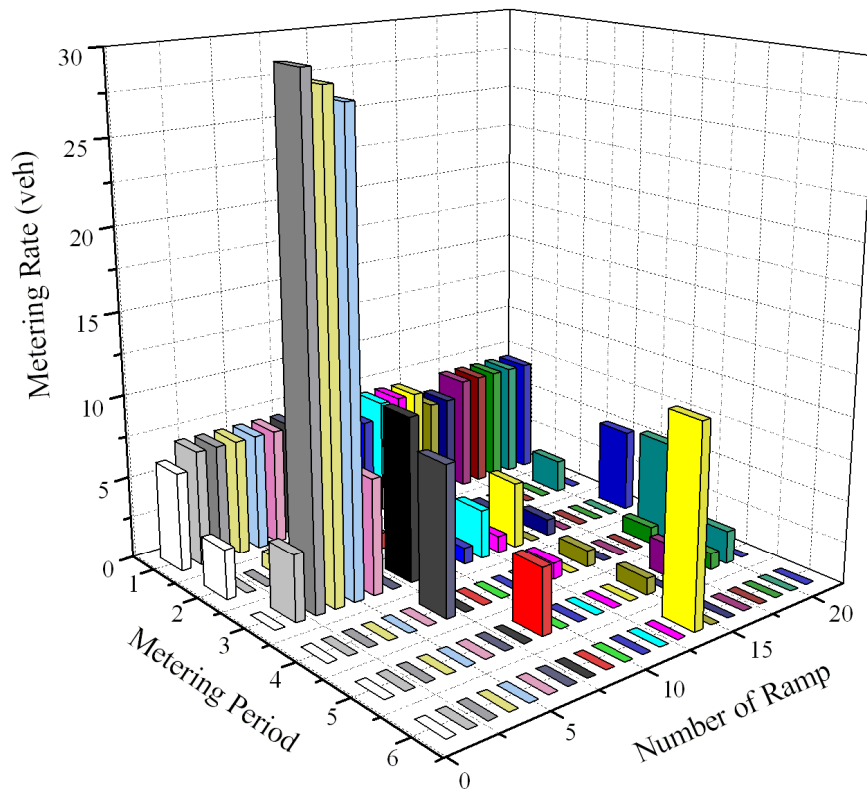
Figure 6.5 Metering rate solution
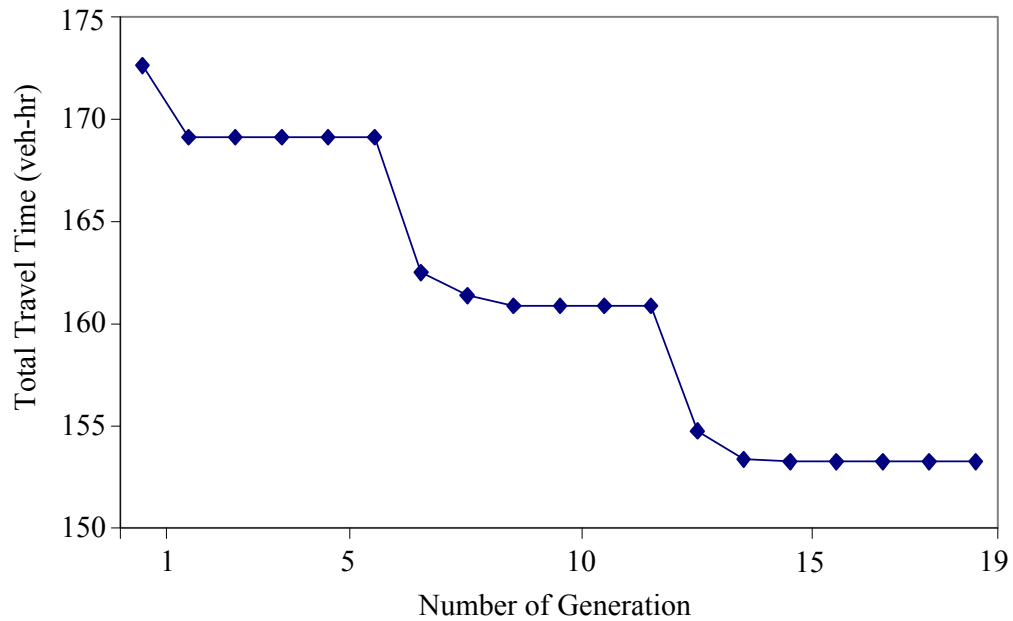
Figure 6.6 The convergence trend of the GA-DSUO for heavier demand

MSA Results

Using the best time-dependent ramp metering rate, i.e. the metering rate that gives the lowest total travel time, the DSUO flow is computed by performing the MSA as proposed in section 6.6.2. When the average absolute error of the algorithm, ∇ does not change for 3 subsequent iterations, the algorithm is terminated and the results are collected. The average absolute error shown in the figure is 1 vehicle. The trend of MSA over iterations is shown in Figure 6.7. It is worth noting that, the maximum absolute error observed is 22 vehicles, which occurs for the vehicles loaded during the last departure time.
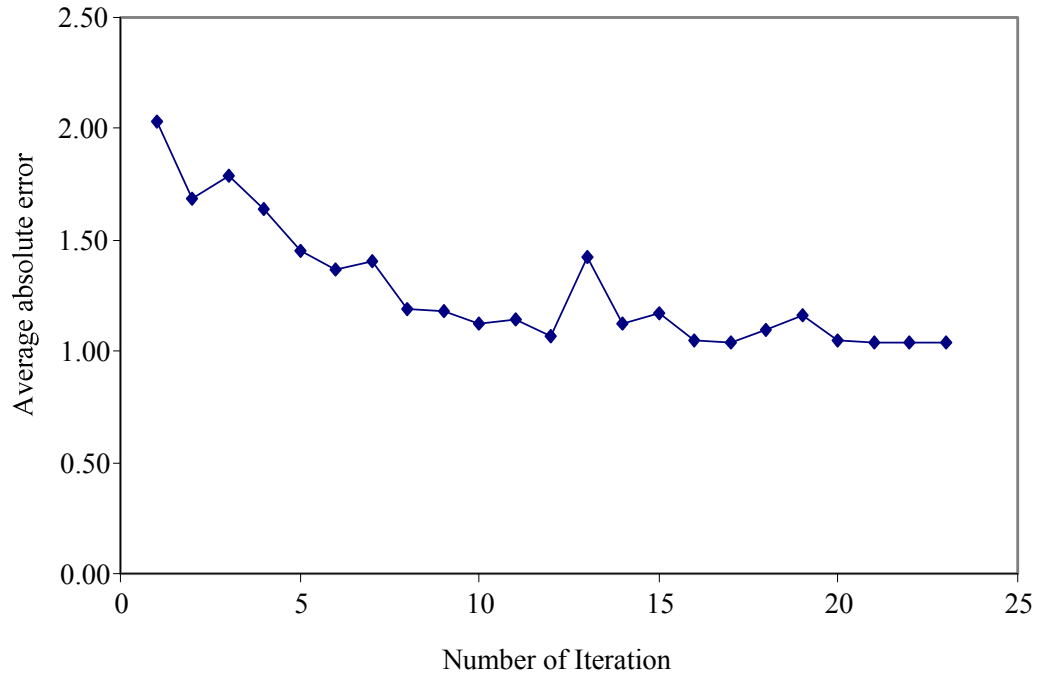
Figure 6.7 The approximated converging pattern for MSA

A closer investigation on the comparison of the DSUO flow obtained for the with-ramp metering case and the without-ramp metering case show that there is a shift of traffic flow from the expressway-ramp system to the arterial streets. Taking OD pair number 2 as an example, it has 20 routes for drivers who would like to travel from origin node 2, which is on the arterial road to node 7, which is on the expressway. For the 20 routes given, drivers have different choices of the on-ramps to choose from to enter the expressway. Table 6.2 presents the entry ramps for each route and the DSUO route flow obtained for the 'with ramp metering' and 'without ramp metering' case. The values shown in columns 3 and 4 of Table 6.2 are the summation of the vehicles departing from node 2 using respective route over the period of time. Note that although some of the routes have the same on-ramp entry points, the routes are different. For example, once drivers enter on-ramp 1 using route 1, they will travel using the expressway mainline to the destinations. Nevertheless, for route 4, drivers will exit the expressway mainline using the off-ramps located at the downstream of

the network system, and continue their journey to the destination using the arterial roads. Similarly, drivers using route 4 exit the expressways through different off-ramp compared to route 6. The same explanation also goes to other routes. Besides, an interesting point to note from Table 6.2 is that, no drivers choose to use route 2. This is because route 2 is an arterial road that connects between 2-7. No drivers perceived it as the shortest route at any of departure time because the design speed of the arterial roads are lower than that set for the expressways.

From Table 6.2, it could be seen that before the implementation of ramp metering, about 60% of the drivers choose to use on-ramps 1 and 2. However, when the ramp metering is applied, it could be seen that, a significant number of drivers have shifted to other routes. A 5% increment in the number of drivers could be observed for routes 7 and 8, while other routes have slight increment in users. This increment is accompanied by a 15% drop in the number of drivers using route 1 and 3. This shows that ramp metering has influenced the route choice decision of drivers. The results found is consistent with Wu (2001) findings. When it is applied, some of the drivers choose other routes, or other on-ramps to enter the expressway mainline. This has confirmed the argument made earlier in this study to support the importance of the incorporation of route choice model (DSUO) in the ramp metering study. The total travel delay before the ramp metering implementation is 109.1 veh-hr, while the best ramp metering solution yields a total travel time of 61.7 veh-hr. This is a saving of 76% in total travel time.

Table 6.2 Comparison of DSUO flow for the with and without ramp metering case

Route number [1]	Entry on-ramp [2]	Exit off-ramp [3]	DSUO flow without ramp metering [4] (veh)	DSUO flow after ramp metering [5] (veh)	Change [5]-[4] (veh)
1	1	N.A.	196	166	-30
2	N.A.	1	0	0	0
3	2	N.A.	98	60	-38
4	1	6	0	5	5
5	4	7	32	35	3
6	1	9	2	6	4
7	5	3	25	45	20
8	6	10	40	52	12
9	2	16	1	5	4
10	4	17	2	5	3
11	4	10	19	19	0
12	3	13	13	15	2
13	7	9	9	10	1
14	19	N.A.	3	9	6
15	3	11	11	11	0
16	5	15	2	4	2
17	17	N.A.	13	13	0
18	5	16	16	16	0
19	4	12	17	17	0
20	8	13	1	7	6
Total			500	500	

6.7 Summary

A ramp metering study that considers origin-destination trip information has been considered in this study. A non-linear single level optimization model is developed to minimize the total travel time of the expressway-ramp-arterial network system while searching the optimal ramp metering rate. A Probit-based dynamic ideal stochastic user optimal (DSUO) assignment model is adopted to describe the drivers' route choice decision behavior in response to the given ramp metering rate. The DSUO assignment model is formulated as a fixed point problem and is considered as

one of the constraints of the optimization model. As such, the objectives of both parties, the system managers and the drivers, could be taken into account. Not all of the available routes for an origin-destination (OD) could be considered, especially in the case of a larger size network, where only partial routes, i.e. $\hat{K}_{rs}$ routes are considered in this study. In addition, the utilization of the MCTM model as the traffic flow model has posed some challenges in determining the exact solution of the Probit-based ideal DSUO. As a result, the outcome from the MSA is an approximation. Hence, the optimization model is solved by meta-heuristics, i.e. genetic algorithm which only require the objective function values without the need for exact relationship or solutions, be known.

The illustrative network example presented shows that MSA could solve the Probit-based DSUO assignment model with sufficient accuracy. In addition, GA could solve the optimization model to give a better solution compared to the initial solution. A closer investigation revealed that DSUO could address the drivers' route choice decision well. A significant diversion of drivers from the expressway-ramp system to the arterial road could be observed. Results show that drivers would consider the queuing condition at the on-ramps and choose to enter the expressway-ramp system through the less congested on-ramps. It is therefore necessary for the system managers to consider the drivers route choice behavior during optimization of ramp metering operations.

CHAPTER 7 A FAIR RAMP METERING OPERATION

7.1 Introduction

There is no doubt that ramp metering improves the expressway-ramp network system by increasing the throughput and reducing the total travel delay. However, the strategy leads to inequity, causing drivers who use on-ramps near to the expressway's bottleneck location suffer more delay than those who enter the expressway at the upstream. The ramp metering equity issue is seldom addressed in the past studies. A before-and-after study carried out by Levinson and Zhang (2004) on the ramp metering in Twin Cities of Minneapolis and St. Paul, Minnesota suggested that there is a need to review current metering algorithm to consider the issue of equity. To deal with the ramp metering equity, first, a performance measure that can capture equity has to be derived. This allows the degree of ramp metering equity can be measured quantitatively. Second, a methodology that can address the equity issue as well as the efficiency of the ramp metering is necessary. This allows both issues be optimized simultaneously in finding the feasible solutions for the metering rate.

This chapter proposes a novel equity index to capture the ramp metering equity issue. The equity index takes the ratio of the minimum to the maximum average delay incurred by the on-ramp members in a group. By maximizing the proposed index, the equity of the ramp metering is guaranteed. In fact, the equity index quantifies the qualitative description of the perfect equitable metering defined by Zhang and Levinson (2005), which implies that there is no difference among drivers

whenever/wherever they access the expressway through on-ramps. The equity index reflects the evenness of the average delay experienced by drivers at the on-ramps of a same group.

Another challenge is the balancing of both system efficiency and equity issue. From a system manager's point of view, ramp metering aims to improve the efficiency of an expressway-ramp network system and can be measured by a system efficiency index, such as the total travel delay. However, drivers, who are the users, are selfish and only concerned about their own delay. This is addressed by the equity issue and its index. Therefore, a multiobjective optimization model that considers both issues simultaneously is used to find the optimal solutions. Since it is a multiobjective model, no transformation relationship is required. This can avoid the difficulties encountered by previous studies, in which a valid transformation relationship between both issues is necessary in capturing the equity issue. A ramp metering that can take into consideration both the efficiency and the equity issue is termed as a fair ramp metering operation.

This chapter adopted the MCTM model discussed in Chapter 5 as the dynamic traffic flow model to propagate traffic flow on the expressway-ramp network system. MCTM is employed in view of its advantages in modeling shock wave, horizontal queue propagation and dissipation and FIFO principle. Since MCTM is a numerical approximation model, an analytical expression of the objective function for the multiobjective optimization model could not be obtained. Hence, a meta-heuristic algorithm, namely the non-sorting genetic algorithm (NSGA-II) (Deb 2001; Deb et al., 2002), is incorporated with the MCTM, is adopted to solve the multiobjective optimization model. The solution algorithm can find a set of Pareto-solutions for the fair ramp metering problem. An illustrative case study of the I210W expressway in

Pasadena, California is used to demonstrate the applicability of the proposed strategy. Note that in this chapter, the driver route choice is not considered.

7.2 Ramp Metering Equity Index and the Fair Ramp Metering Problem

Consider an expressway-ramp network system without arterial roads, $G' = (N, A')$ defined earlier in section 5.2, where N is the set of nodes denoting heads and tails of on-ramps or off-ramps, A' is the set of directed arcs denoting on-ramps, off-ramps and expressway stretches, and $A' \subseteq N \times N$. More specifically, the set of arcs, A', is the union of four mutually disjointed sets - A_{on}^+, A_{on}^-, A_{off} and A_E - where A_{on}^+, A_{on}^-, A_{off} and A_E are the sets of metered on-ramps, un-metered on-ramps, off-ramps and expressway stretches, respectively. This definition is identical to section 5.2, and is repeated herein for the sake of completeness.

Let $[0, T]$ denote the time horizon for the dynamic ramp metering operations. For traffic demand loading in the dynamic traffic flow model, the time horizon is discretized into T equal time interval of δ, namely, $[0, \delta, 2\delta, 3\delta, \cdots, (T-1)\delta]$. Let $Z_a(t)$ be the ramp metering rate from clock tick t to clock tick $t + \delta$, where $t \in \{0, 1, 2, \cdots, T-1\}$, for on-ramp $a \in A_{on}^+$, and $\mathbf{Z}$ be a dynamic ramp metering rate solution denoted by a row vector of all on-ramp time-dependent metering rates , namely

$$\mathbf{Z} = \left(Z_a(t), a \in A_{on}^+, t = 0, 1, \cdots, T-1 \right) \tag{7.1}$$

Given a dynamic ramp metering rate solution $\mathbf{Z}$, the total travel delay of all vehicles on arc $a \in A$ experienced from clock tick t to clock tick $t + \delta$ is a function of the

dynamic ramp metering rate solution, denoted by $D_a(\mathbf{C},t)$. Without loss of generality, it is assumed that the time interval $\delta = 1$ for easy of exposition hereafter.

To measure the efficiency of a dynamic ramp metering rate solution $\mathbf{Z}$, the total system travel delay, denoted by $F_0(\mathbf{Z})$, can be used as a system efficiency index. The total system travel delay is the sum of travel delay encountered by all the vehicles when traveling on the expressway network system over the time horizon $[0,T]$, shown as:

$$F_0(\mathbf{Z}) = \sum_{a \in A'} \sum_{t=0}^{T-1} D_a(\mathbf{Z},t), \qquad a \in A' \qquad (7.2)$$

where $D_a(\mathbf{C},t)$ is the travel delay encountered by vehicles traveling on arc $a \in A'$ of the expressway-ramp network system at time interval t.

The average delay of vehicles experienced at on-ramp $a \in A_{on}^+ \cup A_{on}^-$ over the time horizon $[0,T]$, denoted by $\bar{D}_a(\mathbf{C})$, can be calculated by

$$\bar{D}_a(\mathbf{C}) = \frac{\sum_{t=0}^{T-1} D_a(\mathbf{C},t)}{V_a(T)} \qquad (7.3)$$

where $V_a(T)$ is the number of vehicles that experience at least one time interval delay at on-ramp $a \in A_{on}^+ \cup A_{on}^-$ over the time horizon $[0,T]$. Here, travel delay (in veh-hr) is defined as the extra time drivers spend for the queuing or stopping. It essentially sums up the queuing and waiting times. Therefore, only vehicles that suffer at least one time interval delay will be counted in the denominator of the right hand side term in eqn. (7.3). This manipulation is to avoid under-estimation of the average travel delay.

In addition to the efficiency of a dynamic ramp metering solution, the equity index defined in this study adopted the concept of even distribution of average on-

ramp travel delay for several neighboring on-ramps should be examined from the drivers' point of view. The on-ramps in the expressway-ramp network system is classified into K groups, where K is a positive integer, and each group is represented by set $\bar{A}_{on}^k$, $k = 1, 2, \cdots, K$. Certainly, $\bar{A}_{on}^k$ is a subset of the set of all on-ramps, i.e., $\bar{A}_{on}^k \subseteq A_{on}^+ \bigcup A_{on}^-$. Given a dynamic ramp metering rate solution $\mathbf{Z}$, the ramp metering equity index for a group of on-ramps can be defined by the ratio of the minimum and maximum average travel delay incurred by the on-ramp members in the group, namely:

$$I_k\left(\mathbf{C}\right) = \frac{\min\limits_{a \in \bar{A}_{on}^k}\left\{\bar{D}_a\left(\mathbf{C}\right)\right\}}{\max\limits_{a \in \bar{A}_{on}^k}\left\{\bar{D}_a\left(\mathbf{C}\right)\right\}}, k = 1, 2, \cdots, K \tag{7.4}$$

According to eqn. (7.4), it follows that

$$0 \leq I_k\left(\mathbf{C}\right) \leq 1, k = 1, 2 \cdots, K \tag{7.5}$$

It can be seen that, the average travel delay for each on-ramp in the on-ramp group k is the same if $I_k\left(\mathbf{Z}\right) = 1$. In other words, a dynamic ramp metering rate solution with the equity index of $I_k\left(\mathbf{Z}\right) = 1$ exhibits the perfect equity for drivers using on-ramps within the same group. More interestingly, the nearer the equity index to one, the system ensures a fairer situation. Hence, the index $I_k\left(\mathbf{Z}\right)$ is competent for quantifying the equity issue arising from ramp metering operations.

A fair ramp metering system should impose a ramp metering rate solution with the equity index, defined by eqn. (7.4), as large as possible in order to ensure the equity of the system. While the equity is captured, efficiency of the ramp metering cannot be sacrificed. The fair ramp metering problem thus aims to find a dynamic ramp metering rate solution that can simultaneously minimize the efficiency index expressed by eqn. (7.2) and maximize the equity indexes defined by eqn. (7.4)

associated with several on-ramp groups. Multiple objectives of the fair ramp metering problem may lead to various Pareto-ramp metering solutions.

7.3 Mathematical Model

To formulate the proposed fair ramp metering problem, the multiobjective modeling technique is employed because both efficiency and equity as the multiple objectives should be taken into account. In addition to the multiple objectives, some necessary practical constraints for a feasible ramp metering solution should be elaborated as well.

7.3.1 Constraints for ramp metering rates

To implement the ramp metering in practice, imposing some restrictions on the ramp metering rate solution is necessitated. Although emphases on these restrictions are different, they can be formulated as constraints in the model. Prior to a unified expression for these constraints, three typical ramp metering rate scheme are analyzed. The first constraint is similar to those presented in the previous chapter.

To ensure stability of traffic flow at a metered on-ramp, the ramp metering rate may vary after multiple consecutive time intervals, but not at every time interval. These multiple consecutive periods form a ramp metering period. More specifically, the discretized time horizon $\left[0, \delta, 2\delta, \cdots, (T-1)\delta \right]$ is completely partitioned into M disjoint ramp metering periods, where M is a positive integer, named by $\left(\Delta_{l-1}, \Delta_l \right]$, $l = 0, 1, 2, \cdots, M$, in which $\Delta_0 = 0$ and $\Delta_M = (T-1)\delta$. Having defined these ramps metering periods, the period-dependent ramp rate pattern is obtained.

Period-dependent ramp metering rate scheme governed by the average queuing lengths

In this scenario, ramp metering rate in the current metering period for a metered on-ramp is governed by the average queuing length on the on-ramp in the previous period, namely:

$$Z_a(t) = \rho_a \times H_a(\Delta_{l-1}), \quad \forall t \in (\Delta_{l-1}, \Delta_l], l = 1, 2, \cdots, M, a \in A_{on}^+ \tag{7.6}$$

where $0 \leq \rho_a \leq 1$, referred to as the ramp metering ratio, is a decision variable to be determined, and $H_a(\Delta_{l-1})$ is the average number of vehicles at the on-ramp in the previous metering period $(\Delta_{l-2}, \Delta_{l-1}]$ that can be calculated by:

$$H_a(\Delta_{l-1}) = \frac{\sum\limits_{t \in (\Delta_{l-1}, \Delta_l]} \sum\limits_{i=1}^{l_a} \left[n_{a_i}(t) - \sum\limits_{h_m \in \Gamma^-(a_i)} y_{a_i h_m}(t) \right]}{|\Delta_{l-1} - \Delta_l|} \tag{7.7}$$

To some extent, eqn. (7.7) gives a dynamic feedback control strategy of traffic flow.

Period-dependent ramp metering rate scheme with the reserve effective capacity split ratios

To mitigate traffic congestion on an expressway, limiting number of vehicles on a metered on-ramp to enter the expressway stretch is a reasonable ramp metering strategy. It can be carried out by imposing a friction of the dynamic reserve receiving capacity of the downstream expressway stretch as the ramp metering rate set at each ramp metering period, namely:

$$Z_a(t) = \mu_a \times R_{c_i}(\Delta_{l-1}), \quad \forall t \in (\Delta_{l-1}, \Delta_l], l = 1, 2, \cdots, M, a \in A_{on}^+ \tag{7.8}$$

where $0 \leq \mu_a \leq 1$ is known as the effective receiving capacity split ratio, and $R_{c_i}(\Delta_{l-1})$ is the reserve receiving capacity defined by eqn. (5.3) for merging cell (c, i) of

expressway stretch $c \in A_E$ at which the metered on-ramp $a \in A_{on}^+$ meets the expressway stretch.

Eqn. (7.8) indicates that the number of vehicles at on-ramp $a \in A_{on}^+$ to enter the expressway during ramp metering period l cannot exceed $100\mu_a\%$ of the dynamic reserve receiving capacity of the downstream expressway stretch.

Period-dependent constant ramp metering rate scheme

Taking a constant ramp metering rate over a ramp metering period for a metered on-ramp is an easy but practical way to implement the ramp metering operations. This period-dependent constant ramp metering rate solution can be expressed by:

$$\overline{\chi}_a\left(\Delta_l\right) \le Z_a\left(t\right) \le \widehat{\chi}_a\left(\Delta_l\right), \quad \forall t \in \left(\Delta_{l-1},\Delta_l\right], l=1,2,\cdots,M, a \in A_{on}^+ \qquad (7.9)$$

where $\overline{\chi}_a\left(\Delta_l\right)$ and $\widehat{\chi}_a\left(\Delta_l\right)$ are two pre-determined lower and upper bounds of the constant ramp metering rate in the period $\left(\Delta_l,\Delta_{l-1}\right]$.

Unified expressions

In addition to the above three ramp metering rate constraints above, there may be other constraints needed to be imposed in practice. For easy of exposition and without loss of generality, we can utilize the following generic constraints to express any restriction on the ramp metering rate:

$$g_a\left(Z_a\left(t\right),t\right) \le 0, a \in A_{on}^+, t=0,1,\cdots,T-1 \qquad (7.10)$$

where $g_a\left(Z_a\left(t\right),t\right)$ is a function of ramp metering rate and time t at on-ramp $a \in A_{on}^+$

7.3.2 Multiobjective optimization formulation

The total system delay $F_0(\mathbf{Z})$, shown by eqn. (7.2), is a function of a dynamic ramp metering rate solution $\mathbf{Z}$. Value of function $F_0(\mathbf{Z})$ at any given ramp metering rate solution $\mathbf{Z}$ can be evaluated by the MCTM according to the formula:

$$F_0(\mathbf{Z}) = \sum_{t=0}^{T-1} \sum_{a \in A} \sum_{i=1}^{l_a} \left[n_{a_i}(t) - \sum_{h_m \in \Gamma^-(a_i)} y_{a_i h_m}(t) \right] \tag{7.11}$$

where $n_{a_i}(t)$ is the vehicle's flow in cell (a,i) of arc $a \in A$ at time t in unit of veh/hr and $y_{a_i h_m}(t)$ is the vehicle flow of the cell-based network, which connects cell (a,i) of arc $a \in A$ to cell (h,m) of arc $h \in A$, from clock tick t to clock tick $t+1$. Since a vehicle that is not proceeding at each time interval will encounter one time interval extra travel delay, eqn. (7.11) sums up all the travel delay encountered by vehicles over the specified time horizon.

The average travel delay of vehicles incurred at an on-ramp over the time horizon, shown in eqn. (7.3), can be calculated by the MCTM according to the formula:

$$\bar{D}_a(\mathbf{Z}) = \frac{\sum_{t=1}^{T} \sum_{i=1}^{l_a} \left[n_{a_i}(t) - \sum_{h_m \in \Gamma^-(a_i)} y_{a_i h_m}(t) \right]}{V_a(T)}, a \in A_{on}^k, k = 1, 2, \cdots, K \tag{7.12}$$

where $V_a(T)$ is the volume of vehicle that experience at least one time interval delay at on-ramp $a \in A_{on}^+ \bigcup A_{on}^-$ over time horizon $[0,T]$. In reality, eqn. (7.12) determines the average travel delay of the vehicles at on-ramps by dividing the total travel delay of all vehicles with the total number of vehicles with travel delay.

Based on the ramp metering equity index defined by eqn. (7.4), it is straightforward to define a function for a particular group of on-ramps, $k = 1, 2, \cdots, K$, as follows:

$$F_k(\mathbf{Z}) = 1 - I_k(\mathbf{Z}) = 1 - \frac{\min\limits_{a \in \bar{A}_{on}^k}\{\bar{D}_a(\mathbf{Z})\}}{\max\limits_{a \in \bar{A}_{on}^k}\{\bar{D}_a(\mathbf{Z})\}}, k = 1, 2, \cdots, K \tag{7.13}$$

Eqn. (7.13) implies that maximizing the equity index for a specific group of on-ramps is equivalent to the minimization of the function above. According to eqn. (7.5), it follows that

$$0 \le F_k(\mathbf{Z}) \le 1, k = 1, 2, \cdots, K \tag{7.14}$$

However, finding a ramp metering rate solution that minimizes the total system delay expressed by eqn. (7.11) is not adequate to capture the equity of the system. This is clear that such an objective is biased to vehicles traveling on expressway rather than vehicles at on-ramps since traffic demand on expressway is usually much higher than vehicles at on-ramps. Nevertheless, the objective may ensure the efficiency of the system after the implementation of ramp metering. To capture the equity issue for a group of on-ramps, k, we have to seek a ramp metering rate solution that maximizes the equity index $I_k(\mathbf{Z})$ or minimizes function $F_k(\mathbf{Z})$. This is because $I_k(\mathbf{Z})$ defines the degree of equity of the system. This objective may ensure that the ramps within the same group k may have some degree of even distribution of the average delay.

Since the fair ramp metering problem aims to determine a ramp metering rate solution that takes into account both the efficiency (for the benefit of overall system) and the equity (for the benefit of drivers using on-ramps) issue, it can be thus formulated as the multiobjective minimization model:

175

$$\min_{\mathbf{Z}} \begin{bmatrix} F_0(\mathbf{Z}) \\ F_1(\mathbf{Z}) \\ \vdots \\ F_K(\mathbf{Z}) \end{bmatrix} \tag{7.15}$$

subject to

$$g_a\left(Z_a(t),t\right) \le 0, \ \forall \text{ metered onramp } a \in A_{on}^+, t = 0,1,\cdots,T-1 \tag{7.16}$$

$$n_{a_1}(0) = \sum_{t=0}^{T-1} q_a(t), \forall \text{ onramp } a \in A_{on}^+ \bigcup A_{on}^- \tag{7.17}$$

$$Q_{a_2}(t) = q_a(t), \forall \text{ onramp } a \in A_{on}^+ \bigcup A_{on}^-, t = 0,\cdots,T-1 \tag{7.18}$$

$$y_{a_i a_{i+1}}(t) = \min\left\{S_{a_i}(t), R_{a_{i+1}}(t)\right\}, \ \forall \text{ ordinary link } a_i a_{i+1} \in \Omega_{ordinary}, t = 0,1,\cdots,T-1$$

$$\tag{7.19}$$

$$\begin{cases} y_{b_i c_j}(t) = \beta_{b_i c_j}(t) \cdot \min\left\{S_{b_i}(t), R_{c_j}(t)/\beta_{b_i c_j}(t), R_{d_k}(t)/\beta_{b_i d_k}(t)\right\} \\ y_{b_i d_k}(t) = \beta_{b_i d_k}(t) \cdot \min\left\{S_{b_i}(t), R_{c_j}(t)/\beta_{b_i c_j}(t), R_{d_k}(t)/\beta_{b_i d_k}(t)\right\} \end{cases} \tag{7.20}$$

$$\forall \text{ paired diverging links } \left(b_i c_j, b_i d_k\right) \in \Omega_{diverge}, t = 0,1,\cdots,T-1$$

$$n_{a_i}(t+1) = n_{a_i}(t) + \sum_{h_m \in \Gamma^+(a_i)} y_{h_m a_i}(t) - \sum_{h_m \in \Gamma^-(a_i)} y_{a_i h_m}(t), i = 1,2,\cdots,l_a, a \in A, t = 0,1,\cdots,T-1$$

$$\tag{7.21}$$

$$\begin{cases} M_2 \cdot z_{c_i}(t) + \varepsilon \le S_{e_j}(t) + S_{g_p}(t) - R_{c_i}(t) \le M_1 \cdot \left(1 - z_{c_i}(t)\right) \\ z_{c_i}(t) = 0 \text{ or } 1 \\ y_{e_j c_i}(t) = \left[1 - z_{c_i}(t)\right] \cdot S_{e_j c_i}(t) + z_{c_i}(t) \cdot \left[\sum_{\tau=1}^{\tau_{e_j}(t)} y_{e_j c_i \tau}\left(t, \Phi_{e_i g_p c_j}\left(t, R_{c_j}(t)\right)\right)\right] \\ y_{g_p c_i}(t) = \left(1 - z_{c_i}(t)\right) \cdot S_{g_p c_i}(t) + z_{c_i}(t) \cdot \sum_{\tau=1}^{\tau_{g_p}(t)} y_{g_p c_i \tau}\left(t, \Phi_{e_i g_p c_j}\left(t, R_{c_j}(t)\right)\right) \end{cases} \tag{7.22}$$

$$, \forall \text{ paired merging links } \left(e_j c_i, g_p c_i\right) \in \Omega_{merge}, t = 0,1,\cdots,T-1$$

where M_1 is a large constant; M_2 is a large negative constant; and ε is a very small positive constant. For the purpose of illustration, one could imagine these three constants as: $M_1 \to +\infty$; $M_2 \to -\infty$; $\varepsilon \to 10^{M_2}$.

Constraint (7.16) is the generic expression for the ramp metering rate constraints. Constraints (7.17)-(7.22) are an alternative concise representation of the MCTM. Constraint (7.17) indicates that the first cell of an on-ramp at time $t = 0$ stores the total demand that intends to enter the expressway network like a big parking slot (Lo, 1999). Constraint (7.18) implies that the inflow capacity of the second cell of the on-ramp is set to the exogenous dynamic demand. Vehicles will enter the network according to the dynamic demand if space is available in the second cell. These two constraints mimics dynamic traffic demand generated at an on-ramp. The purpose of binary variable $z_i(t)$ involved in eqn. (7.22) is to express the dynamic traffic flow on the pair of merge links in concise mathematical expressions rather than "if-then" logical conditions shown in eqns. (5.7) and (5.14) (Lo, 1999). This is shown as:

$$z_{c_i}(t) = 1 \text{ if and only if } S_{e_j}(t) + S_{g_p}(t) \leq R_{c_i}(t) \tag{7.23}$$

$$z_{c_i}(t) = 0 \text{ if and only if } S_{e_j}(t) + S_{g_p}(t) > R_{c_i}(t) \tag{7.24}$$

Multiobjective minimization model (7.15)-(7.22), in general, does not possess a universal optimal solution due to certain trade-off among the multiple objectives. In fact, the scalar concept of "optimality" cannot be directly applied for a multiobjective optimization model. Because of the seminal work of Edgeworth and Pareto in late nineteenth century, the Pareto-optimality has been utilized to characterize a solution of a multiobjective optimization model (Deb, 2001). A feasible ramp metering rate solution satisfying the Pareto-optimality condition for the multiobjective optimization

model (7.15)-(7.22) is referred to as the Pareto-optimal ramp metering rate solution, which is illustrated thereafter.

7.3.3 Pareto optimal ramp metering solutions

Let Γ be the set of all feasible ramp metering rate solutions for the multiobjective minimization model (7.15)-(7.22), namely

$$\Gamma = \left\{ \mathbf{Z} = \left(Z_a(t), a \in A_{on}^+, t = 0,1, \cdots T-1 \right) \middle| g_a\left(Z_a(t), t \right) \le 0, a \in A_{on}^+, t = 0,1, \cdots, T-1 \right\} \quad (7.25)$$

Any two feasible ramp metering rate solutions are comparable in terms of the Pareto-dominance relation defined by the multiple objective functions shown in eqn. (7.15).

Definition 1: (Pareto-dominance) A feasible ramp metering rate solution $\mathbf{Z}_1$ is said to dominate another feasible ramp metering solution $\mathbf{Z}_2$ if and only if

$$F_k(\mathbf{Z}_1) \le F_k(\mathbf{Z}_2), k = 0,1,2 \cdots, K \quad (7.26)$$

and there is at least one $k \in \{0,1,2,...,K\}$ such that

$$F_k(\mathbf{Z}_1) < F_k(\mathbf{Z}_2) \quad (7.27)$$

In view of multiple objective functions, the Pareto-dominance is a partial mathematical ordering relation between two ramp metering rate solutions. Based on the Pareto-dominance relation, the Pareto-optimality condition for the multiobjective minimization model (7.15)-(7.22) can be defined as follows (see, Steuer, 1986).

Definition 2: (Pareto-optimality) Let $\mathbf{Z} \in \Gamma$ be a feasible ramp metering rate solution:

(i)   The feasible ramp metering solution $\mathbf{Z}$ is said to be *nondominated* regarding a subset $\Gamma' \subseteq \Gamma$ if and only if there is no solution in Γ' which dominates $\mathbf{Z}$.

(ii)   The feasible ramp metering solution $\mathbf{Z}$ is called Pareto-optimal ramp metering rate solution if and only if $\mathbf{Z}$ is nondominated regarding the whole set Γ.

In general, the Pareto-optimal ramp metering rate solution is not unique and it cannot be improved with respect to any objective without worsening at least one other objective. Let Γ^* denote a set of all Pareto-optimal ramp metering rate solutions, and the corresponding vectors of the objective function values, $\left\{ \left(F_0(\mathbf{Z}), F_1(\mathbf{Z}), \cdots, F_K(\mathbf{Z}) \right) \middle| \mathbf{Z} \in \Gamma^* \right\}$, is called Pareto-front.

7.4 Solution Algorithm

The $K+1$ objective functions in the multiobjective minimization model (7.15)-(7.22) do not possess conventional analytical expressions, and calculation of these objective function values at any ramp metering rate solution necessitates the dynamic traffic flow pattern formulated by MCTM or constraints (7.17)-(7.22). For any given ramp metering rate solution, however, MCTM can be solved numerically rather than analytically. These two features determine that an evolutionary algorithm based solution method is the only choice to find the set of Pareto-optimal ramp metering rate solutions because it only requires values of functions involved in the objective and constraints. Among a few categories of evolutionary algorithms, multiobjective genetic algorithm (GA) are the most popular solution methods for solving multiobjective optimization problems (Jones et al. 2002).

The first multiobjective GA was proposed by Schaffer (1985). Subsequently, more than thirteen multiobjective GAs have been developed (Coello, 2000), including niched Pareto genetic algorithm (NPGA) (Horn et al., 1994), random weight-based genetic algorithm (RWGA) (Murata, 1995), non-dominated sorting genetic algorithm (NSGA) (Srinivas and Deb, 1995) and fast non-dominated sorting genetic algorithm (NSGA-II) (Deb et al. 2002). Generally, multiobjective GAs differ based on their

fitness assignment procedures, elitism choice strategies and population diversity mechanisms. According to the recent comparison made by Konak et al. (2006), NSGA-II is an efficient and well-tested multiobjective GA. Therefore, NSGA-II is adopted for solving multiobjective minimization model (7.15)-(7.22) .

NSGA-II operates with a collection of strings, called a population. A string corresponds to a unique solution of a problem of interest in the solution space by a string decoding scheme. For the multiobjective minimization model (7.15)-(7.22), a dynamic ramp metering rate solution $\mathbf{Z} = \left(Z_a(t), a \in A_{on}^+, t = 0,1,\cdots,T-1 \right)$ can be easily encoded into a binary string, of which a specified portion represents a metering rate for a particular on-ramp at time t . More specifically, the metering ratios, $\left\{ \rho_a, a \in A_{on}^+ \right\}$, or the reserve receiving capacity split ratios, $\left\{ \mu_a, a \in A_{on}^+ \right\}$, for the period-dependent ramp metering rate schemes shown in eqns. (7.6) and (7.8) can be encoded as the binary strings. With such a string encoding scheme, the corresponding dynamic ramp metering rate solution can be decoded by means of eqns. (7.6) or (7.8). Given a dynamic ramp metering rate solution decoded from the binary string, MCTM can be solved numerically by incorporating the dynamic ramp metering rate solution. The $K+1$ objective function values with respect to the dynamic ramp metering rate solution can be evaluated accordingly. The NSGA-II embedding with MCTM for solving multiobjective optimization model (7.15)-(7.22) is described below in brief. For details, readers are encouraged to refer to Srinivas and Deb, (1995) and Deb et al. (2002).

7.4.1 NSGA-II embedding with MCTM

Step 1: (Initialization) Randomly create a parent population P_0 of size L in the dynamic ramp metering rate solution space. Set the number of generation, $\omega = 0$.

Step 2: (Generate an initial offspring population) Randomly select string from population P_0 to perform crossover and mutation to generate offspring population B_0 of size L.

Step 3: (Stopping criterion checking) If a stopping criterion is satisfied, stop and return to P_ω. Otherwise, go to Step 4.

Step 4: Set $E_\omega = P_\omega \bigcup B_\omega$.

Step 5: (Invoke MCTM) Decode each string in set E_ω into a dynamic ramp metering rate solution and then run MCTM to evaluate the $K+1$ objective function values corresponding the dynamic ramp metering rate solution, according to eqns. (7.11)-(7.13).

Step 6: (Allocate fitness value) Apply the fast non-dominate sorting algorithm (Deb et al. 2002) to identify all the non-dominated fronts in set E_ω, denoted by $\bar{E}_1, \bar{E}_2, \cdots, \bar{E}_{M_\omega}$ where M_ω is an positive integer, in terms of the $K+1$ objective function values evaluated in Step 4.

Step 7: (Maintain elitist strings) (i) Let $P_{\omega+1} = \phi$ (ii) For fronts $m = 1, \ldots M_\omega$ do following steps:

Step 7.1: (Crowding distance assignment) Calculate the crowding distance* of each string, i.e. the average distance of two solutions on either side of this solution along each of the objective, in the non-dominated front $\bar{E}_m$.

Step 7.2: (Create parent population for next generation) Create $P_{\omega+1}$ as follows:

Case 1: If $\left|P_{\omega+1}\right|+\left|E_m\right| \leq L$, then set $P_{\omega+1} = P_{\omega+1} \cup f_g$

Case 2: If $\left|P_{\omega+1}\right|+\left|E_m\right| > L$, then add the least crowded $L-\left|P_{\omega+1}\right|$ solutions from E_m to set $P_{\omega+1}$.

Step 8: (Crossover) Set $B_{\omega+1} = \phi$ and generate an offspring population $B_{\omega+1}$ of size L as follows:

Step 8.1: (Parent selection with diversity mechanism). Use binary tournament selection method (Goldberg, 1989) based on the crowding distance to select parents from $P_{\omega+1}$.

Step 8.2: Use a crossover operator to generate offspring to add them to set $B_{\omega+1}$.

Step 9: (Mutation) Mutate each sttring in set $B_{\omega+1}$ with a predefined mutation rate. Let the number of generations $\omega = \omega+1$ and go to Step 3.

*Note: The crowding distance is calculated by:

$$E_m[i]_{\text{distance}} = E_m[i]_{\text{distance}} + \frac{\left(E_m[i+1]_{.k} - E_m[i-1]_{.k}\right)}{f_k^{\max} - f_k^{\min}} \qquad (7.28)$$

where $E_m[i]_{.k}$ refers to the k th objective function value of the i th solution in the set of front m. $f_k^{\max}$ and $f_k^{\min}$ are the maximum and minimum values of k th objective function. For detail, please refer to Deb et al. (2002).

7.5 Numerical Example

To numerically evaluate the proposed model and solution method, a stretch of Interstate 210W expressway-ramp network system comprising 21 metered on-ramps and 18 off-ramps of Pasadena, California are adopted. Google Earth is utilized to

estimate length of each on-ramp and identify the number of lanes for the expressways mainline and the ramps. Figure 7.1 schematically depicts the expressway network system with eight mainline segments.

The time horizon of interest is assumed to be one hour and the time interval, δ is assumed to be 10 seconds. Figure 7.2 gives the cell-transmission network encoded for the expressway-ramp network system shown in Figure 7.1. The length of a cell in the cell-transmission network is the distance a vehicle can travel with free flow speed in a time interval. Parameters used by the MCTM - free flow speed for expressway mainline section (100km/hr), free flow speed for ramp (60km/hr), backward shock wave speed (28 km/hr) and jam density (17 vehicles/km/lane) - are taken from Munoz et al (2004) who have calibrated the cell-transmission network. Because free flow speeds of the mainline and the ramps are different, size of the expressway mainline cell (0.28km) is not the same as the ramp's cell (0.17km).
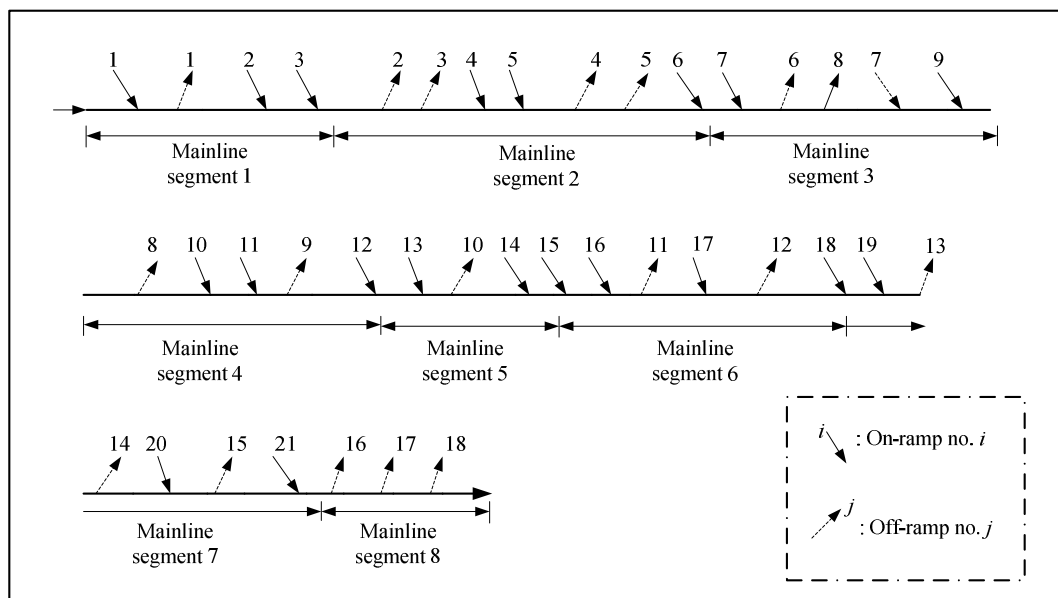

Figure 7.1 Example network with expressway's mainline divided into segments

The average hourly traffic demand on the mainline is assumed as 2000 veh/lane while the average travel demand at an on-ramp is between 720veh/hr and 2000veh/hr depending on the number of lanes the on-ramp has. The two-lane on-ramps have higher traffic demand than one-lane on-ramps. Traffic flow split ratio between the mainline and an off-ramp is arbitrarily assumed between 0.4 and 0.6. Capacity of the expressway mainline is set to 2200 veh/hr/lane and ramp capacity is 2000 veh/hr for one-lane ramp with free flow speed of 60 km/hr and 2400 veh/hr for two-lane ramp according to Highway Capacity Manual (2000). Given an average hourly traffic demand for an on-ramp, it is assumed that dynamic traffic is generated from the first cell of the on-ramp, shown in eqn. (7.17), by obeying the Poisson distribution, i.e. the time headway between two consecutive vehicles follows the exponential distribution.

To investigate impact of various parameters involved in formulating the multiobjective minimization model, a benchmark scenario is configured below. Note that the expressway mainline is divided into eight segments indicated in Figure 7.1, and the eighth segment does not have any on-ramp and each of the other seven segments has three on-ramps respectively. Therefore, three vicinity on-ramps in each segment are classified in a group for the benchmark scenario. In other words, the benchmark scenario possesses seven equity indices, namely $K = 7$.

Regarding the ramp metering constraints, the benchmark scenario applies the period-dependent ramp metering rate scheme governed by the average queuing lengths, shown in eqn. (7.6). The length of each ramp metering period is five minutes, namely the ramp metering rates are updated every five minutes although the time interval in MCTM is set to be 10 seconds. Thus, there are totally twelve ramp metering periods over the one hour time horizon. According to eqn. (7.6), each ramp metering

ratio, $0 \leq \rho_a \leq 1, a \in A_{on}^+$, is encoded by a 7-bit binary string. With a total of 21 on-ramps, the length of a string is 147 genes in the NSGA-II embedding with the MCTM.
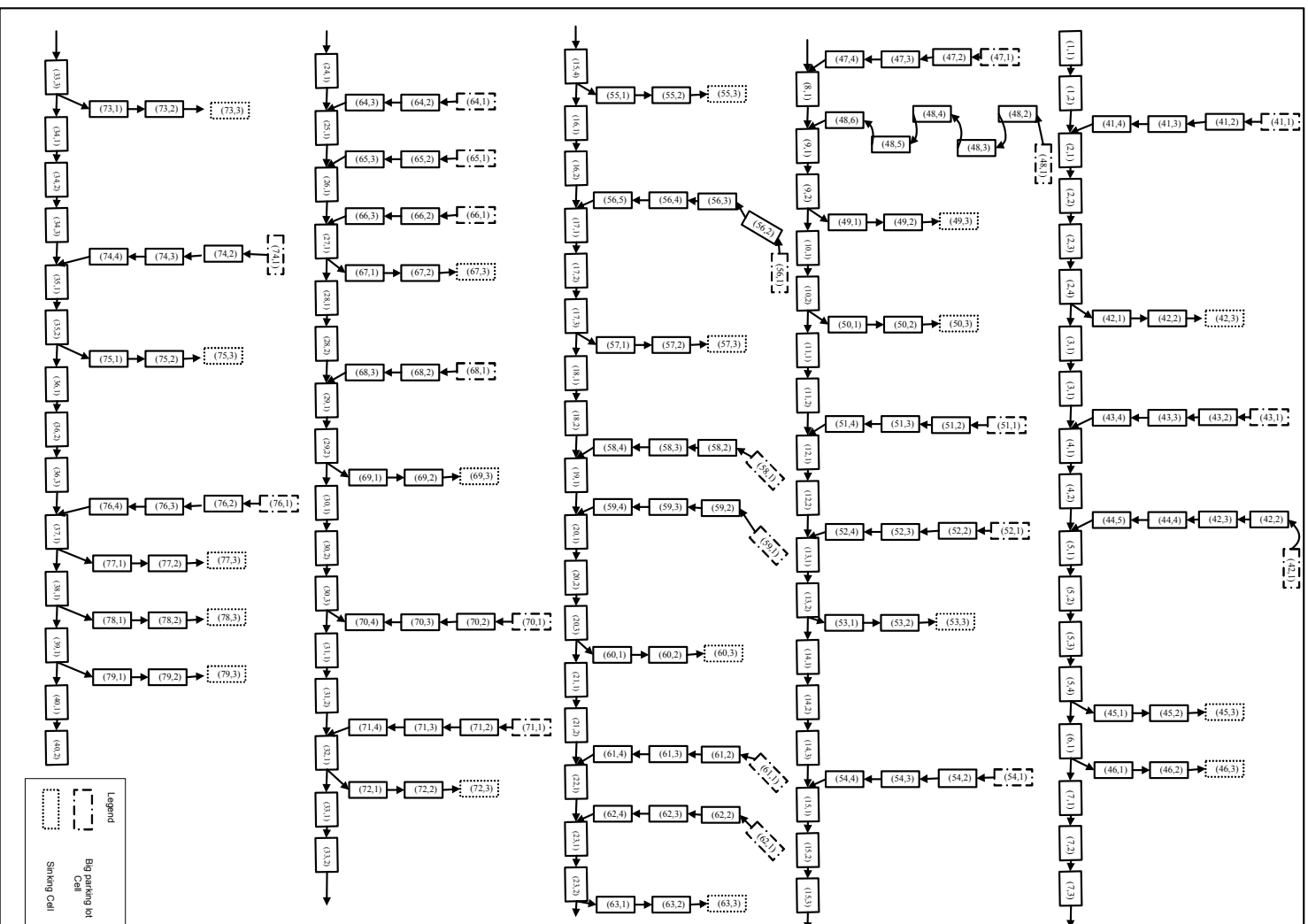
Figure 7.2 Cell-transmission network for I210W

The parameters used by the NSGA-II embedding with the MCTM are set as follows – the population size is 100, the maximum number of generations (i.e., stopping criterion) is 30, and the crossover probability and the mutation probability are 0.7 and 0.03, respectively.

A Pareto-optimal dynamic ramp metering rate solution denoted by $\mathbf{Z}^P$ correspond to the total system travel delay $F_0\left(\mathbf{Z}^P\right)$ and K equity index values, $I_k\left(\mathbf{Z}^P\right), k=1,2,\cdots,K$. For the purpose of comparison among the Pareto-optimal solutions, the average equity index for a Pareto-optimal dynamic ramp metering rate solution is defined below.

$$\overline{I}\left(\mathbf{C}^P\right) = \frac{\sum_{k=1}^{K} I_k\left(\mathbf{C}^P\right)}{K} \tag{7.29}$$

Eqn. (7.29) is an indicator to quantify the average performance of ramp metering equity index among groups. For a particular solution, some groups may have lower equity index value while some may have higher. Hence, a consistent performance indicator is necessary to judge the quality of the results.

7.5.1 Numerical Results for the Benchmark Scenario

There are a total of 89 Pareto-optimal dynamic ramp metering rate solutions obtained for the benchmark scenario by implementing the NSGA-II embedding with the MCTM. However, many of these are close with each other in terms of the total system travel delay (i.e., system efficiency index) and the seven equity indices. Among these 89 solutions, 16 dynamic ramp metering rate solutions which are more distinct with each other, and their total system travel delays and the seven equity indices are presented in Table 7.1. These 16 solutions are sorted in ascending order of the total system travel delay. The best three solutions in terms of the system efficiency are

highlighted in bold font in the table, which yield an average equity index value of 0.613 and system travel delay of 1086.4 veh-hrs. If the values in the table are sorted according to the average equity index, the first three most equitable solutions are actually the last three solutions in their current position in Table 7.1, shown by the bold italic font. The average total system travel delay of these solutions is 1142.9 veh-hrs with the average equity value of 0.777. Based on their locations in Table 7.1, this clearly provides insight on the trade-off relationship of the equity and efficiency issues. In such a case, a 5% reduction in total system benefit is a price pay for an improvement of 21% in equity. This trade-off relationship is to some extent similar to those found by Levinson and Zhang (2004).

Given the large set of the Pareto-optimal dynamic ramp metering rate solution, this offers great flexibility for system managers to decide the trade-off that they could accept. This could be done according to some established guidelines or criteria. Using the proposed methodology, system engineers could actually observe and quantify the price they traded off in order to have a more equitable ramp metering system. Figure 7.3 shows the best period-dependent dynamic ramp metering rate pattern in terms of the average equity index, which can guide traffic control operators to set an appropriate metering rate solution.

Table 7.1 Total system travel delay $F_0\left(\mathbf{Z}^P\right)$ and equity index $I_k\left(\mathbf{Z}^P\right)$ of 16 Pareto-optimal dynamic ramp metering rate solutions for the benchmark scenario

$\mathbf{Z}^P$ no.	$F_0\left(\mathbf{Z}^P\right)$ (veh-hrs)	Equity index $I_k\left(\mathbf{Z}^P\right)$							Average equity index $\bar{I}\left(\mathbf{Z}^P\right)$
		$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$	$k=7$	
1	**1085.6**	**0.289**	**0.285**	**0.794**	**0.774**	**0.678**	**0.681**	**0.800**	**0.614**
2	**1086.2**	**0.304**	**0.285**	**0.794**	**0.774**	**0.678**	**0.681**	**0.800**	**0.616**
3	**1087.5**	**0.289**	**0.285**	**0.794**	**0.783**	**0.614**	**0.700**	**0.797**	**0.610**
4	1088.8	0.289	0.276	0.794	0.801	0.661	0.681	0.797	0.615
5	1089.1	0.288	0.363	0.794	0.768	0.668	0.686	0.797	0.623
6	1093.5	0.289	0.348	0.794	0.777	0.661	0.681	0.797	0.621
7	1100.7	0.289	0.493	0.794	0.777	0.662	0.681	0.797	0.642
8	1105.6	0.582	0.285	0.794	0.774	0.678	0.681	0.797	0.656
9	1108.4	0.391	0.492	0.794	0.774	0.590	0.701	0.797	0.649
10	1111.3	0.322	0.974	0.794	0.774	0.590	0.700	0.797	0.708
11	1112.3	0.286	0.973	0.794	0.777	0.59	0.701	0.797	0.703
12	111.16	0.289	0.974	0.794	0.783	0.662	0.683	0.797	0.712
13	1134.4	0.582	0.898	0.794	0.816	0.662	0.681	0.797	0.747
14	*1132.8*	*0.582*	*0.973*	*0.794*	*0.777*	*0.668*	*0.681*	*0.797*	*0.753*
15	*1147.5*	*0.843*	*0.902*	*0.794*	*0.814*	*0.666*	*0.680*	*0.797*	*0.785*
16	*1148.3*	*0.827*	*0.973*	*0.794*	*0.816*	*0.662*	*0.681*	*0.797*	*0.793*
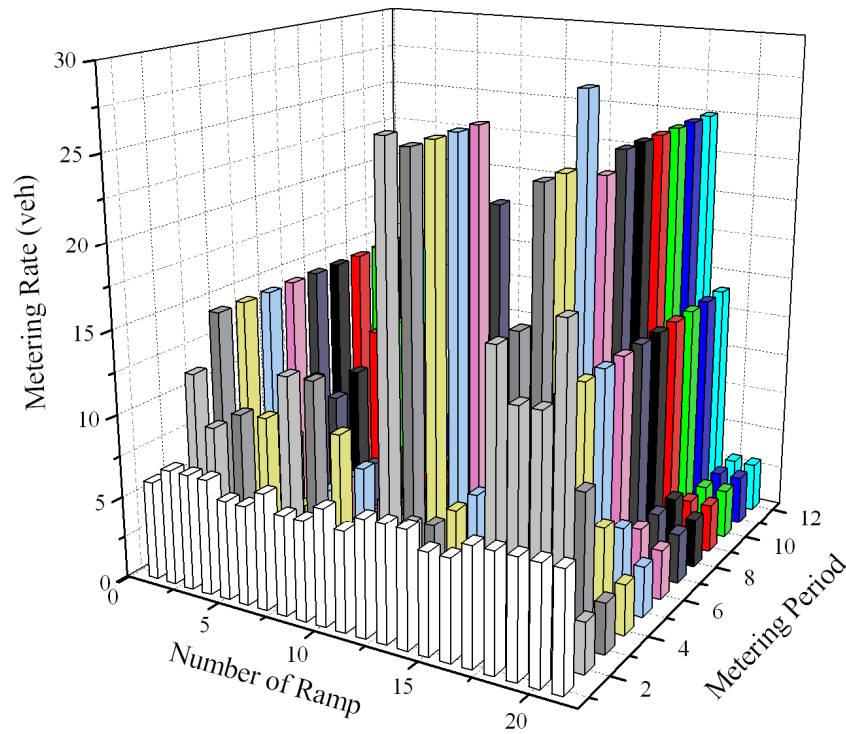
Figure 7.3 Metering rate for benchmark case

7.5.2 Impact of Equity Issue

To further demonstrate the impact of the equity issue, the benchmark scenario is used to numerically compare the system optimal dynamic ramp metering rate solution which minimizes the total system travel delay alone and the Pareto-optimal dynamic ramp metering rate solutions that take into account both system efficiency and the equity issues. The numerical result shows that the Pareto-optimal dynamic ramp metering rate solution with the best equity index yield an average total system travel delay of 9.7% more than the ramp metering that only takes into consideration of the system efficiency issue. Figure 7.4 plots the on-ramp average travel delays for the system optimal solution and the Pareto-optimal solution with the best average equity index.

Figure 7.4 On-ramp average travel delay for the two ramp metering rate solutions

It clearly indicates that for the system optimal dynamic ramp metering solution, the distribution of average travel delay among on-ramps is very uneven. Some on-ramps have larger average travel delay while others have little delay. On the other hand, the Pareto-optimal dynamic ramp metering solution wit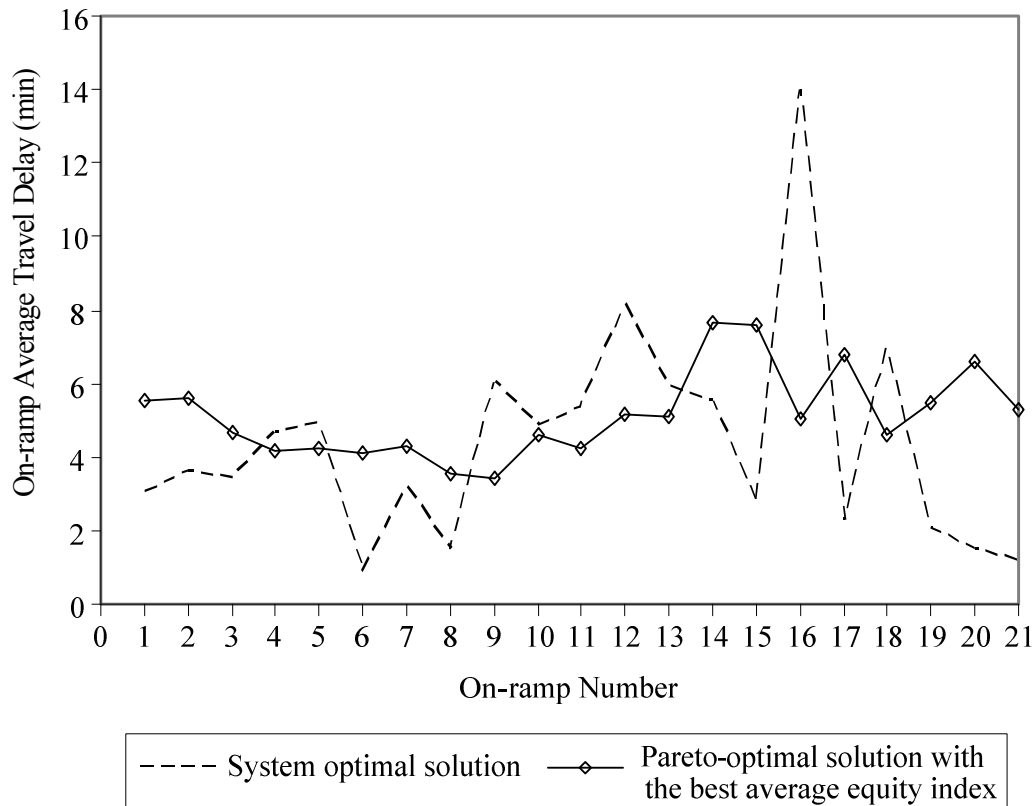h the best average equity index shows that the average ramp travel delay is distributed more evenly among set of ramps. It can be observed that the average travel delay encountered by drivers within same group is within certain range. In addition, for the Pareto-optimal solution, the average waiting time of the drivers on the ramp in average is 5 minutes, with the maximum waiting time of 8 minutes, which are evaluated by using the MCTM. This suggests that by imposing a maximum waiting time for each individual vehicle may be a feasible method to capture the equity issue as well. This finding also proves that the equity index proposed in this study is able to capture the equity issue of ramp metering and that metering is less effective (with higher total delay). For the case where only system efficiency is considered, the average waiting time at on-ramp is 4 minutes, with the maximum average waiting time of 14 minutes.

7.5.3 On-ramp Grouping Effect

A close study on the formulation reveals the fact that, its performance is dependent on the grouping of the on-ramp. Intuitively, one can expect that a smaller number of group members in a group will give a better equity index performance, as shown in Table 7.2. Table 7.2 shows the values for the best three metering rate in terms of efficiency and equity respectively for different number of members in a group. By multiplying the value of average equity index column in Table 7.2 into percentages, one can obtain the degree of equity. It can be seen that for the group of 3 nearby on-ramps, it can achieve an average of 77.7% equity. For a grouping of 4 and 5 members,

7 members, and 10 members, show an average of 73.8%, 63.6%, and 59.1% of equity respectively, taking the average value of the three best solutions in terms of equity for each group of solutions. It can be observed that the percentage of equity deteriorates with increasing members in a group. This is logical since there are many factors, such as geometrical layout and ramp meters arrival rate, affecting the equity characteristics of ramp metering. Fewer members in a group may reduce the variability and thus higher equity may be achieved. In addition, those ramps that are far apart may not have same characteristics and should behave equally. The grouping of the on-ramps could be done in several ways. For example, the groups can be suggested based on the engineering judgment or requirements of the authority. It could also be done by grouping the on-ramps with similar historical arrival rate. Besides, proper optimization models could be developed to find the best grouping that yield the optimal equity value.

Table 7.2 Total system travel delay $F_0\left(\mathbf{Z}^P\right)$ and equity index $I_k\left(\mathbf{Z}^P\right)$ for the Pareto-optimal dynamic ramp metering rate solutions with different on-ramp grouping strategies

On-ramp grouping strategy	$F_0\left(\mathbf{Z}^P\right)$ (veh-hrs)	Equity index, $I_k\left(\mathbf{Z}^P\right)$							Average equity index $\bar{I}\left(\mathbf{Z}^P\right)$
		$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$	$k=7$	
3 on-ramps ($K=7$)									
#1	1085.6	0.289	0.285	0.794	0.774	0.678	0.681	0.800	0.614
#2	1086.2	0.304	0.285	0.794	0.774	0.678	0.681	0.800	0.616
#3	1087.5	0.285	0.285	0.794	0.783	0.614	0.700	0.797	0.610
*1	1132.8	0.582	0.973	0.794	0.777	0.668	0.681	0.797	0.753
*2	1147.5	0.843	0.902	0.794	0.714	0.666	0.680	0.797	0.785
*3	1148.3	0.827	0.973	0.794	0.816	0.662	0.681	0.797	0.793
4 on-ramps ($K=5$)									
#1	1056.5	0.283	0.241	0.276	0.666	0.696			0.433
#2	1059.9	0.260	0.294	0.276	0.614	0.770			0.443
#3	1063.1	0.279	0.294	0.276	0.614	0.728			0.441
*1	1152.5	0.755	0.821	0.690	0.677	0.770			0.743
*2	1152.7	0.755	0.821	0.705	0.680	0.689			0.730
*3	1162.1	0.245	0.179	0.284	0.322	0.269			0.740
5 on-ramps ($K=4$)									
#1	1053.8	0.259	0.188	0.531	0.705				0.421
#2	1055.0	0.259	0.188	0.562	0.723				0.433
#3	1057.6	0.260	0.188	0.478	0.748				0.418
*1	1150.8	0.804	0.794	0.562	0.723				0.721
*2	1152.9	0.804	0.794	0.566	0.723				0.722
*3	1155.0	0.760	0.794	0.566	0.748				0.717
7 on-ramps ($K=3$)									
#1	1054.1	0.256	0.289	0.653					0.399
#2	1056.8	0.261	0.289	0.653					0.401
#3	1058.1	0.285	0.289	0.638					0.404
*1	1144.4	0.760	0.467	0.668					0.632
*2	1145.4	0.760	0.468	0.678					0.635
*3	1147.8	0.760	0.483	0.681					0.641
10 on-ramps ($K=2$)									
#1	1052.2	0.212	0.529						0.371
#2	1053.3	0.212	0.567						0.389
#3	1056.2	0.217	0.566						0.391
*1	1143.8	0.611	0.567						0.589
*2	1146.4	0.619	0.563						0.591
*3	1146.6	0.619	0.566						0.592

\# Best dynamic ramp metering rate solution in terms of the total system delay
* Best dynamic ramp metering rate solution in terms of the average equity index

7.5.4 Impact of Ramp Metering Constraints

To test the impact of ramp metering constraints for the multiobjective minimization model (7.15)-(7.22), a scenario with the period-dependent ramp metering rate scheme with the reserve receiving capacity ratio is created by replacing the ramp metering constraint expressed by eqn. (7.6) with eqn. (7.8). Eqn. (7.8) differs from the benchmark scenario because it allocates a portion of the reserve receiving space of the mainline strength for the on-ramp vehicles. This can be interpreted as giving some priority to the on-ramp vehicles. This scenario gives a set of 63 Pareto-optimal dynamic ramp metering solution by means of the NSGA-II embedding with the MCTM. It is found that most of these solutions are also close to each other. 10 out of these 63 solutions are presented in Table 7.3.

By comparing two Pareto-optimal solutions with the best average equity index shown in Table 7.1 and Table 7.3, it can be seen that the solution gives a total travel delay of 1542.1 veh-hrs, which is 34.3% less efficient. Vehicles at on-ramps are favored more over vehicles on expressway since a portion of reserve capacity of the expressway stretch is allocated to the ramp. The amount of allocation remains the same even if the arrival rate at the on-ramp is lower than the metering rate. In such a case, neither the on-ramp nor the mainline vehicles gain from the ramp metering scheme. This has caused excessive delay to the overall system. On the other hand, since priority has been given to the on-ramps, the on-ramp delay should be lower compared to the benchmark case. From Figure 7.5, it is observed that the average on-ramp travel delay is lesser compared to the benchmark case. The best average equity index for the scenario with eqn. (7.8) is 0.845, which is about 6.2% more equitable compared to the benchmark scenario. It is interesting to note that from Figure 7.5, the maximum average ramp delay is kept below 6 minutes, which is 2 minutes lesser than that for

benchmark scenario. This gives some insight that the ramp metering equity issue could be addressed if the maximum average ramp delay is considered during the development of the ramp metering algorithms. With a more equitable dynamic ramp metering solution, a less efficient ramp metering solution is obtained.

Table 7.3 Total system travel delay $F_0\left(\mathbf{Z}^P\right)$ and equity index $I_k\left(\mathbf{Z}^P\right)$ of 10 Pareto-optimal dynamic ramp metering rate solutions for the scenario period-dependent ramp metering rate scheme with the reserve receiving capacity split ratio

| $\mathbf{Z}^P$ no.	$F_0\left(\mathbf{Z}^P\right)$ (veh-hrs)	Equity index $I_k\left(\mathbf{Z}^P\right)$							Average equity index $\bar{I}\left(\mathbf{Z}^P\right)$
		$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$	$k=7$	
1	1503.2	0.813	0.892	0.831	0.811	0.826	0.851	0.839	0.838
2	1505.7	0.187	0.895	0.831	0.827	0.798	0.844	0.843	0.836
3	1507.1	0.813	0.895	0.831	0.827	0.825	0.854	0.839	0.841
4	1509.2	0.813	0.895	0.831	0.827	0.798	0.851	0.843	0.837
5	1514.2	0.806	0.895	0.829	0.827	0.834	0.851	0.843	0.840
6	1532.5	0.811	0.896	0.854	0.827	0.829	0.847	0.843	0.844
7	1533.8	0.813	0.896	0.855	0.827	0.834	0.851	0.843	0.845
8	1535.5	0.805	0.892	0.848	0.827	0.823	0.854	0.840	0.841
9	1540.0	0.813	0.896	0.855	0.827	0.767	0.854	0.843	0.837
10	1542.1	0.813	0.896	0.855	0.826	0.834	0.851	0.843	0.845
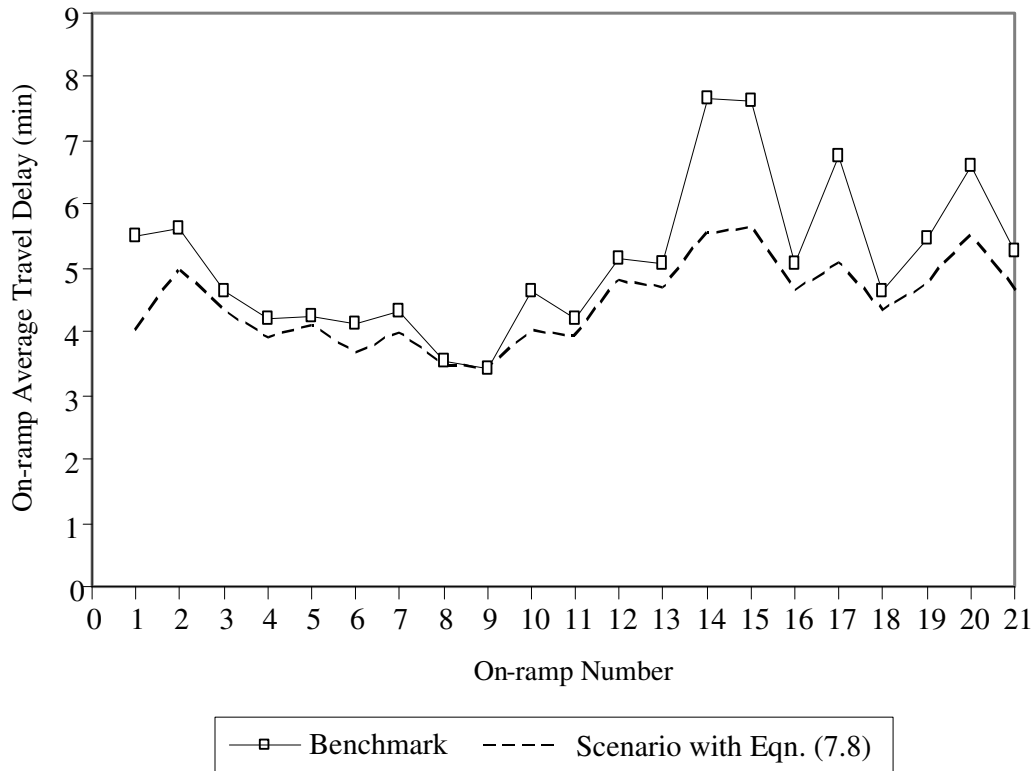
Figure 7.5 On-ramp average travel delay for the benchmark scenario and the scenario with the period-dependent ramp metering rate scheme with the reserve receiving capacity ratio

7.5.5 The Maximum Generation Effect

The results shown in the above sections are obtained by setting the maximum generation as 30. In this sub-section, it is shown numerically that a maximum generation of 30 is sufficient to produce reliable results. Figure 7.6 shows the convergence of the proposed methodology when a population size of 100 strings is used. To show the outcome in a 2-dimensional graph, all the on-ramps are grouped into one group. It can be observed from the figure that, for the results with 10 generation, the points are lesser with higher objective function values. But when the number of generation proceeds, the Pareto solution set increases and shift downwards to solutions with lower objective function values. This means that over generations, the

solution sets are enlarged with better solutions obtained. However from generation 30 onwards, the solution set converges with lesser changes in the Pareto solutions obtained. It could be seen that comparing the solutions of generation 30 and 50 as shown in Figure 7.6, the solutions overlapped with each other, with only a little different in solution observed. This shows that the Pareto solution set obtained for 30 generations is sufficient and comparable to the solutions obtained using 50 generations.
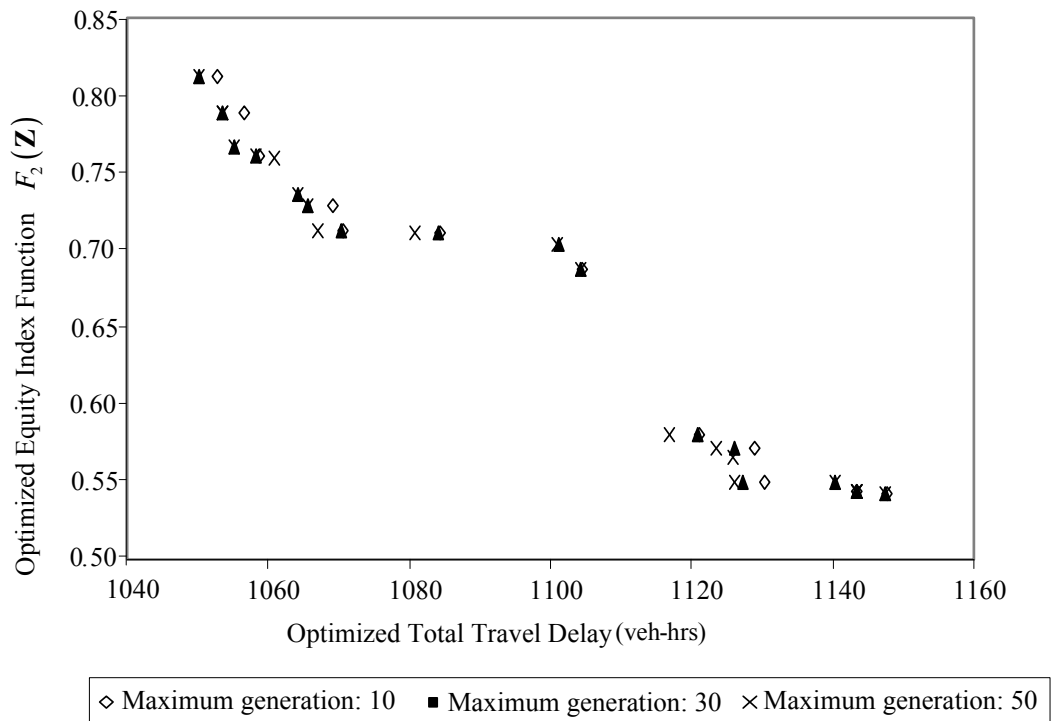

Figure 7.6 Convergence trend of the solution algorithm with different maximum generation setting

7.5.6 The Population Size Effect

Similarly, a sensitivity analysis test is carried out to test the population size effect. Figure 7.7 shows the solutions given by the proposed methodology when different population sizes of strings are used. From the figure, it can be observed that

the population size has little effect on the quality of the Pareto solutions. The solution set obtained using population size of 100 and 150 is almost the same.
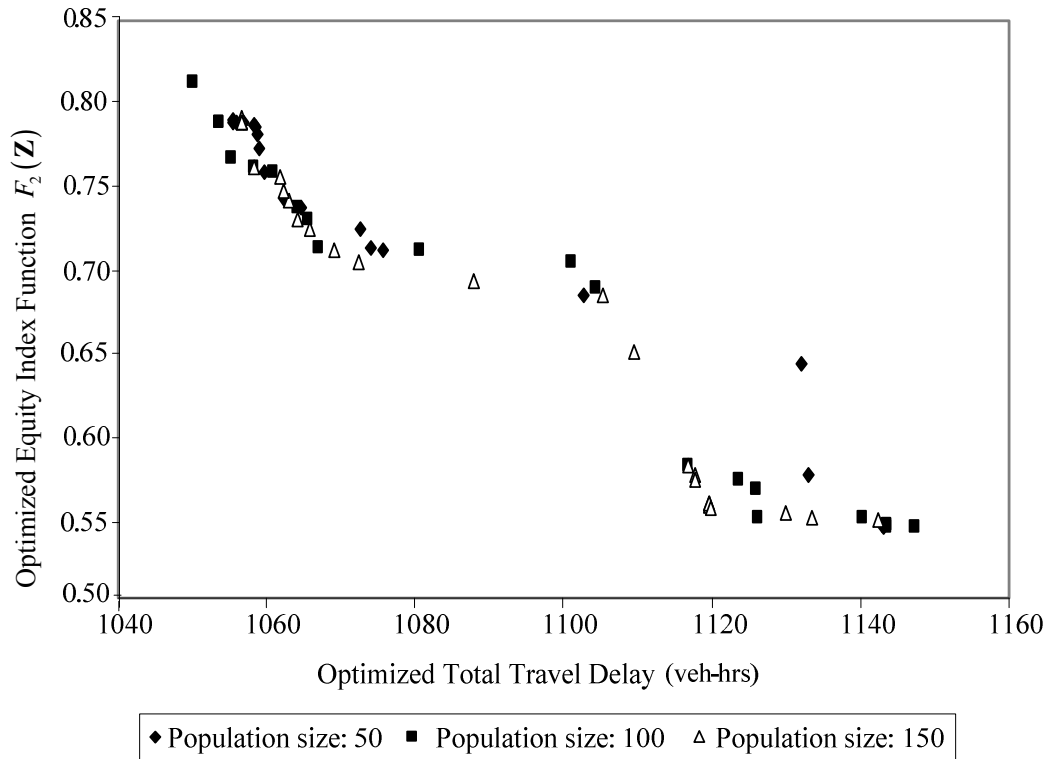

Figure 7.7 Sensitivity analysis test on the population size effect

7.5.7 Remarks

It can be seen that a perfect equity is unable to be achieved using the proposed methodology. This is because there are many factors affecting the performance of an on-ramp such as geometrical layout (i.e. the number of lanes), arrival rate, and location of the ramp from the prevailing bottleneck location, number of vicinity ramps and their distances. Based on the Pareto-optimality conditions, the proposed methodology tries its best to optimize the dynamic ramp metering rate solutions taking into account the system efficiency and the equity issue. In addition, the proposed methodology cannot guarantee the perfect equity across the groups. In this study, the spatial equity issue of

ramp metering is considered. It is one of the ways to capture ramp metering equity issue. Under such strategy, drivers are not distinguished according to their destinations. Hence, both short trips and long trips drivers are considered equal in the study. Another possible way to define ramp metering equity is to differentiate the drivers according to destinations and also the number of occupants in the vehicle.

The benchmark scenario requires about 2 hours of computational time to find the set of Pareto-optimal dynamic ramp metering solutions on a desktop personal computer with Intel Pentium IV 3.2Ghz and Ram memory of 3GB. In view of the computational time, the proposed methodology is more appropriate to use during offline decision-making stage. Another point worth mentioning is that the set of Pareto-optimal dynamic ramp metering solutions may vary with the time horizon of interest. The longer the time horizon, the more computational time will be used by the methodology proposed in this paper.

7.6 Summary

There are three contributions made in this study. First, a novel ramp metering equity index is proposed. By properly quantifying the on-ramps in an expressway-ramp network system into groups, the equity index is shown to have the ability to capture the ramp metering equity issue. A more equitable ramp metering solution could be obtained if the equity index is maximized. Second, by formulating the fair ramp metering problem as a multiobjective optimization model, both efficiency and equity issue can be balanced simultaneously. Solving the model using the meta-heuristics algorithm, NSGA, sets of Pareto solutions are obtained. The proposed methodology thus offers great flexibility for the engineers or transportation policy makers to decide the trade-off between these objectives. The proposed model is then

solved by the hybrid NSGA with the MCTM. Third, MCTM that could model the dynamic traffic flow condition and the on-ramp queue situation in a more appropriate manner is adopted in this study. The system total travel delay and the on-ramp average travel delay are derived from the model. In addition, it shows that MCTM could be employed in the ramp metering study.

The results obtained by using the example network are consistent with the previous finding where the efficiency and equity issues are found to be partially contradictory. Trade-off is needed from the efficiency in order to obtain a more equitable ramp metering solutions. This consistency has confirmed to some extent the validity of the equity index and the methodology proposed. In addition, various different scenarios are carried out in order to show the effect of some important conditions to the methodology proposed. Sensitivity analyses on the population size and the maximum generation parameters are also carried out.

CHAPTER 8 CONCLUSIONS

8.1 Outcomes and Contributions

This thesis addressed the need to investigate three dynamic traffic control and management strategies, namely contraflow operations, ATIS-based expressway-arterial corridors control operations, and ramp metering operations. Current literature studies show that there are many limitations in these strategies. For example, there is no systematic methodology to deal with the contraflow scheduling and lane configuration problems. Besides, the ramp metering equity issues are not properly addressed in the existing studies. This thesis worked on eliminating these limitations by proposing improved methodologies. For example, a novel equity index and a good methodology are proposed to address the fair ramp metering. Microscopic traffic simulation model and analytical model are used to model the dynamic traffic flow on a network. They are employed during the design, analysis and the evaluation of these strategies. It is shown that they are appropriate as tools to describe the minute-by-minute change in traffic condition.

Contraflow operations are cost-effective traffic congestion mitigation strategies. They are used during the evacuation of hurricane attack and have increasingly being adopted to mitigate city congestion. The contraflow scheduling problem and the lane configuration problems discussed in this thesis are two practical and important concerns during the offline planning of the strategy. The bilevel programming model formulation together with the hybrid GA-PARAMICS as the solution algorithm can be adopted to solve both contraflow scheduling and lane configuration problems. By adopting such formulation, both system managers' objective and drivers' route choice

decision can be captured. The upper level problem of the formulation gives a solution that can characterize for the overall system improvement measured by the total travel time; while the lower level problem measures the drivers' response to the given solution by performing their own route choice decision. In addition, the operational and safety constraints are considered implicitly in the upper level problem. Adoption of the microscopic traffic simulation model, PARAMICS, in the evaluation of the proposed strategy is appropriate because the detail of the network layout and the drivers' behavior could be captured. In addition, PARAMICS provide user-friendly environment to work in and offers great flexibility for users. An illustrative case study performed using the proposed strategies have shown that it is feasible and workable. Nevertheless, the results of the sensitivity analysis provide some insights that proper choice of the parameters, i.e. either network input parameters or the GA parameters are necessary since they will affect the outcome of the proposed strategy.

To alleviate traffic congestion on the expressway-arterial corridor without ramp meters, this thesis proposed a novel online traffic management strategy that employed ATIS means. By assuming that the real time traffic flow information, such as the density can be obtained from the detectors mounted on the expressway segments, the Expressway Mainline Control Mechanism (EMC) is triggered when the level of traffic congestion has reached the pre-determined level. The congestion information is disseminated through the ATIS means, such as variable message board or radio. It is anticipated that drivers would be rerouted to alternative routes that are less congested. Moreover, the strategy will also prevent the on-ramps and off-ramps from becoming too congested. This is guaranteed by embedding the mechanisms, namely OnC and OffC. Obviously, such strategy requires some assumptions that drivers comply with the ATIS advice. In view of the need to capture the drivers' behavior in detail, this

thesis adopted the microscopic traffic simulation model, PARAMICS. In addition, it is shown how the drivers' compliance behavior can be simulated in the environment of PARAMICS. The case study adopted in evaluation of the proposed strategy show that the strategy is effective in reducing the total system travel time of the expressway-arterial corridor. Nevertheless, the effectiveness of the control strategy is sensitive to the network and parameters tested.

Ramp metering operation is another crucial strategy needed to alleviate traffic congestion on the expressway-ramp-arterial network. Most of the current studies optimize the efficiency of the operation without taking into consideration of drivers' route choice behavior. However, it is important to consider drivers' route choice behavior because studies have shown that the application of ramp metering could cause traffic diversion to the arterial roads. In view of this, this thesis applied the Probit-based dynamic stochastic user optimal (DSUO) principle to capture the drivers' route choice decision. The DSUO is embedded as a constraint in the optimization model, which aims to minimize the total travel time. The numerical example network shows the applicability of the proposed strategy.

In addition, this thesis also addressed the issue of fair ramp metering. A fair ramp metering operation addresses the ability to capture both the efficiency and the equity issue. A novel equity index, which takes the ratio of the minimum to the maximum average on-ramp delay, is proposed as a performance measure to quantify the ramp metering equity issue. In addition, a multiobjective optimization model is proposed to formulate the fair ramp metering problem. The results of the illustrative case study confirm that the proposed methodology is able to address the problem well. The set of Pareto solutions obtained from the optimization model shows that the efficiency and the equity are two partial contradicted objectives. Trade-off needs to be

decided between these objectives. The proposed methodology offers the flexibility for system managers to decide the trade-off that could be accepted based on some pre-determined criteria.

While the previous two strategies adopt microscopic traffic simulation model for evaluation, the analytical model, i.e. modified cell transmission model (MCTM) is used to evaluate the ramp metering operation. MCTM is adopted because it could propagate traffic flow over time and space, model the horizontal queue propagation and dissipation, capture shock wave and FIFO discipline. Compared to most existing ramp metering studies, the MCTM could model the on-ramp queuing condition more appropriately since it can capture horizontal queue. This is important in ramp metering studies since the modeling of the on-ramp queuing condition could affect the effectiveness as well as the equity of the proposed methodology. The MCTM proposed in this thesis is modified from the original CTM proposed by Daganzo (1994, 1995). A special treatment on the merging cells has been done for the MCTM, compared to the original CTM. The MCTM is also used in determining the actual route travel time in performing the DSUO. However, using MCTM to compute the actual route travel time has posed some challenges. This is due to the fact that the actual travel time maybe non-differentiable under certain situation. As such, a tolerance-based DSUO is obtained.

In short, the contributions of this thesis are such as follows:

1. Improve existing three important strategies to alleviate traffic congestion problem.

2. Propose algorithms and systematic methodologies to mitigate limitations in current studies.

3. Address important issues that have been neglected in previous studies, i.e. the ramp metering equity issue.

4. Show the importance of dynamic traffic flow models adopted in evaluation and analysis of traffic control and management strategies.

8.2 Recommendations for Future Work

This thesis provides many potential future research topics .For contraflow operations, scheduling and lane configuration problem are addressed separately in this thesis. When searching for the individual optimal, the other problem needs to be fixed with assumed value. For example, when finding the optimal schedule, the lane configuration needs to be pre-determined. Hence, to mitigate this restriction, both problems can be addressed under a same framework and solved simultaneously. Besides, other meta-heuristics algorithm can be adopted to find the optimum solutions, for example tabu search or simulated annealing. In addition, parallel computing could be applied to reduce the computational time of GA-PARAMICS methodology.

For the ATIS-based expressway-arterial corridor traffic control strategy, comparison between the proposed strategies with other real-time corridor control techniques, such as route guidance can be performed. In addition, future study can be carried out to study the impact zone of the bottleneck locations. Assume that a traffic bottleneck occurred at downstream of the expressway, how far would the impact of congestion be propagated? A trial-and-error approach or a proper optimization model could be developed to determine the impact zone. Engineers may choose to disseminate the congestion information only to those drivers who are affected.

For ramp metering operation, it has been shown that optimization of the ramp metering model with consideration to drivers' route choice behavior is necessary since some drivers may be rerouted due to the ramp metering operations. There are some limitations in employing the MCTM to compute the actual route travel time for DSUO

flow computation. In the future, some other models that can address this limitation could be considered. A further investigation can be carried out to study the monotone property and the non-differentiable properties of the actual route travel time computed by MCTM. In addition, a more robust methodology to consider the route choice set can also be considered. After eliminating these limitations, some conventional method for solving non-linear programming model could be adopted to solve the fixed point formulation can be tested.

Besides, this thesis has addressed the ramp metering equity issue, which is worth studying in the future research. The proposed methodology requires on-ramps to be categorized in groups when searching for equity. This thesis provides some insight that the number of member in a group would affect the performance of the equity index. Currently, the method adopted is based on the on-ramps' locations. Other methods, such as the on-ramps demand or other factors can be considered. It is also interesting to find that whether the duplication of on-ramps grouping is allowed. Besides, research work can also focus on finding a meaningful and relevant relationship to transform the equity and efficiency issue under a same function. A comparison with the proposed multiobjective framework is essential.

Finally, another interesting subject that is left for the future work is the field validation and verification for the proposed strategies and methodologies. By doing this, the benefit of these strategies in terms of travel time savings could be validated.

REFERENCES

Abbas, M.M., and McCoy, P.T. 1999. Optimizing variable message sign locations on freeway using genetic algorithms. *Paper presented in the 78th Annual meeting of Transportation Research Board*, Washington D.C. CD-ROM.

Abdulaal, M., and LeBlanc, L.J. 1979. Continuous equilibrium network design models. *Transportation Research Part B*, 13 (1), pp. 19-32.

Ahmed, A-R, Khanal, M., and Rust, P.S. 2002. Evaluation of freeway diversion route plans using microscopic simulation and real time data: The case of Idaho's Treasure Valley Corridor. *Paper presented in the 81^{sd} Annual Meeting of Transportation Research Board*, Washington D.C. CD-ROM.

Al-Deek, H.M, Martello, M.M., May, A.D., and Sanders,W. 1988. *Potential benefits of in-vehicle information systems in a real life freeway corridor under recurring and incident-induced congestion.* Path Research Report UCB-ITS-PRR-88_2, Institute of Transportation Studies, University of California, Berkeley.

Allsop R. 1974. Some possibilities for using traffic control to influence trip distribution and route choice. In D.J. Buckley, ed, *Transportation and Traffic Theory: Proceedings of the Sixth* International *Symposium on Transportation and Traffic Theory*, Sydney, Australia, 26th-28th August, 1974. New York: Elsevier. pp. 345-374.

Balakrishna, R., Koutsopoulos, H.N., Ben-Akiva, M., Ruiz, B.M.F., and Mehta,M. 2004. A simulation-based evaluation of advanced traveler information systems. *Paper presented in the 83rd Annual Meeting of Transportation Research Board*, Washington D.C. CD-ROM.

Beegala, A., Hourdakis, J., and Michalopoulus, P.G. 2005. A methodology for performance optimization of ramp control strategies through micro-simulation. *Transportation Research Record,* 1925, pp. 87-98.

Bellemans, T., De Schutter, B., and De Moor, B. 2003. Anticipative model predictive control for ramp metering in freeway networks. *Proceedings of the American Control Conference*, 5, pp.4077-4082.

Berinde, V. 2007. *Iterative approximation of fixed points*. 2nd. Edition. New York: Springer.

Ben-Akiva, M. 1996. *Development of a deployable real time dynamic traffic assignment system, Task D Interim report: Analytical developments for DTA system*. ITS program. Cambridge, MIT ITS Program.

Ben-Akiva, M., De Palma, A. and Kanaroglou, P. 1986. Dynamic model of peak period traffic congestion with elastic arrival rates. *Transportation Science*, 20(3), pp. 164-181.

Ben-Akiva, M., De-Palma, A., and Kaysi, I. 1991. Dynamic network models and driver information systems. *Transportation Research Part A*, 25(3), pp 251-266.

Ben-Akiva, M., Cuneo, D., Hasan, M., Jha, M., and Yang, Q. 2003. Evaluation of freeway control using a microscopic simulation laboratory. *Transportation Research Part C*, 11, pp. 29-50.

Benmohamed, L., and Meerkov, S.M. 1994. Feedback control of highway congestion by a fair on-ramp metering. *Proceedings of the 33rd Conference on Decision and Control*. pp. 2437-2442.

Benson, B.G. 1996. Motorist attitudes about content of variable-message signs. *Transportation Research Record*, 1550, pp. 48-56.

Bertsimas, D., and Sim, M. 2000.The price of robutness. *Operations Research*, 52, pp. 35-53.

Birge, J.R. and Ho, J.K. 1993. Optimal flows in stochastic dynamic networks with congestion. *Operations Research*, 41(1), pp. 203-216.

Bovy, P.H.L., and Catalano, S.F. 2007. Stochastic route choice set generation: Behavioral and probabilistic foundations. *Transportmetrica*, 3 (3), pp. 173-189.

Boyce, D. 1984. Urban-transportation network-equilibrium and design models: Recent achievements and future prospects. *Environment and Planning A*, 16 (11), pp. 1445-1474.

Boyce, D., Lee, D.H., and Janson, B.N. 1998. A variational inequality model of an ideal dynamic user-optimal route choice problem. In M.G.H. Bell, ed, *Transportation networks: Recent methodological advances*, Amsterdam: Pergamon. pp. 289-302.

Boyce, D., Lee, D.H., and Ran, B. 2001. Analytical models of the dynamic traffic assignment problem. *Networks and Spatial Economics*, 1. pp. 377-390.

Brockfeld, E., Kuhne, R.D., and Wagner, P. 2005. Calibration and validation of microscopic traffic flow models. *Transportation Research Record*, 1934, pp. 179-187.

Cambridge Systematics Inc. 2000. *Twin Cities Ramp Metering Effectiveness Study: "Before" and "After" Qualitative Research with Travelers*. Technical Report. <http://www.dot.state.mn.us/rampmeterstudy/reports.html >Assessed May 10, 2007.

Cantarella, G.E. and Cascetta, E. 1995. Dynamic processes and equilibrium in transportation networks: Towards a unifying theory. *Transportation Science*, 29 (4), pp. 305-329.

Cantarella, G.E., and Vitetta, A. 2006. The multi-criteria road network design problem in an urban area. *Transportation*, 33, pp. 567-588.

Cantarella, G.E., Pavone, G., and Vitetta, A. (2006) Heuristics for urban road network design: Lane layout and signal settings. *European Journal of Operational Research*, 175, pp. 1682-1695.

Carey, M. 1986. A constraint qualification for a dynamic network loading based on travel time functions. Transportation Science, 20(1), pp. 55-58.

Carey, M. 1987. Optimal time varying flows on congested networks. *Operations Research*, 35(1), pp. 58-69.

Carey, M. 1992. Nonconvexity of the dynamic traffic assignment problem. *Transportation Research Part B*, 26 (2), pp. 127-132.

Cascetta, E. and Cantarella, G.E. 1991. A day-to-day and within-day dynamic stochastic assignment model. *Transportation Research Part A*, 25 (5), pp. 277-291.

Cascetta, E., Russo, F., and Vitetta, A. 1997. Stochastic user equilibrium assignment with explicit path enumeration: Comparison of models and algorithms. In M. Papageorgiou and A. Pouliezo, eds. *Proceedings of the 8th International Federation of Automatic Control On Transportation Systems*. Chania, Greece, pp. 1078-1084.

Ceylan, H. and Bell, M.G.H. 2004. Traffic signal timing optimization based on genetic algorithm approach, including drivers' routing. *Transportation Research Part B*, 38, pp. 329-342.

Chang, T-H., and Li, Z.Y. 2002. Optimization of mainline traffic via an adaptive co-ordinated ramp-metering control model with dynamic OD estimation. *Transportation Research Part C*, 10, pp. 99-120.

Chatterjee, K., Hounsell, N.B., Firmin, P.E, and Bonsall, P.W. 2002. Driver response to variable message sign information in London, *Transportation Research Part C*, 10 (2), pp. 149-169.

Chen, H.K., and Hsueh, C.F. 1998. Discrete-time dynamic user-optimal departure time/route choice model. *Journal of Transportation Engineering*, 124(3), pp. 245-254.

Chen, H.K., Lee, D.H., and Fu, C.T. 1994. The application of a route diversion strategy to a vehicle route guidance system using multiple driver classes. *Transportation Planning and Technology*, 18, pp. 81-105.

Chen, H-K., Chou, C-Y. and Lai, C-T. 2004. A bi-level dynamic signal timing optimization problem. *Proceedings of the IEEE International Conference on Networking, Sensing & Control*, 2, pp. 856-861.

Chen, K., and Miles, J.C., 1999. *ITS Handbook 2000: Recommendations from the World Road Association (PIARC)*. London: Aztech House.

Cheu, R.L., Jin,X., Ng, K.C., and Ng, Y.L. 1998. Calibration of FRESIM for Singapore expressway using Genetic Algorithm. *Journal of Transportation Engineering*, 124(6), pp. 526-535.

Cheu, R.L., Wang, Y. and Fwa, T.F. 2004. Genetic algorithm-simulation methodology for pavement maintenance scheduling. *Computer-aided Civil and Infrastructure Engineering*, 19(6), pp.446-455.

Chien, C-C., Zhang, Y.P, and Ioannou, P.A. 1997. Traffic density control for Automated Highway Systems. *Automatica*, 33 (7), pp. 1273-1285.

Chu, L. and Yang, X. 2003. Optimization of the ALINEA ramp metering control using genetic algorithm with micro-simulation. *Paper presented in the 82nd Transportation Research Board Annual Meeting*, Washington, D.C. CD-ROM.

Chu, L. Y., Henry, X.L., Recker, W., and Zhang, H. M. 2004a. Performance evaluation of adaptive ramp-metering algorithms using microscopic traffic simulation model. *Journal of Transportation Engineering*, 130(3), pp. 330-338.

Chu, L., Henry, X.L., and Recker, W. 2004b. Using microscopic simulation to evaluate potential Intelligent Transportation System strategies under nonrecurrent congestion. *Transportation Research Record*, 1886, pp. 76-84.

Chu, L., Henry, H.X., Oh, J., and Recker, W. 2006. A calibration procedure for microscopic traffic simulation. *Paper presented in the 85th Transportation Research Board Annual Meeting*. Washington D.C. CD-ROM.

Coello, C. A. C. 2000. An updated survey of GA-based multiobjective optimization techniques. *ACM Computing Surveys*, 32(2), pp. 109-143.

Cova, T.J., and Johnson, J.P. 2003. A network flow model for lane-based evacuation routing. *Transportation Research Part A*, 37, pp. 579-604.

Daganzo, C. F. 1981. Estimation of gap acceptance parameters within and across the population from direct roadside observation. *Transportation Research Part B*, 15, pp.1-15.

Daganzo, C.F. 1994. The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory. *Transportation Research Part B*, 28(4), pp 269-287.

Daganzo, C.F. 1995. The cell transmission model: Part II: Network traffic. *Transportation Research Part B*, 29 (2), pp. 79-93.

De Palma, A., Ben-Akiva, M., Lefevre, C. and Litinas, N. 1983. Stochastic equilibrium model of peak period of traffic congestion. *Transportation Science*, 17, pp. 430-453.

Deb, K. 2001. *Multiobjective optimization using evolutionary algorithms*. New York: John Wiley& Sons.

Deb, K., Pratap, A., Agarwal, S., and Meyarivan, T. 2002. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation,* 6(2), pp. 182-197.

Diakaki, C., Papageorgiou, M., and Aboudolas, K. 2002. A multivariable regulator approach to traffic responsive network wide signal control. *Control Engineering Practice*, 10, pp. 183-195.

Dijkstra, E.W. 1959. A note on two problems in connection with graphs. *Numerical Mathematics*, 1, pp. 269-271.

Dowling, D., Skabardonis, A., Halkias, J., Hale, G.M. and Zammit, G. 2004. Guidelines for calibration of microsimulation models: Framework and applications. *Transportation Research Record,* 1876, pp. 1-9.

Drezner, Z. and Wesolowsky, G. 1997. Selecting an optimum configuration of one-way and two-way routes. *Transportation Science*, 31(4), pp. 386-394.

Elloumi, N., Haj-Salem, H., and Papageorgiou, M. 1994. METACOR: A macroscopic modeling tool for urban corridor. *Paper presented at the Triennal Symposium on Transportation Analysis*. Capri, Italy.

Erera, A., Daganzo, C., and Lovell, D. 1999. *The access control problem on capacitated FIFO networks with unique OD path is hard*. California PATH program. Institute of Transportation Studies, UCB-ITS-PRR-99-35.

Friesz, T.L, Bernstein, D., Mehta, N.J., Tobin, R.L., and Ganjalizadeh, S. 1989. Dynamic network traffic assignment considered as a continuous time optimal control problem. *Operations Research*, 37(6), pp. 893-901.

Friesz, T.L, Bernstein, D., Smith, T.E., Tobin, R.L., and Wie, B.W. 1993. A variational inequality formulation of the dynamic network user equilibrium problem. *Operations Research*, 41(1), pp.179-191.

Fu, L., and Rilett, L.R. 1998. Shortest path problems in traffic networks with dynamic and stochastic link travel times. *Transportation Research Part B*, 32 (7), pp. 499-516.

Gallo, G., and Pallottino, S. 1984. Shortest path methods in transportation models. In: M. Florian, ed, *Transportation planning models*, Amsterdam: Elsevier, pp. 227-256.

Gardes, Y., May, A.D., Dahlgren, J., and Skabardonis, A. 2002. Freeway calibration and application of the PARAMICS model. *Paper presented in the 81st Transportation Research Board Annual Meeting*. Washington D.C. CD-ROM.

Gazis, D., Herman, R., and Rothery, R. 1961. Nonlinear follow-the-leader models of traffic flow. *Operations Research*, 9, pp. 545.

Gen, M., and Cheng, R. 1996. *Genetic algorithms and engineering design*. New York: John Wiley.

Gipps, P.G. 1986. A Model for the Structure of Lane Changing Decisions, *Transportation Research Part B*, 20, pp. 403-414.

Glover, F. 1986. Future paths for integer programming and links to an artificial intelligence. *Computers and Operations Research*, 13, pp. 533-549.

Goldberg, D. 1989. *Genetic Algorithms in search, optimization and machine learning*. Reading Mass: Addison-Wesley.

Goldberg, D.E. and Miller, B.L. 1995. *Genetics algorithms, tournament selection and the effects of noise*. IlliGAL Report No. 95006. University of Illinois. Chicago.

Gomes, G. and Horowitz, R. 2006. Optimal freeway ramp metering using the asymmetric cell transmission model. *Transportation Research Part C,* 14, pp. 244-262.

Goodwin, P.B. 1996. Empirical evidence on induced traffic-A review and synthesis. *Transportation*, 23(1), pp 35-54.

Hall, R. 1986. The fastest path through a network with random time-dependent travel time. *Transportation Science*, 20 (3), pp. 182-188.

Hasan, M., Jha, M., and Ben-Akiva, M. 2002. Evaluation of ramp control algorithms using microscopic traffic simulation. *Transportation Research Part C,* 10, pp. 229-256.

Hegyi, A., De Schutter, B., Hellendoorn, H., and Van Den Boom, T. 2002. Optimal coordination of ramp metering and variable speed control-an MPC approach. In *Proceedings of the American Control Conference*. pp. 3600-3605.

Hellinga, B. R., and Van-Aerde, M. 1995. Examining the potential of using ramp metering as a component of an ATMS. *Transportation Research Record*, 1494, pp. 169-172.

Hellinga, B.R. 1996. *Requirements for the calibration of traffic simulation models*. University of Waterloo.

Highway Capacity Manual. 2000. Transportation Research Board, National Research Council, Washington, D.C.

Hills, P.J. 1996. What is induced traffic? *Transportation*, 23(1), pp 5-16.

Holland, J.H. 1975. *Adaptation in natural and artificial systems*. Ann Arbor: University of Michigan Press.

Horn, J., Nafpliotis, N. and Goldberg, D. E. 1994. A niched Pareto generic algorithm for multiobjective optimization. In: *Proceedings of the first IEEE conference on*

evolutionary computation, IEEE world congress on computational intelligence, pp. 82-87.

Homepage of Land Transportation Authority of Singapore. 2006. <http://www.lts.gov.sg> Accessed July 09, 2006.

Horowitz, J.L. 1984. The stability of stochastic equilibrium in a two-link transportation network. *Transportation Research Part B*, 18, pp. 13-28.

Hou, Z., and Xu, J-X. 2003. Freeway traffic density control using iterative learning control approach. *Proceedings of Intelligent Transportation Systems*, 2, pp. 1081-1086.

Hourdakis, J., Michalopoulos, P.G. and Kottommannil, J. 2003. A practical procedure for calibrating microscopic traffic simulation models. *Transportation Research Record,* 1852, pp. 130-139.

Huynh, N., Chiu, Y-C., and Mahmassani, H.S. 2003 Finding near-optimal locations for variable message signs for real-time network traffic management. *Transportation Research Record,* 1856, pp. 34-53.

ITS America. 2008. <http://itsa.org> Assessed January 10, 2008.

Jacobson, L., Henry, K., and Mehyar, O. 1989. Real time metering algorithm for centralized control. *Transportation Research Record*, 1232, pp. 17-26.

Janson, B.N. 1991. Dynamic traffic assignment for urban road networks. *Transportation Research Part B*, 25(2), pp. 143-161.

Jayakrishnan, R. and Mahmassani, H.S. 1990. Dynamic simulation-assignment methodology to evaluate in-vehicle information strategies in urban traffic networks. *In Proceedings of Winter Simulation Conference*, pp. 763-769.

Jayakrishnan, R., Mahmassani, H.S. and Hu, T. 1994. An evaluation tool for Advanced Traffic Information and Management Systems in urban networks. *Transportation Research Part C*, 2(3). pp. 129-147.

Jayakrishnan, R., Oh, J., and Sahraoui, A. 2001. Calibration and path dynamics issues in microscopic simulation for Advanced Traffic Management and Information Systems. *Transportation Research Record,* 1771, pp.9-17.

Jones, D. F., Mirrazavi, S. K. and Tamiz, M. 2002. Multiobjective meta-heuristics: An overview of the current state-of-the art. *European Journal of Operational Research*, 137(1), pp. 1-9.

Kachroo, P. and Ozbay, K. 1998. Solution to the user equilibrium dynamic traffic routing problem using feedback linearization. *Transportation Research Part B*, 32(5), pp. 343-360.

Kachroo,P. and Ozbay, K. 2003. *Feedback ramp metering in Intelligent Transportation Systems*. New York: Kluwer Academics.

Kim, K.O. and Rilett, L.R. 2003. Simplex-based calibration of traffic micro-simulation models with Intelligent Transportation Systems data. Transportation *Research Record*, 1855, pp. 80-89.

Kim, S., Shekhar, S., and Min, M. 2008. Contraflow transportation network reconfiguration for evacuation route planning. *IEEE Transactions On Knowledge And Data Engineering*, 20(8), pp. 1115-1128.

Kirkpatrick, S., Jr. Gelatt, C. D. and Vecchi, M. P. 1983. Optimization by Simulated Annealing, *Science*, 220, pp. 671-680.

Kita, H. 1993. Effect of merging lane length on the merging behavior at expressway On-ramps. In C.F.Daganzo, ed, *Transportation and Traffic Theory*, Amsterdam: Elsevier Science Publishers, pp. 37-51.

Konak, A., Coit, D.W. and Smith, A.E. 2006. *Multi-objective optimization using genetic algorithms: A tutorial. Reliability Engineering and System Safety*, 91, pp. 992-1007.

Kotsialos, A., and Papageorgiou, M. 2004. Nonlinear optimal control applied to coordinated ramp metering. *IEEE Transactions on Control Systems Technology*, 12(6), pp. 920-933.

Kotsialos, A., Papageorgiou, M., Mangeas, M., and Hadj-Salem, H. 2002. Coordinated and integrated control of motorway networks via nonlinear optimal control. *Transportation Research Part C,* 10(1). pp. 65-84.

Kouvelis, P. and Yu, G. 1997. Robust discrete optimization and its applications. Norwell, MA: Kluwer Academic Publishers.

Lahiti, S., Gan, A.C., and Shen, Q. 2002. Using simulation to estimate speed improvements from simple ramp metering at on-ramp junction. *Paper presented in the 81st Transportation Research Board Annual Meeting*. Washington, D.C.CD-ROM.

Lau, R. 1997. *Ramp metering by zone: The Minnesota algorithm*. Minnesota Dept. of Transportation. USA.

Lee, D.H., Yang, X., and Chandrasekar, P. 2004a. Parameter calibration for PARAMICS using Genetic Algorithm. *Paper presented in the 83rd Transportation Research Board Annual Meeting*. Washington D.C. CD-ROM.

Lee, D.-H., Wang, H., Cheu, R. L. and Teo, S. H. 2004b. Taxi dispatch system based on current demands and real-time traffic condition. *Transportation Research Record*, 1882. pp. 193-200.

Levinson, D., and Zhang, L. 2004. Evaluating the effectiveness of ramp meters: Evidence from the Twin Cities ramp meter shut-off. In D. Gillen, and D.

Levinson, eds, *Assessing the benefits and costs of ITS*, The Netherlands: Kluwer Academic.

Levinson, D., and Zhang, L. 2006. Ramp meters on trial: Evidence from the Twin Cities metering holiday. *Transportation Research Part A,* 40. pp. 810-828.

Levinson, D., Harder, K., Bloomfield, J., and Carlson, K. 2006. Waiting tolerance: Ramp delay vs freeway congestion. *Transportation Research Part F*, 9, pp. 1-13.

Lighthill, M.J., and Whitham, J.B. 1955. On Kinematic waves. I: Flow movement in long rivers; II: Four traffic flow on long crowded roads. *Proceedings of Royal Society A*, 229, pp. 281-345.

Lim, E., and Wolshon, B. 2005. Modeling and performance assessment of contraflow evacuation. *Journal of Transportation Research Record*, 1922, pp. 118-128.

Lim, Y., and Heydecker, B. 2005. Dynamic departure time and stochastic user equilibrium assignment. *Transportation Research Part B*, 39, pp. 97-118.

Lim, Y., and Kim, H. 2005. A shortest path algorithm for real road network based on path overlap. *Journal of Eastern Asia Society for Transportation Studies*, 6, pp. 1426-1438.

Little, J.D.C. 1966. The synchronization of traffic signals by mixed integer linear programming. *Operational Research*, 14, pp. 568-594.

Lo, H. 1999. A dynamic traffic assignment formulation that encapsulates the cell transmission model. In: A. Ceder, ed, *Transportation and Traffic Theory: Proceedings of the 14th International Symposium on Transportation and Traffic Theory.* New York: Pergamon. pp. 327-395.

Lo, H., and Chen, A. 2000. Traffic equilibrium problem with route-specific costs: Formulation and algorithms. *Transportation Research Part B*, 34, pp. 493-513.

Lo, H., and Szeto, W.Y. 2002a. A methodology for sustainable traveler information services. *Transportation Research Part B*, 36, pp. 113-130.

Lo, H., and Szeto, W.Y. 2002b. A cell-based dynamic traffic assignment model: Formulation and properties. *Mathematical and Computer Modeling*, 35, pp. 849-865.

Lo, H., and Szeto, W.Y. 2002c. A cell-based variational inequality formulation of the dynamic user optimal assignment problem. *Transportation Research Part B*, 36, pp. 421-443.

Lu, J., Ban, X., Qiu, Z., and Ran, B. 2005a. A robust optimization model for route guidance based on ATIS. *Transportation Research Record*, 1935, pp. 1-7.

Lu, H., Ma, W. T. and Jayakrishnan, R. 2005b. Distributed modeling framework for large-scale microscopic simulation. *Paper presented in the 84th Annual Meeting, Transportation Research Board* , Washington D.C. CD-ROM.

Ma, W., Cheu, R.L. and Lee, D-H. 2004. Scheduling of lane closures using genetic algorithms with traffic assignments and distributed simulations. *Journal of Transportation Engineering*, 130, pp. 322-329.

Ma., J.T., Zhang, H.M., and Dong, H. 2006. Calibration of departure time and route choice parameters in micro simulation with macro measurements and Genetic Algorithm. *Paper presented in the 85th Transportation Research Board Annual Meeting*. Washington D.C. CD-ROM.

Mahut, M., Florian, M., Tremblay, N., Campbell, M. Patman, D., and McDaniel, Z.K. 2004. Calibration and application of a simulation-based dynamic traffic assignment model. *Transportation Research Record*, 1876, pp. 101-111.

Matson, J.B., and Daniel, J.R. 1998. Validation of CORSIM for the Atlanta ramp metering operation. *Paper presented at the 75th Transportation Research Board Annual Meeting*. Washington, D.C. CD-ROM.

McQueen, B., Schuman, R. and Chen, K. 2002. *Advanced traveler information systems*. Boston, MA: Artech House.

Meng, Q. 2000. Bilevel transportation modeling and optimization. PhD thesis. The Hong Kong University of Science and Technology, Hong Kong.

Meng, Q., Yang, H., and Bell, M.G.H. 2001. An equivalent continuously differentiable model and a locally convergent algorithm for the continuous network design problem. *Transportation Research Part B*, 35, pp. 83-105.

Meng, Q., Lee, D-H., Yang, H. and Huang, H-J. 2004. Transportation network optimization problems with stochastic user equilibrium constraints. *Transportation Research Record*, 1882, pp. 113-119.

Merchant, D.K. and Nemhauser, G.L. 1978. A model and an algorithm for the dynamic traffic assignment problems. *Transportation Science*, 12(3), pp. 183-199.

Merritt, E. 2003. Calibration and validation of CORSIM for Swedish road traffic conditions. *Paper presented in the 82nd Transportation Research Board Annual Meeting*. CD-ROM.

Messmer, A. and Papageorgiou, M. 1994. Automatic control methods applied to freeway network traffic. *Automatica*, 30, pp. 691-702.

Messmer, A., Papageorgiou, M., and Mackenzie, N., 1998. Automatic control of variable message signs in the interurban Scottish highway network. *Transportation Research Part C*, 6, pp. 173-187.

Michandani, B.P., and Soroush, H. 1986. Routes and flows in stochastic networks. In: G., Angrealta, F., Mason, P., Serafini, eds. *Advanced school on stochastic in combinatorial optimization*, New Jersey: World Scientific, pp. 129-177.

Michandani, P., and Head, L. 2001. A real time traffic signal control system: Architecture, algorithms, and analysis. *Transportation Research Part C*, 9, pp. 415-432.

Migdalas, A. 1995. Bi-level programming in traffic planning: models, methods and challenge. *Journal of Global Optimization*, 7, pp. 381-405.

Morin, J-M. 1995. Aid-t-decision for variable message sign control in motorway networks during incident condition, In *Proceedings of 4th ASCE International Conference Applications of Advanced Technologies in Transportation Engineering*, pp. 378-382.

Munoz, L., Sun, X., Sun, D., Gomes, G., and Horowitz, R. 2004. Methodological calibration of the Cell Transmission Model. In *Proceeding of the 2004 American Control Conference*, Boston, Massachusetts, June 30-July 2, 2004. pp. 798-803.

Murata, T., and Ishibuchi, H. 1995. MOGA: multi-objective generic algorithms. In: *Proceedings of the 1995 IEEE international conference on evolutionary computation*, pp. 289-294.

Murthy, I., and Sakar, S. 1996. A relaxation-based pruning technique for a class of stochastic shortest path problems. *Transportation Science*, 30(3), pp. 220-236.

MV2. 1997. *Evaluation of efficiency of variable message sign policy (in France).* Report to Direction Regionale Authority.

Nagurney, A. 1993. Network economics: A variational inequality approach. Kluwer Academic Publishers, Norwell, Massachusetts, USA.

Newell, G.F. 1961. Nonlinear effects in the dynamics of car following. *Operations Research*, 9(2), pp. 209-229.

Ozbay, K., Yasar, I., and Kachroo, P. 2004. Comprehensive evaluation of feedback-based freeway ramp-metering strategy by using microscopic simulation: Taking ramp queues into account. *Transportation Research Record*, 1867, pp. 89-96.

Papageorgiou, M. 1983. A hierarchical control system for freeway traffic. *Transportation Research Part B*, 17(3), pp. 251-261.

Papageorgiou, M., and Kotsialos, A. 2001. Efficiency versus fairness in network-wide ramp metering. *Proceedings of the IEEE Intelligent Transportation Systems Conference*, pp. 1189-1194.

Papageorgiou, M. and Kotsialos, A. 2002. Freeway ramp metering: An overview. *IEEE Transactions on Intelligent Transportation Systems*, 3(4), pp. 271-281.

Papageorgiou, M., Blosseville, J.M., and Habib, H.S. 1990. Modeling and real-time control of traffic flow on the southern part of Boulevard Peripherique in Paris: Part II: coordinated on-ramp metering. *Transportation Research Part A,* 24, pp. 361-370.

Papageorgiou, M., Habib, H.S., and Blosseville, J.M. 1991. ALINEA: A local feedback control law for on-ramp metering. *Transportation Research Record* ,1320, pp. 58-64.

Papageorgiou, M., Gower, P., Messmer, A., and Morin, J.M. 1994. Control strategies for variable message signs. In R. Srinivasan, F.H. Landau, C.M. Hein, P.P. Jovanis, eds, *Towards an Intelligent Transportation System: Proceedings of 1st World Congress Applications Transportation Telematics and IVHS*. London: Artech House. pp. 1229-1236.

Papageorgiou, M., Salem, H., and Middelham, F. 1997. ALINEA local ramp metering: Summary of field results. *Transportation Research Record*, 1603. pp. 90-98.

Park, E.S., Lim, J.H., Suh, I.H., and Bien, Z. 1984. Hierarchical optimal control of urban traffic networks. *International Journal of Control*, 40, pp. 813-829.

Payne, H.J. 1971. Models of freeway traffic and control. *Simulation Council Proceedings*, 1. pp. 51-61.

Peeta, S., and Mahmassani, H.S. 1995a. System optimal and user equilibrium time-dependent traffic assignment in congested networks. *Annals of Operations Research*, 60, pp. 81-113.

Peeta, S., and Mahmassani, H.S. 1995b. Multiple user classes real-time traffic assignment for on-line operations: A rolling horizon solution framework. *Transportation Research Part C*, 3(2), pp. 83-98.

Peeta, S., Poonuru, K., and Sinha, K., 2000a. Evaluation of mobility impacts of advanced information systems. *Journal of Transportation Engineering*, 126 (3), pp. 212-220.

Peeta, S., Ramos, J.L., and Pasupathy, R. 2000b. Content of variable message signs and on-line driver behavior. *Transportation Research Record*, 1725, pp. 102-108.

Qi, H. 2002. *Expressway incident detection using probe vehicles*. Master thesis. National University of Singapore, Singapore.

Quadstone Ltd. 2005. PARAMICS Modeller Reference Manual.

Ran, B. and Shimazaki, T. 1989. A general model and algorithm traffic assignment problems. *Proceedings of the Fifth World Conference on Transportation Research*, Yokohama, Japan.

Ran B., and Boyce, D. 1994. Dynamic urban transportation network models: Theory and implications for intelligent vehicle-highway system. In *Lecture notes in economics and mathematical systems 417*,Berlin: Springer-Verlag

Ran, B. and Boyce, D. 1996a. *Modeling dynamic transportation networks: An intelligent transportation system oriented approach*. 2nd Edition. Berlin: Springer.

Ran, B. and Boyce, D. 1996b. A link-based variational inequality formulation of ideal dynamic user optimal route choice problem. *Transportation Research Part C*, 4(1). pp.1-12.

Ran B., Boyce, D., and Leblanc, L. 1993. A new class of instantaneous dynamic user-optimal traffic assignment models. *Operations Research*, 41(1). pp. 192-202 .

Richards, P.I. 1956. Shockwaves on the highway. *Operations Research*, 4, pp. 42-51.

Robertson, D.I. 1969. TRANSYT: A traffic network study tool. *RRL Report, LR253,* Transport and Road Research Laboratory, Crowthorne.

Russo, F., and Vitetta, A. 2006. A topological method to choose optimal solutions after solving the multi-criteria urban road network design problem. *Transportation*, 33, pp. 347-370.

Schaffer, J. D. 1985 Multiple objective optimization with vector evaluated generic algorithm. In: J.J. Grefenstette, ed., *Proceedings of the 1st international conference on genetic algorithms and their applications*, pp. 93-100. USA: Lawrence Erlbaum Associates.

Schrank, D. and Lomax, T. 2007. *The 2007 Urban Mobility Report*. Texas Transportation Institute.

Shen, G., Dai, H., Liu, X., Wang, Z., and Sun, Y. 2003. Urban expressway traffic flow modeling and control using artificial neural networks. *Proceedings of IEEE Intelligent Transportation Systems Conference*, 1, pp. 836-841.

Sheffi, Y. 1985. *Urban transportation networks: Equilibrium analysis with mathematical programming methods.* New Jersey: Prentice Hall.

Shimizu, K., Ishizuka, Y. and Bard, F. 1997. *Nondifferentiable and two-level mathematical programming*. Boston: Kluwer Academic Publishers.

Smith, M.J. 1994. A new dynamic traffic model and the existence and calculation of dynamic user equilibria on congested capacity-constrained road networks. *Transportation Research Part B*, 27, pp. 49-63.

Srinivas, N., and Deb, K. 1995. Multiobjective function optimization using nondominated sorting genetic algorithms. *Evolutionary Computing*, 2(3), pp. 221-248.

Steuer, R.E. 1986. *Multiple criteria optimization: Theory, computation, and application*. New York: John Wiley.

Sun, T.Z., Benekohal, R.F., and Waller, S.T. 2006. Bi-level programming formulation and heuristic solution approach for dynamic traffic signal optimization. *Computer-aided Civil and Infrastructure Engineering*, 21, pp. 321-333.

Sun, X., and Horowitz, R. 2006. A set of new traffic-responsive ramp metering algorithms and microscopic simulation results. *Paper presented in the 85th Transportation Research Board Annual Meeting*. Washington, D.C.CD-ROM.

Swann, J., Routledge, I.W., Parker, J., and Tarry, S. 1995. Results of practical applications of variable message signs (VMS): A64/A1 accident reduction scheme and Forth Estuary Driver Information and Control System (FEDICS). In: *Traffic management and road safety*. Proceedings of Seminar G held at the 23rd PTRC European Transportation Forum, University of Warwick, 11-15th September, P394. London: PTRC Education and Research Services Ltd., pp. 149-167.

Szeto, W.Y. 2003. *Dynamic traffic assignment: Formulations, properties and extensions*. PhD thesis. The Hong Kong University of Science and Technology, Hong Kong.

Szeto, W.Y. and Lo, H. 2004. *On the solution existence of dynamic traffic assignment with physical queues*. Working paper.

Szeto, W.Y. and Lo, H. 2005. Non-equilibrium dynamic traffic assignment. In Mahmassani, H.S., ed, *Proceedings of the 16th International Symposium on Transportation and Traffic Theory*. New York: Elsevier. pp. 427-445.

Theodoulou, G. and Wolshon, B. 2004. Alternative methods to increase the effectiveness of freeway contraflow evacuation. *Transportation Research Record*, 1865, pp. 48-56.

Tong, C.O. and Wong, S.C. 2000. A predictive dynamic traffic assignment model in congested capacity-constrained road networks. *Transportation Research Part B*, 34, pp. 625-644.

Tsuji, H., Takahashi, R., Kawashima, H., and Yamamoto, Y. 1985. A stochastic approach for estimating the effectiveness of a route guidance system and its related parameters. *Transportation Science*, 19(4). pp. 333-351.

Tuydes, H., and Ziliaskopoulos, A. 2004. Network re-design to optimize evacuation contraflow. *Paper presented in the 83rd Transportation Research Board Annual Meeting*, Washington, D.C. CD-ROM.

Tuydes, H., and Ziliaskopoulos, A. 2006. Tabu-based heuristic approach for optimization of network evacuation contraflow. *Transportation Research Record*, 1964, pp. 157-168.

Vythoulkas, P.C. 1990. A dynamic stochastic assignment model for the analysis of general networks. *Transportation Research Part B*, 24(6), pp. 453-469.

Wan, B., Rouphail, N., and Sacks, J. 2006. Part 3: Traffic simulation, cell transmission, and measurement: Failure detection and diagnosis in microsimulation traffic models. *Transportation Research Record*, 1965, pp. 163-170.

Wang, Y., Papageorgiou, M., and Messmer, A. 2002. A predictive feedback routing control strategy for freeway network traffic. *Proceedings of the American Control Conference*, 5, pp. 3603-3611.

Wardrop, J.G. 1952. Some theoretical aspects of road traffic research. *Proceedings, Institution of Civil Engineers*, 11(1), pp 325-378.

Watling, D., and Van Vuren, T. 1993. The modeling of dynamic route guidance systems. *Transportation Research Part C*, 1(2), pp. 159-182.

Wattleworth, J.A. and Berry, D.S. 1965. Peak-period control of a freeway system-some theoretical investigations. *Highway Research Record,* 89(1). pp. 1-25.

Wie, B.W., Tobin, R.L., Friesz, T.L., and Bernstein, D. 1995. A discrete time, nested cost operator approach to the dynamic network user equilibrium problem. *Transportation Science*, 29(1). pp.79-92.

Wolshon, B., and Lambert, L. 2006. Reversible lane systems: Synthesis of practice, *Journal of Transportation Engineering*, 132 (12), pp. 933-944.

Wong, S.C., and Yang, H. 1997. Reserve capacity of a signal-controlled road network. *Transportation Research Part B*, 31 (5), pp. 397-402.

Wong, C.K., and Wong, S.C. 2002. Lane-based optimization of traffic equilibrium settings for area traffic control. *Journal of Advanced Transportation*, 36(3), pp. 349-386.

Wu, J. 2001. *Traffic diversion resulting from ramp metering*. Master Degree Thesis, The University of Wisconsin, Milwaukee.

Wu, J-H., Cen, Y., and Florian, M. 1998. The continuous dynamic network loading problem: A mathematical formulation and solution method. *Transportation Research Part B*, 32, pp. 173-187.

Xu, J-X., and Xiang, Y. 2005. A learning approach for freeway traffic control. *International Conference on Control and Automation*, 2, pp. 887-892.

Xue, D. and Dong, D. 2002. An intelligent contraflow control method for real-time optimal traffic scheduling using artificial neural network, fuzzy pattern recognition, and optimization. *IEEE Transaction on Control Systems Technology*, 8(1), pp. 183-191.

Yang, H., and Bell, M.G.H. 1998a. Models and algorithms for road network design: A review and some developments. *Transportation Review*, 18, pp. 257-278.

Yang, H. and Bell, M. G. H. 1998b. A capacity paradox in network design and how to avoid it. *Transportation Research Part A,* 32, pp. 539-545.

Yang, H., and Huang, H-J. 2005. *Mathematical and economic theory of road pricing*. New York: Elsevier Science.

Yang, H., and Yagar, S. 1994. Traffic assignment and traffic control in general freeway arterial corridor systems. *Transportation Research Part B*, 28 (6), pp. 463-486.

Yin, Y., Liu, H. and Benouar, H. 2004. A note on equity of ramp metering. *Proceedings of IEEE Intelligent Transportation Systems Conference*. pp. 497-502.

Zhang, H., S.G., Ritchie and W.W., Recker. 1996. Some general results on the optimal ramp control problem. *Transportation Research Part C*, 4(2). pp. 51-69.

Zhang, H.M., Ritchie, S.G., and Jayakrishnan, R. 2001. Coordinated traffic-responsive ramp control via nonlinear state feedback. *Transportation Research Part C,* 9. pp. 337-352.

Zhang, L., and Levinson, D. 2004. Optimal freeway ramp control without origin-destination information. *Transportation Research Part B*, 38. pp. 869-887.

Zhang, L., and Levinson, D. 2005. Balancing efficiency and equity of ramp meters. *Journal of Transportation Engineering,* 131(6). pp. 477-481.

Zhou, W.W., Livolsi, P., Miska, E., Zhang, H., Wu, J., and Yang, D. 1993. An intelligent traffic responsive contraflow lane control system. *Proceedings of the IEEE-IEE Vehicle navigation & Information systems Conference*, pp. 174-181.

Ziliaskopoulos, A. K., and Lee, S. 1996. A cell transmission based assignment-simulation model for integrated freeway/surface street systems. *Paper presented in the 75th Transportation Research Board Annual Meeting.* Washington, D.C. CD-ROM.

Ziliaskopoulos, A.K., and Waller, S.T. 2000. An internet-based Geographical Information System that integrates data, models and users for transportation applications. *Transportation Research Part C*, 8, pp. 427-444.

ACCOMPLISHMENT DURING PHD STUDY

1. Awards Earned

Two awards are earned during my PhD study in National University of Singapore.

1. President Graduate Fellowship, National University of Singapore

The award is honored to the postgraduate students who show exceptional promise in research and coursework study on the basis of competition among eligible candidates.

2. The Best Scientific Paper: Second Prize in the 14th World Congress on Intelligent Transportation Systems in Beijing, China.

The paper is selected from about 500 papers submitted to the conference. The title of the paper is: *An application of intelligent noise filtering techniques in demand forecasting for carsharing systems*.

2. List of Publications

Journal Papers

1. Meng, Q., **Khoo, H.L.,** and Cheu, R.L. 2008. Microscopic traffic simulation models based optimization approach for the contraflow lane configuration problem. *ASCE Journal of Transportation Engineering*, 134 (1), pp. 41-49.

2. Meng, Q., and **Khoo, H.L.** 2008. Optimizing contraflow scheduling problem: Model and algorithm. *Journal of Intelligent Transportation Systems*, 12(3), pp. 126-138.

3. Meng, Q., **Khoo, H.L.,** and Cheu, R.L. 2007. Urban expressway-arterial corridor on-line control system based on Advanced Traveler Information System. *Transportation Research Record*, 2000, pp. 44-50.

4. **Khoo, H.L.**, and Meng, Q. 2007. An optimal contraflow lane configuration scheme with stochastic user equilibrium constraints. *Journal of Eastern Asia Society for Transportation Studies*, 7, pp. 600-611.

5. **Khoo, H.L.**, Fung, C.H., Lee, D-H., and Meng, Q. 2007. An application of intelligent noise filtering techniques in demand forecasting for carsharing systems. *International Journal of ITS Research*, 6(1), pp. 3-10.
   (The conference version of this paper won the second prize of the Best Scientific Paper in the 14th World Congress on Intelligent Transportation Systems in Beijing, China).

6. **Khoo, H.L.**, and Meng, Q. 2008. Optimization of contraflow operations: Model and algorithms. *Journal of Institute Engineers of Singapore*, 1(4).

7. Meng, Q., and Khoo, H.L. 2007. A Pareto-Optimization approach for a fair ramp metering. Submitted to *Journal of Transportation Research Part C*. Review pending.

8. Meng, Q., and Khoo, H.L. 2007. Self-similar characteristics of vehicle arrival pattern on highways. Submitted to *ASCE Journal of Transportation Engineering*. Review pending.

Conference Papers

1. Meng, Q., **Khoo, H.L.,** Huang, Y., and Cheu, R.L. 2006. An MPEG model for the optimal contraflow operation problem with UE constraints. Proceedings of the Ninth International Conference on Applications of Advanced Technology in Transportation; Chicago, Illinois USA, August 13-16, pp. 834-839.

2. **Khoo, H.L.,** and Qiang Meng 2006. Self similarity behavior in Highway Traffic Arrival Pattern. Proceedings of the 11th International Conference of Hong Kong Society for Transportation Studies; Hong Kong, December 9-11, pp.371-380.

3. Meng, Q., and **Khoo, H.L. 2007.** Microscopic traffic simulation based for contraflow operations. The 11th World Conference on Transportation Research; University of Berkeley, California, June 24-28. Presented.

4. **Khoo, H.L.,** Fung, C.H., Xu, J-X., Lee, D-H., Meng, Q., and Lim, J.S. 2007. An application of intelligent noise filtering techniques in demand forecasting for carsharing systems. The 14th World Congress on Intelligent Transportation Systems; Beijing, China; October 9-13. Published in web version. *(This paper won the second prize of Best Scientific Paper in the conference).*

5. **Khoo, H.L.,** and Meng, Q, 2007. An integrated framework for vehicle emission impact analysis. The 5th Asia Pacific Conference Transportation and the Environment, 7-8 December, Singapore. Presented.

6. **Khoo, H.L.**, 2006. A Tabu Search-simulation methodology for contraflow operation optimization. The MUTRF 2006, Bangi, Kuala Lumpur, Malaysia.