ABSTRACT

Title of Document: INTEGRATING OF ARTERIAL SIGNAL AND FREEWAY OFF-RAMP CONTROLS FOR COMMUTING CORRIDORS

Xianfeng Yang, Doctor of Philosophy, 2015

Directed By: Gang-Len Chang, Professor, Department of Civil & Environmental Engineering

Congestion at the downstream intersections of a main arterial often causes the formation of traffic queue from the off-ramp back to the freeway mainline, and the resulting queue spillback will then substantially reduce the freeway capacity. To contend with such a critical issue, this study proposes an integrated operating system which consists of O-D demand estimation functions, pre-timed signal optimization models, and real-time signal control strategies.

The primary function of O-D estimation is to identify the traffic demand pattern at the freeway off-ramp and its connected arterial segment. Since the congestion patterns are usually caused by the green time allocation between off-ramp turning flows and local through traffic, conventional algorithms for signal design geared to best facilitate the arterial flows are likely to fall short of providing sufficient green duration to heavy off-ramp flows. Hence, to prevent queue spillover at freeway off-ramps, the system first detects the primary traveling paths of both off-ramp and local

arterial flows so as to maximize the efficiency of the signal progression system and to best utilize the capacity of the intersections within the impact region of a freeway interchange.

Based on the identified demand patterns, the second part of the system constructs two sequential models to optimize the pre-timed signal plan for the target intersections. With the objective of maximizing intersection capacity, the first model is developed to optimize the green splits and cycle length at each impacted intersection. Then, the second model optimizes the signal offset and phase sequence for each intersection to satisfy the progression needs of the identified critical path-flows.

To contend with the fluctuation of traffic flows, this study further develops a real-time control strategy which consists of three core modules, including off-ramp queue estimation, arterial adaptive control, and off-ramp priority control. If no freeway breakdown caused by off-ramp queue spillover is predicted, the arterial adaptive control module will be implemented to dynamically adjust the intersection signal timings and offsets. Otherwise the system will activate its off-ramp priority control module to offer green extensions and progression priority to the off-ramp flows.

INTEGRATING OF ARTERIAL SIGNAL AND FREEWAY OFF-RAMP CONTROLS FOR COMMUTING CORRIDORS

By

Xianfeng Yang

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2015

Advisory Committee:
Professor Gang-Len Chang, Chair
Professor Ali Haghani
Associate Professor Qingbin Cui
Associate Professor Lei Zhang
Professor Michael Fu

Dedication

To my wife, Yan Wang, my father, Xin Yang, my mother, Shunzhen Chen, and my brother, Haifeng Yang

for their love and support.

Acknowledgements

First and foremost, I would like to express my deepest sense of gratitude to my advisor, Dr. Gang-Len Chang, for his persistent guidance and inspiration throughout my Ph.D. study at the University of Maryland, College Park. His commitment to research and unconditional support to students will inspire me for my future career.

I would like to express my gratitude to the members of my doctoral examination committee, Dr. Ali Haghani, Dr. Cinzia Cirillo, Dr. Lei Zhang, Dr. Michael Fu, and Dr. Qingbin Cui for their constructive comments and valuable suggestions on my research work.

Also, very special thanks go to my colleagues in Traffic Safety and Operations Lab, Yang (Carl) Lu, Yao Cheng, Sung Yoon Park, Mark Franz, Dr. Woon Kim, Hyeonmi Kim, Liu Xu, Anna Petrone, Dr. Chien-Lun Lan, and Dr. Xin Zhang. Furthermore, I would like to thank all dear friends of mine, Dr. Wenxin Qiao, Dr. Yang Lu, Chenfeng Xiong, Dr. Shanjiang Zhu, Dr. Yongjie Lin, Dr. Shuliang Pan, Dr. Lei Feng, and Dr. Xiang He, for their support and help during my Ph.D. study.

Finally, I would like to express my deepest appreciation to my wife, Dr. Yan Wang, my father, Xin Yang, my mother, Shunzhen Chen, and my brother, Haifeng Yang. Without their unlimited encouragement and support, this dissertation would not have been completed.

Table of Contents

Chapter 1 : Introduction 1

1.1 Research Background 1

1.2 Objectives 5

1.3 Organization 6

Chapter 2 : Literature Review 11

2.1 Introduction 11

2.2 Pre-timed Arterial Signal Optimization Models 12

2.2.1 Signal Optimization at Isolated Intersections 12

2.2.1 Signal Optimization for Local Arterials 14

2.3 Real-time Signal Control Models 22

2.3.1 Actuated Signal Control 23

2.3.2 Adaptive Signal Control 26

2.4 Integrated Control Strategies 30

2.4.1 Integrated Corridor Control 30

2.4.2 Off-ramp control strategies 33

2.5 Discussions 36

Chapter 3 : System Framework of Integrated Signal Control 39

3.1 Introduction 39

3.2 Research Background and Key Issues 39

3.3 System Framework 42

Chapter 4 : Origin-Destination Matrix Estimations 46

4.1 Introduction 46

4.2 Model-I Formulations 50

4.3 Model-II Formulations 54

4.4 Model-III Formulations 56

4.5 Estimation Algorithm 61

4.5.1 Estimation Algorithm for Model-I 61

4.5.2 Estimation Algorithm for Model-II 65

4.5.3 Estimation Algorithm for Model-III 67

4.6 Numerical Test 68

4.6.1 Experimental Design 68

4.6.2 Model Evaluation and Results 71

4.6 Closure 81

Chapter 5 : A Pre-timed Corridor Signal Control Model 83
5.1 Introduction 83
5.2 Signal Timing Optimization 83
5.3 Multi-path progression model 87
5.3.1 Critical issues in a multi-path progression model 87
5.3.2 Model Formulation 89
5.4 Numerical Test-I 101
5.5 Numerical Test-II 106
5.6 Closure 117
Chapter 6 : An Integrated Real-time Control System 119
6.1 Introduction of the System 119
6.2 Off-ramp Queue Estimation 123
6.2.1 Formulations for Model-I 129
6.2.2 Formulations for Model-II 130
6.3 Arterial Adaptive Signal Control 131
6.3.1 Intersection Signal Timings Adjustment 133
6.3.2 Adaptive Signal Progression Control 137
6.4 Dynamic Off-ramp Priority Control 141
6.4.1 Intersection Signal Timing Adjustment with Off-ramp Priority 141
6.4.2 Signal Progression Design with Off-ramp Priority 144
6.5 Numerical Tests 145
6.6 Closure 155
Chapter 7 : Conclusions 157
7.1 Research Summary and Contributions 157
7.2 Future Research 160
Bibliography 163

List of Tables

Table 2.1 Summary of Adaptive Control Systems 30
Table 4.1 Key notations for O-D Estimations 51
Table 4.2 Estimation Algorithm using extended Kalman filter..... 64
Table 4.3 Time-varying in-flow at each source node 70
Table 4.4 Estimation Accuracy of O-D Flows..... 72
Table 4.5 The ground-true and identified critical paths by proposed models 78
Table 5.1 The three-hour demand patterns for the three intersections 109
Table 5.2 The optimization results from different models (unit: seconds)..... 110
Table 5.3 Percentage difference between simulation and field volume data..... 112
Table 5.4 Adjusted VISSIM parameters..... 112
Table 5.5 Arterial performance under the control of different models..... 117
Table 6.1 Key notations in this study..... 121
Table 6.2 The two-hour demand patterns for the three intersections 146
Table 6.3 Network performance with different control systems 155

List of Figures

Figure 1.1 Installed Detectors at National Highway No. 1, Zhubei, Taiwan 2
Figure 1.2 Speed obtained by detectors at the upstream of off-ramp 2
Figure 1.3 Speed obtained by detectors at the downstream of off-ramp 3
Figure 1.4 Dissertation Organizations 7
Figure 2.1 Key notations in the MAXBAND model 16
Figure 2.2 SCATS system architecture 27
Figure 2.3 The RHODES system architecture (Mirchandani, 2004) 29
Figure 3.1 An arterial segment and volume distribution in Chupei, Taiwan 40
Figure 3.2 Critical traffic paths at the interchanged area 40
Figure 3.3 A Modeling Framework of the Proposed System 43
Figure 4.1 A typical local arterial segment 50
Figure 4.2 Numbering scheme and notations for intersection and turning flows 55
Figure 4.3 Flow diverging and conservation at two adjacent intersections *l* and *l-1*.. 57
Figure 4.4 The formation of through queue with different arrival patterns 58
Figure 4.5 The geometric layout of the study site 69
Figure 4.6 Arterial topology of the study site 69
Figure 4.7 Signal timings and initial phase sequence 70
Figure 4.8 Estimation accuracy for link flow by Model-I 74
Figure 4.9 Estimation accuracy for turning flow by Model-I 74
Figure 4.10 Estimation accuracy for OD flow by Model-I 75
Figure 4.11 Estimation accuracy for link flow by Model-II 75
Figure 4.12 Estimation accuracy for turning flow by Model-II 76
Figure 4.13 Estimation accuracy for OD flow by Model-II 76
Figure 4.14 Estimation accuracy for link flow by Model-III 77
Figure 4.15 Estimation accuracy for turning flow by Model-III 77
Figure 4.16 Estimation accuracy for OD flow by Model-III 78
Figure 4.17 time-dependent flows for O-D 9→12 79
Figure 4.18 time-dependent flows for O-D 6→12 80
Figure 4.19 time-dependent flows for O-D 9→1 80
Figure 4.20 time-dependent flows for O-D 6→4 80
Figure 5.1 the deterministic queuing process at the off-ramp 85
Figure 5.2 An illustrative example for the multi-path progression (A) 87
Figure 5.3 An illustrative example for the multi-path progression (B) 88
Figure 5.4 Three possible cases for one particular progress path 88
Figure 5.5 Key notations in the proposed models 90
Figure 5.6 Illustration of the arterial for experimental analysis 101
Figure 5.7 Signal timings and the initial phase sequences 102
Figure 5.8 The resulting green bands obtained by Model-I 103
Figure 5.9 The resulting green bands obtained by Model-II 104
Figure 5.10 The resulting green bands obtained by Model-III 105
Figure 5.11 The geometric layout of the study site 107
Figure 5.12 Five identified critical paths at the study site 108
Figure 5.13 The resulting green bands obtained by MAXBAND and M3 111

Figure 5.14 Travel time passing the bottleneck 113
Figure 5.15 The time-dependent travel time along path 1 115
Figure 5.16 The time-dependent travel time along path 2 115
Figure 5.17 The time-dependent travel time along path 3 116
Figure 5.18 The time-dependent travel time along path 4 116
Figure 5.19 The time-dependent travel time along path 5 116
Figure 6.1 The framework of the integrated control system 120
Figure 6.2 Location of dual-zone detectors on the target off-ramp 123
Figure 6.3 An illustrative example for the presence data format 123
Figure 6.4 The recording of number of passing vehicles by the short detection zone 124
Figure 6.5 The detection of queue formation and clearance by long detection zone 124
Figure 6.6 Identification errors caused by short detection zone 125
Figure 6.7 Identification errors caused by long detection zone 125
Figure 6.8 Queuing vehicles can be fully discharged during the green phase 126
Figure 6.9 Queuing vehicles cannot be fully discharged during the green phase 127
Figure 6.10 Queue spillover at downstream link which affect the discharging process 127
Figure 6.11 Flowchart of the entire queue estimation process 128
Figure 6.12 Time slots within the target signal cycle 129
Figure 6.13 Time slots within the target signal cycle 130
Figure 6.14 Illustration of the detection system at each intersection 132
Figure 6.15 Green band of an outbound path between two intersections 138
Figure 6.16 Green band of an inbound path between two intersections 138
Figure 6.17 Comparison of estimated and actual queue length at the off-ramp 148
Figure 6.18 The estimation errors of the off-ramp queue estimation model 149
Figure 6.19 Green extension time granted to the off-ramp flows 150
Figure 6.20 The time-dependent travel time along the freeway marline 151
Figure 6.21 The time-dependent travel time along path 1 152
Figure 6.22 The time-dependent travel time along path 2 153
Figure 6.23 The time-dependent travel time along path 3 153
Figure 6.24 The time-dependent travel time along path 4 153
Figure 6.25 The time-dependent travel time along path 5 154

Chapter 1 : Introduction

1.1 Research Background

Over the past several decades, traffic congestion in primary commuting corridors has emerged as a serious social problem that affects the service quality of roadway infrastructure and increase the environmental pollution due to emissions. Hence, how to improve the operational efficiency of traffic control systems and best utilize the mobility of existing roadway networks has long been recognized as a vital issue by the traffic community.

Due to the interdependent nature between a freeway segment and its neighboring arterial in urban corridors, congestion formation at either the freeway or the local arterial often result in a long queue at the ramps, and consequently propagate to block the interchanged area. To contend with such issues, a large body of studies related to concurrent control of freeway and local arterial has been reported in the literature. Most of such studies, however, focused on the on-ramp metering controls and their coordinated operations (Papageorgiou et al., 1991; Papamichail et al., 2010). The equally critical issue of off-ramp control and its coordination with local traffic signal control, in contrast, has not received adequate attentions yet.

The complex interactions between off-ramp flows and arterial congestion can best be illustrated with the field data collected at one freeway segment in Chupei, Taiwan (see Figure 1.1). During PM peak hours, its freeway exiting flows to Guangming 6^{th} Rd are extremely high, and often cause an oversaturated condition at the off-ramp. Due to the impact of downstream signals, the traffic queue on the off-

ramp often builds up quickly and spill back to the freeway mainline. To record the freeway traffic conditions over the entire peak hours, the study group has installed loop detectors at both upstream and downstream of the off-ramp to collect vehicle speeds and traffic flow rates. Also, to observe the freeway queue evolution during peak hours, the study group has further placed a video camera near the bottleneck area, as shown in Figure 1.1.

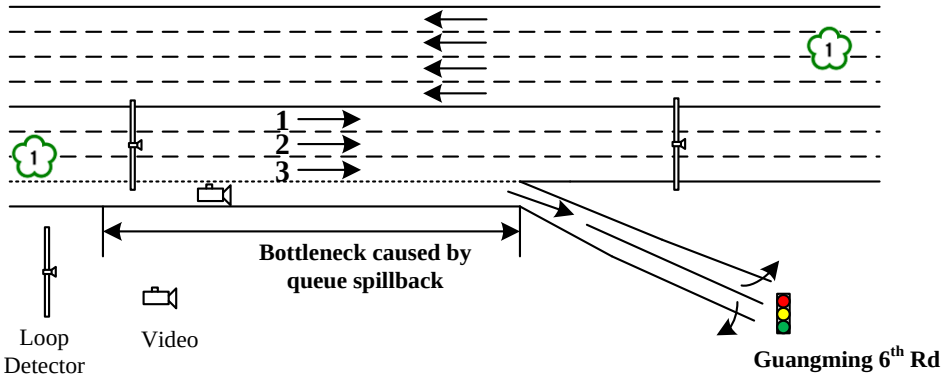


Figure 1.1 Installed Detectors at National Highway No. 1, Zhubei, Taiwan

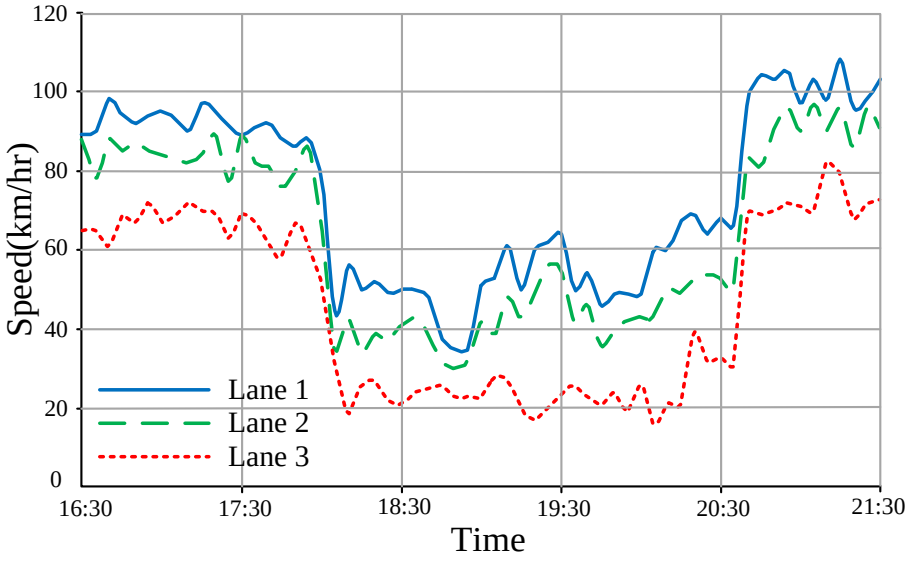


Figure 1.2 Speed obtained by detectors at the upstream of off-ramp

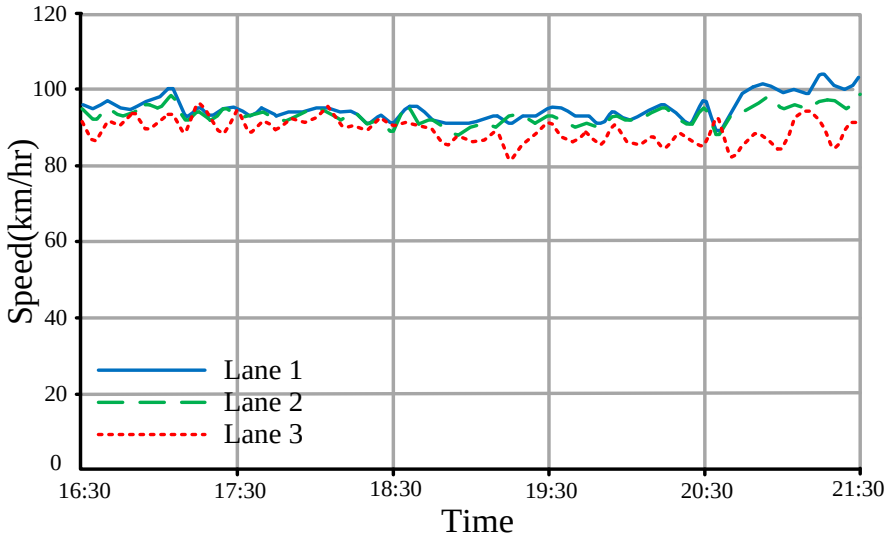


Figure 1.3 Speed obtained by detectors at the downstream of off-ramp

Figure 1.2 shows the time-dependent speed profile on each lane at the upstream of the off-ramp. During the time period of 17:00 – 20:30, one can observe significant speed drops on all three travel lanes. Noticeably, lane 3, nearest to the off-ramp exit, has dropped its speed to 20 km/h. However, after the traffic flows passed the off-ramp interchanged area, their traveling speeds on all three lanes could quickly recover to the free flow speed (i.e., 90 km/h), as evidenced by Figure 1.3.

As described by Lovell (1997), most drivers do not tend to segregate themselves by destination well in advance of an off-ramp, but rather make most of their lane-changing decisions at the last moment. Hence, when the exit queue of an off-ramp spread itself to the freeway, the operational efficiency of the through mainline flows will be significantly restricted (Daganzo et al., 1999; Muñoz and Daganzo, 2002; Jia et al, 2004). Considering a partial blockage of the right lane, Newell (1999) proposed a model to evaluate the delays on a freeway when queues from an exit ramp spill back to the freeway mainline. Cassidy et al (2002) studied the

exiting queue of an off-ramp using field data from video-tapes, and found that a bottleneck with a diminished capacity arises on a freeway segment when off-ramp queue spilled back to its mandatory exit lane.

To mitigate the freeway congestion caused by the excessive off-ramp queue, one category of studies proposed lane-assignment strategies based on travelers' destinations (Daganzo et al., 2002), or detour plans to guide traffic flows to less congested ramps (van den Berg et al., 2006; Günther et al, 2012). In addition, Hagen et al. (2006) studied the problem of queues at freeway off-ramps and developed a tool box which can offer a set of potential strategies to reduce the queues. Aside from these strategies, a more efficient way to mitigate such off-ramp queue spillover is to control the traffic signal at its connecting local arterial. In review of the literature, it is noticeable that a large body of studies has been done on optimizing the corridor control for freeway off-ramp and local arterials (Messer, 1998; Tian et al., 2002; Li et al., 2009; Lim et al., 2011; Pei and Zhou, 2013; Yang et al., 2014).

However, despite the significant research advances reported in literature, several critical issues remain unsolved. For example, a shorter length of off-ramp may require a relatively longer green time or shorter signal cycle length to prevent it from overflow during the peak period. Hence, one shall fully account for such geometrical constraints when designing the signal timings. Moreover, for the commonly-observed arterial where congestion patterns are caused by the weaving of off-ramp flows and local traffic, the existing methods, geared to control isolated off-ramp intersections or minimize the overall network delay, are likely to fall short of providing sufficient offsets for the off-ramp flows. This is due to that those off-ramp flows often need to

first take turning exercises and then proceed through movements along the arterial. Failing to account for the turning needs of heavy off-ramp flows may result in queue spillback at the both freeway off-ramp and arterial turning bays, and consequently impede the progression of local traffic flows.

1.2 Objectives

In summary, prior to the implementation of an effective signal control system at the off-ramp interchanged area, many key theoretical and operational issues await further explorations. Some of those in high priority include:

- How to design an integrated control system which can maximize the operational efficiency of both the freeway mainline segment and its neighboring arterial;
- How to analyze the demand pattern at the interchanged area so as to identify the most critical traffic paths;
- How to facilitate the progression of primary traffic flows moving on those critical paths;
- How to develop an optimal pre-timed signal control system that can best coordinate the off-ramp and local arterial traffic flows;
- How to deal with the uncertainty of off-ramp arriving flows in practice and reliably estimate their dynamic queue evolution;
- How to develop a proper real-time signal control system in responds to the traffic fluctuations on both freeway mainline and local arterial.

To address all above issues, this dissertation has developed an integrated corridor control system which consists of three major parts: O-D matrix estimation, pre-timed signal optimization, and real-time signal adaptive control. More specifically, the primary objectives of this proposed system are to:

- propose a reliable O-D estimation model to identify the destinations and travel paths of all primary traffic flows at the interchanged area;
- formulate a pre-timed signal optimization model that can offer sufficient green time to the off-ramp flows and concurrently facilitate all primary flows moving on the critical paths to pass through the congested arterial segment;
- develop a real-time control system with multi-level control strategies to estimate the queue evolution on the freeway off-ramp, contend with the uncertainty of vehicle arrival patterns on local arterial, and provide signal priority to the off-ramp flows when detecting potential off-ramp queue spillover.

1.3 Organization

Based on the proposed research objectives, this dissertation has organized the primary research tasks into seven chapters. The core of those tasks and their interrelations are illustrated in Figure 1.4.

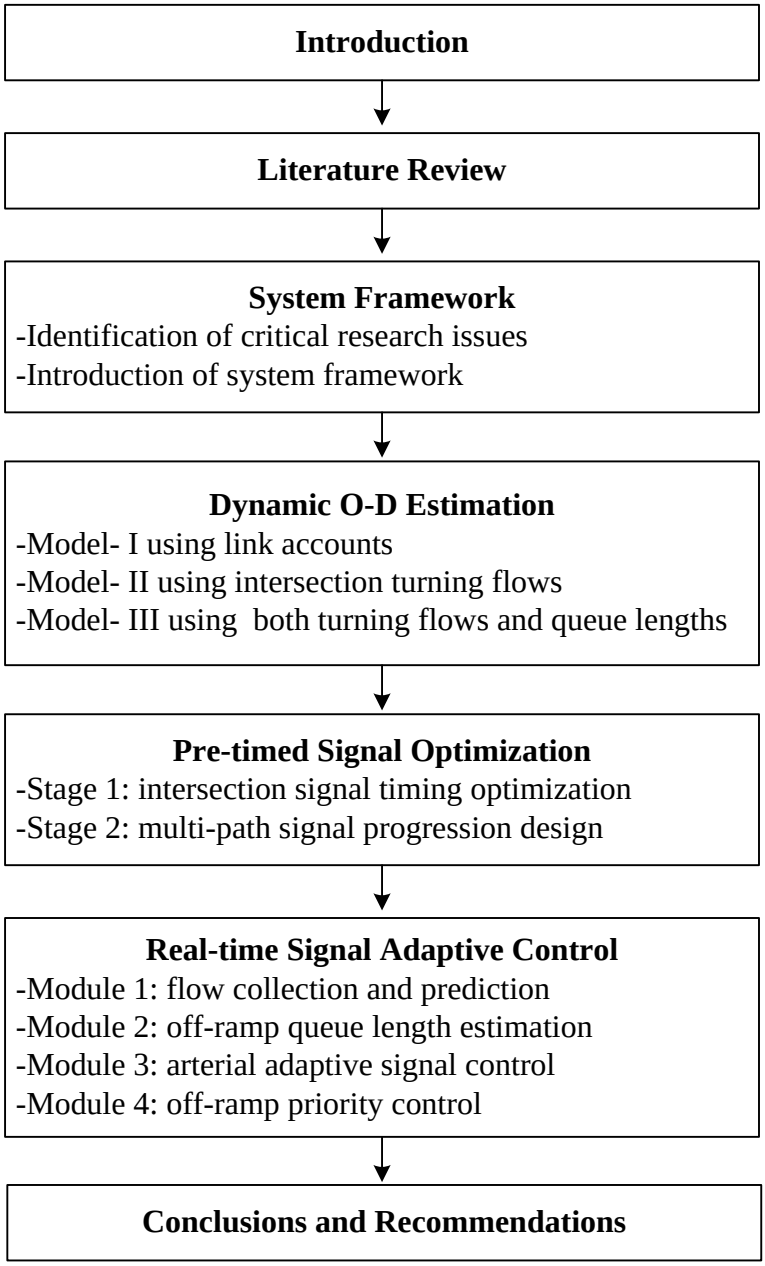


Figure 1.4 Dissertation Organizations

As shown in Figure 1.4, the remaining chapters of this dissertation are organized as follows:

- **Chapter 2** presents a comprehensive literature review of existing studies on both pre-timed and real-time signal control models for urban networks,

including intersection-level signal optimization strategies, arterial-level signal coordination plans, and integrated control methods. The review will identify the advantages of those studies and explore additional areas for further enhancements.

- **Chapter 3** illustrates the framework of the proposed integrated control system and the relations between its principal control components. Such system includes three primary stages: O-D estimation, pre-timed signal design, and real-time signal control. Also, this chapter specifies the required system inputs, the key functional features, and primary outputs at each stage.

- **Chapter 4** presents three models with different input needs for estimating the dynamic origin-destination flows at signalized arterials. Based on the principle of flow conservations, the first model is focused on capturing the relations between link counts and dynamic O-D flows, and the second model is designed to directly take turning flows at each intersection as its primary input. Since the number of measurements is typically less than that of the unknown time-varying O-D flows under such an underdetermined system, this study has further enhanced the second model with an additional set of measurements, the real-time queue length information, to improve the estimation accuracy.

- **Chapter 5** introduces two sequential optimization models to design the pre-signal plan at each impacted intersection. The first model aims to optimize the green splits for the target intersections and their common cycle length with a

properly specified queue length constraint on the off-ramp. The second model presents a multi-path progression strategy to offer green bands to all identified path-flows contributing to the high volume in the freeway interchange and its surrounding arterial segments. Depending on the embedded assumptions and the application needs, this dissertation has proposed three sets of formulations for the need of multi-path signal progression design. Using a control objective of maximizing total green bandwidth, the first set is a direct extension of MAXBAND under a predetermined phasing plan, but using the path-flow data to yield the progression band for each identified path flow. The second set further takes the phase sequence at each intersection as a decision variable, and concurrently optimizes their sequences with offsets for all intersections in the target arterial. Since multiple path-flows may compete for the signal progression and providing a near-zero green band is practically non-productive, the third model is proposed to maximize the total system efficiency by only offering the progression to the most critical path-flows.

- **Chapter 6** develops a real-time system with adaptive signal control to deal with the traffic flow fluctuations in practice. With the flow detection and prediction functions, the real-time system includes three primary modules such as off-ramp queue estimation, arterial adaptive signal control, and freeway off-ramp priority control. The off-ramp queue length estimation module is used to predict whether or not a potential queue spillover will occur in the following signal cycles. Based on the detected flow data, the arterial adaptive control module functions to dynamically adjust the intersection

signal timings and offsets so as to reduce the resulting intersection delay and provide signal progression to those heavy path-flows. If potential queue spillover is predicted by the system, its off-ramp priority control module would then be activated to offer green extension and progression priority to the off-ramp flows.

- **Chapter 7** summarizes the key contributions of this dissertation and indicates the future research directions, including: development of an optimal traffic control model to concurrently account for the delay of traffic flows on the freeway and local arterial; integration of both on-ramp and off-ramp control strategies, such as Variable Speed Limit (VSL) and Ramp-Metering (RM), for a large-scale corridor traffic; enhancement of the current real-time signal control system with advanced information/communication technologies.

Chapter 2 : Literature Review

2.1 Introduction

This chapter presents the review of related studies over the past decades on the subject of traffic signal optimization and control for urban networks. Also, the remaining critical issues to be investigated are summarized.

To facilitate the presentation, this chapter classifies all key reviewed studies into the following three categories:

- **Pre-timed Arterial Signal Optimization Models:** most of such studies focus on optimizing signal plans at isolated intersections or arterials to improve the operational efficiency of the pre-timed traffic control system, and to prevent the formation of local bottlenecks;

- **Real-time Arterial Signal Control Models:** primary efforts on this subject are to utilize the real-time traffic information from detectors, and then dynamically adjust the signal timings to contend with the traffic fluctuations in real-world applications;

- **Integrated Control Models:** researchers working on this subject intend to take the operational efficiency of both freeway mainline flows and local arterial traffic into account and offer the system-wide optimal control at the target corridor.

2.2 Pre-timed Arterial Signal Optimization Models

Although the advance in traffic detection technology has promoted the implementation of real-time signal control, the pre-timed system remains widely used in most cities, especially in developing countries, due to its low deployment and operational costs. A typical pre-timed control system assigns the right-of-way to different traffic movements according to a pre-determined signal plan, including phase sequence, green splits, cycle length, and offsets at each intersection. Among the large body of pre-timed signal optimization models, the commonly-used objective functions include minimization of traffic delay, number of stops, or pre-defined performance index, and maximization of intersection capacity or total green bandwidths.

2.2.1 Signal Optimization at Isolated Intersections

The objective of signal optimization at isolated intersections is to improve their intersection operational efficiency when the impacts to vehicle arrival patterns caused by neighboring intersections are negligible. As early as 1950s, Matson et al. (1955) proposed an optimization method to determine the signal timings with an assumed uniform vehicle arrival pattern. To account for the stochastic nature of traffic in practice, Webster (1956) developed a model to estimate the average intersection delay with Poisson arrivals. With the objective of minimizing the total delay, an empirical equation is further proposed to determine the optimal cycle length:

$$C_o = \frac{1.5\sum L_i + 5}{1.0 - \sum x_i} \tag{2.1}$$

where C_o denotes the optimal cycle length; $\sum L_i$ is the lost time per cycle; $\sum x_i$ is the ratio of critical lane volume and saturation flow rate. Note that Eq. (2.1) is only valid when the intersection is operated with under-saturated condition (i.e., the intersection critical lane volume is below the saturation flow rate).

Similarly, Miller (1963) presented a delay model for vehicle arrival patterns with any variance-to-mean ratio. Then, based on the delay model, some researchers developed the optimal signal plan by differentiating the intersection delay with respect to cycle length and green splits. Reprehensive studies following the same line include Allsop (1971, 1972, 1976), Tully (1976) and Burrow (1987).

To contend with the impacts of over-saturated conditions on the signal design, Gazis (1964) presented an over-saturated control strategy that first allocates the maximum green to the major road and the minimum green to the minor road. Then, the green time will be switched between these two to balance the residual queues. Michalopoulos and Stephanopolos (1977a, 1977b) modified Gazis' approach, and proposed the so-called "bang-bang" control to find an optimal point to switch the signal timings between different approaches. Then, Chang and Lin (2000) improved the Michalopoulos and Stephanopolos model into a discrete version. With the performance index, defined as a combination of delay and vehicle stops, the TRANSYT model (Robertson, 1969) offers the methodology to progressively optimize the signal plan with the objective of minimizing one of the pre-defined performance indexes.

Another group of studies utilized the mathematical programing methods to optimize the signal plans for isolated intersections. For example, based on the assumption that the traffic demand matrix can be multiplied with a common flow multiplier, μ, to represent the maximum amount of the increased volume that would still allow the intersection to perform reasonably well (Silcock, 1997, Wong et al., 2003, Yang et al., 2014), the optimization problem can be converted to an issue of determining the maximal multiplier μ_{max}. Hence, when the obtained μ_{max} is smaller than 1.0, the target intersection will be operated under over-saturated condition. Note that these models are formulated as a mixed-linear-integer programming problem which can be solved to achieve the global optimal. With the objective of minimizing the total delay, Lan (2004) adopted the Highway Capacity Manual (TRB, 2000) delay equation and formulated a nonlinear programming model to optimize the signal plan for near-saturated or over-saturated conditions.

2.2.1 Signal Optimization for Local Arterials

If neighboring intersections are mutually dependent, the signal optimization methods for isolated intersections may fall short of efficiency due to the formation of traffic platoons from upstream signals. In review of the literature on arterial traffic signal control, one may classify the existing studies into two distinct categories: maximizing traffic progression and minimizing total vehicle delay. The core logic of most studies in the former is to synchronize signals of common cycle length with the optimized offsets on an arterial to facilitate the movements of vehicles over consecutive intersections. Morgan and Little (1964) are the pioneers who first presented a model to maximize the total two-way progression bandwidth on an

arterial. Following the same principle, Little (1966) further proposed an advanced model to concurrently optimize the common cycle length, progression speeds, and offsets with integer programming. Then, an enhanced version, MAXBAND (Little et al., 1981), offers a rigorous method to concurrently generate the offsets between adjacent signals, optimize the prevailing speed at each link, and determine the proper left-turn phases. Its key variables are shown in Figure 2.1, and the primary formulations are presented below:

$$Max\ (b + k\bar{b}) \tag{2.2}$$

s.t.

$$(1-k)\bar{b} \geq (1-k)kb \tag{2.3}$$

$$1/C_2 \leq z \leq 1/C_1 \tag{2.4}$$

$$w_i + b \leq 1 - r_i \quad \forall i = 1,...,n \tag{2.5}$$

$$\bar{w}_i + \bar{b} \leq 1 - \bar{r}_i \quad \forall i = 1,...,n \tag{2.6}$$

$$\begin{aligned}(w_i + \bar{w}_i) - (w_{i+1} + \bar{w}_{i+1}) + (t_i + \bar{t}_i) + \delta_i L_i - \bar{\delta}_i \bar{L}_i - m_i \\ = (r_{i+1} - r_i) + (\tau_i + \bar{\tau}_i) + \delta_{i+1} L_{i+1} - \bar{\delta}_{i+1} \bar{L}_{i+1} \quad \forall i = 1,...,n-1\end{aligned} \tag{2.7}$$

$$(d_i / f_i)z \leq t_i \leq (d_i / e_i)z \quad \forall i = 1,...,n \tag{2.8}$$

$$(\bar{d}_i / \bar{f}_i)z \leq \bar{t}_i \leq (\bar{d}_i / \bar{e}_i)z \quad \forall i = 1,...,n \tag{2.9}$$

$$(d_i / h_i)z \leq (d_i / d_{i+1})t_{i+1} - t_i \leq (d_i / g_i)z \quad \forall i = 1,...,n-1 \tag{2.10}$$

$$(\bar{d}_i / \bar{h}_i)z \leq (\bar{d}_i / \bar{d}_{i+1})\bar{t}_{i+1} - \bar{t}_i \leq (\bar{d}_i / \bar{g}_i)z \quad \forall i = 1,...,n-1 \tag{2.11}$$

$$b, \bar{b}, z, w_i, \bar{w}_i, t_i, \bar{t}_i \geq 0 \quad \forall i = 1, ..., n \tag{2.12}$$

$$m_i \text{ integer}; \delta_i, \bar{\delta}_i \text{ binary integers} \quad \forall i = 1, ..., n \tag{2.13}$$

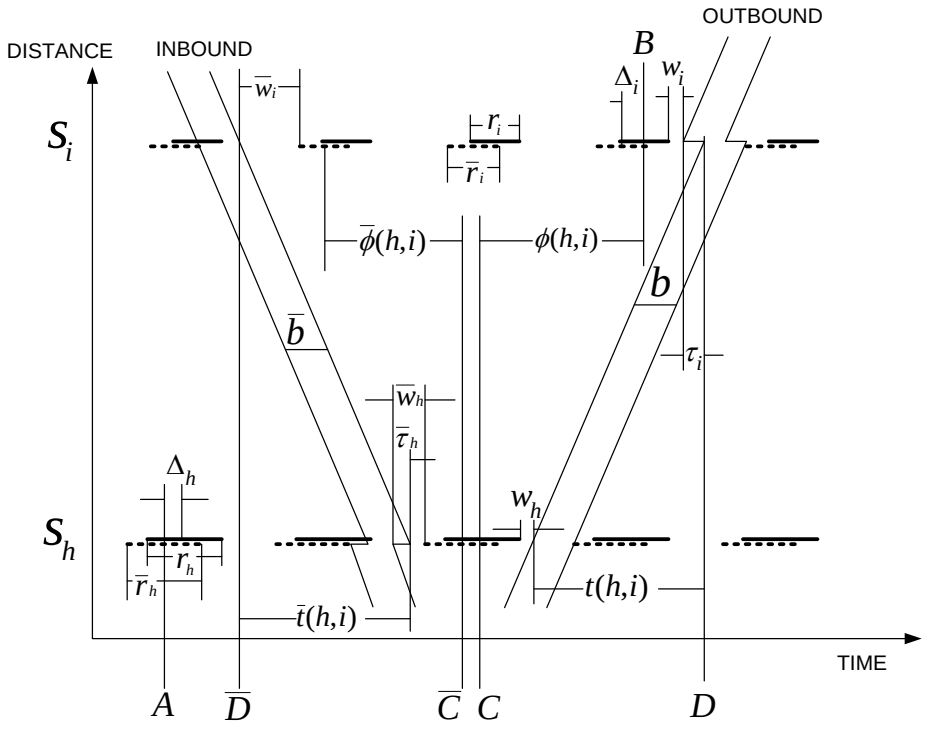


Figure 2.1 Key notations in the MAXBAND model

Parameters included in above formulations are: r_i, the common red time at signal i; $L_i(\bar{L}_i)$, the time allocated to the left-turn movements; C_1 and C_2, the boundaries of the cycle length; $e_i f_i(\bar{e}_i \bar{f}_i)$, the lower and upper limits for the outbound (inbound) speeds; $g_i h_i(\bar{g}_i \bar{h}_i)$, the lower and upper limits for the outbound (inbound) speed change.

Decision variables include: bandwidth ($b, \bar{b}$), cycle length (z), time between the start of a green phase and the boundary of its green band ($w_i, \bar{w}_i$), prevailing speed ($t_i, \bar{t}_i$), and integer variables ($\delta_i, \bar{\delta}_i, m_i$).

The objective function (2.2) for MAXBAND is to maximize the weighted sum of the two-way bandwidths. Constraint (2.3) allocates the progression preference to either the inbound or outbound direction. Constraint (2.4) limits the upper and lower bounds of the selected cycle length. The directional interference constraints in Eqs (2.5)-(2.6) can ensure the green bandwidth to be within the available green time. The loop integer constraint in Eq. (2.7) is specified to guarantee that the signals will not stop traffic flows in the green bands. The variation of travel times (a proxy of speed) is constrained by Eqs (2.8)-(2.11).

Grounded on the core logic of MAXBAND, Gartner et al. (1991) formulated a bandwidth optimization model, named MULTIBAND, to reflect the need of different bandwidths for links with different volumes. The key formulations are cited as follows:

$$Max \frac{1}{n-1}(a_i b_i + \bar{a}_i \bar{b}_i) \tag{2.14}$$

s.t.

$$(1-k_i)\bar{b}_i \geq (1-k_i)k_i b_i \quad \forall i = 1,...,n-1 \tag{2.15}$$

$$1/C_2 \leq z \leq 1/C_1 \tag{2.16}$$

$$(1/2)b_i \le w_i \le 1 - r_i - (1/2)b_i \quad \forall i = 1,...,n-1 \tag{2.17}$$

$$(1/2)b_i \le w_{i+1} \le 1 - r_{i+1} - (1/2)b_i \quad \forall i = 1,...,n-1 \tag{2.18}$$

$$(1/2)\bar{b}_i \le \bar{w}_i \le 1 - \bar{r}_i - (1/2)\bar{b}_i \quad \forall i = 1,...,n \tag{2.19}$$

$$(1/2)\bar{b}_i \le \bar{w}_{i+1} \le 1 - \bar{r}_{i+1} - (1/2)\bar{b}_i \quad \forall i = 1,...,n \tag{2.20}$$

$$\begin{aligned}(w_i + \bar{w}_i) - (w_{i+1} + \bar{w}_{i+1}) + (t_i + \bar{t}_i) + \delta_i L_i - \bar{\delta}_i \bar{L}_i - m_i \\ = (r_{i+1} - r_i) + (\tau_i + \bar{\tau}_i) + \delta_{i+1} L_{i+1} - \bar{\delta}_{i+1} \bar{L}_{i+1} \quad \forall i = 1,...,n-1\end{aligned} \tag{2.21}$$

$$(d_i / f_i)z \le t_i \le (d_i / e_i)z \quad \forall i = 1,...,n \tag{2.22}$$

$$(\bar{d}_i / \bar{f}_i)z \le \bar{t}_i \le (\bar{d}_i / \bar{e}_i)z \quad \forall i = 1,...,n \tag{2.23}$$

$$(d_i / h_i)z \le (d_i / d_{i+1})t_{i+1} - t_i \le (d_i / g_i)z \quad \forall i = 1,...,n-1 \tag{2.24}$$

$$(\bar{d}_i / \bar{h}_i)z \le (\bar{d}_i / \bar{d}_{i+1})\bar{t}_{i+1} - \bar{t}_i \le (\bar{d}_i / \bar{g}_i)z \quad \forall i = 1,...,n-1 \tag{2.25}$$

$$b_i, \bar{b}_i, z, w_i, \bar{w}_i, t_i, \bar{t}_i \ge 0 \quad \forall i = 1,...,n \tag{2.26}$$

$$m_i \text{ integer}; \delta_i, \bar{\delta}_i \text{ binary integers} \quad \forall i = 1,...,n \tag{2.27}$$

where *i* is the index of intersection.

For the same purpose but with different formulations, Chaudhary et al. (2002) also developed a progression optimization program, named PASSER. To ensure the effectiveness of the optimized progression control, Tian and Urbanik (2007) developed a partition technique to facilitate the progression on the more important direction and keep sufficient green-band within the subsets of intersections. To

account for the progression time uncertainty, Li (2014) proposed a set of formulations to assure the robustness of offsets for signal synchronization.

Besides those studies focused on the signal progression design at arterial level, some researchers also extended the models to grid networks. For example, Chang et al. (1988) implemented the MAXBAND model to a multi-arterial closed network and developed an optimization program, named MAXBAND-86. Similarly, another network version named MULTIBAND-96 was developed by Stamatiadis and Gartner (1996). However, such extensions may significantly increase the computation complexity due to the expanded size of integer variable set. To overcome this problem, Gartner and Stamatiadis (2002, 2004) proposed a two-step solution procedure that can improve the computing efficiency of existing progression models. Their proposed first step is to select a set of priority routes, each carrying a large volume, and then subsequently design signal progression for each selected priority route. The second step is to solve the progression optimization for the entire network by freezing the decision variables associated with each selected route. Following the same procedures to test each priority route in the set, one can compare its resulting objective function and then determine the optimal solution. Notably, this model was designed to generate green bands to traffic movements along those key routes rather than solely along the arterial. In design of signals for diverging diamond interchanges (DDI), Yang et al. (2014) proposed a progression model which can concurrently provide green bands to the off-ramp flows and local through traffic. However, since DDI are operated with simple two-phase signals, such a signal-progression model is only applicable for DDI's unique geometric features and limited intersections.

Existing studies in the second category focused mainly on minimizing the total delay for intersections within the control boundaries, where various versions of TRANSYT (Robertson, 1969) and TRANSYT 7-F (Wallace et al., 1988) are perhaps the most commonly adopted tool by the traffic control community. The original TRANSYT model was developed by the Transport Research Laboratory in the United Kingdom. Then, TRANSYT 7 was "Americanized" by the Federal Highway Administration (FHWA) and renamed as 7-F. With an embedded macroscopic simulation methodology, the TRANSYT 7-F program is able to model three traffic flow patterns: arrival flow pattern (IN-pattern), saturation flow pattern (GO-pattern), and departure flow pattern (OUT-pattern), where the IN-pattern is represented with the following equation:

$$IN_{it} = \sum_{j}^{n} F_{ij}(P_{ij} \cdot OUT_{jt'}) \tag{2.28}$$

where IN_{it} denotes the arrival flows on link i during time step t; F_{ij} is a smoothing function related to platoon dispersion for flow to link i from link j; P_{ij} is the proportion of leaving flow from the feeding link that arrives the subject link; $OUT_{jt'}$ is the leaving flows at link j during time step t'; t' equals t minus the travel time from link i to link j; n is the number of links that feed link i.

The GO-pattern is the flow rate at each step that would leave the link if there is enough traffic flows to saturate the green phase (saturation flow rate). Also, queue spillback due to saturated downstream links and/or adjacent turning bays is considered by reducing the GO-pattern in the affected links. The OUT-pattern is the

profile of traffic actually leaving the stop line, which could be computed with the following equations:

$$OUT_{it} = M_{i,t-1} + q_{it} - M_{it} \tag{2.29}$$

$$M_{i,t} = Max\{(M_{i,t-1} + q_{i,t} - s_{i,t}), 0\} \tag{2.30}$$

where $M_{i,t}$ is the number of vehicles in the queue at link i during time step t; $q_{i,t}$ is the number of vehicles arriving at link i during time step t; and $s_{i,t}$ is the corresponding saturation flow rate.

With the similar simulation-optimization solution method, traffic researchers have also produced various models for design of arterial traffic signals. Examples of such studies include a set of mesocopic optimizers by Yun and Park (2006) and Stevanovic et al. (2007). A set of GA-based methods to identify cycle length, green splits, offsets, and phase sequences was also proposed by Hadi and Wallace (1993) and Park et al. (1999). In addition, the Cell Transmission Model (CTM), developed by Daganzo (1994), has also been employed by researchers to optimize traffic signals. By dividing the target roadway into homogeneous segments (cells), a CTM is capable of replicating kinematic waves, queue information and dissipation (Lo, 1999; Lo et al., 2001; and Lo and Chan 2001).

In the same category of delay minimization but not using simulation-based models, a variety of signal optimization methods is available in the literature (e.g., Aboudolas et al., 2010, Li, 2012). For example, to prevent intersection blockage, Liu and Chang (2011) proposed an optimization model to remove the blocking effects at

local arterials. D'Ans and Gazis, (1976) and Papageorgiou (1995) promoted the use of store-and-forward control models to perform real-time signal optimization. Kashani and Saridis (1983) offered a set of queue-and-dispersion models for arterial signal optimization. Yin (2008) and Yang et al. (2013) developed a robust optimization model to design pre-timed signal timings.

2.3 Real-time Signal Control Models

Due to the traffic fluctuations in practice, the pre-timed signal system designed with historical demand patterns may fall short of efficiency. To contend with such problem, traffic engineers and scholars also promote the use of real-time signal control. With traffic data from deployed traffic detectors, a real-time signal control system is capable of collecting the real-time traffic information and adjusting signal timings in a dynamical manner. In practice, there exist the following two types of real-time signal control systems:

- **Actuated Signal Control:** based on real-time measured vehicle arrivals from detectors, such a system can dynamically extend the green time of the current phase to accommodate approaching vehicles, or to switch the green to other phases when no coming vehicle is detected;

- **Adaptive Signal Control:** with advanced detection devices and information technologies, most systems in this category are designed to execute a signal optimization models, considering the projected short-term traffic patterns from all intersection approaches, to dynamically adjust the signal timings.

2.3.1 Actuated Signal Control

As one of the most commonly deployed traffic control systems, actuated signal control is capable of accommodating variable phase sequences (e.g., optional protected left-turn phase), variable green time, and variable signal cycle length. Compared with the pre-timed control system, such controls can, in general, reduce passenger delay and increase intersection capacity (Boillot et al. 1992).

Depending on the detected traffic patterns and executing function, generally there are two types of such systems: semi-actuated control and fully-actuated control. In a semi-control system, traffic detectors are placed only on the side-streets. When approaching vehicles are detected, a service "call" will be sent to the system to switch the green to the side-street after reaching the pre-determined thresholds. In a fully-actuated control system, traffic detectors are installed on all intersection approaches. And there are three key parameters need to be determined for operations: 1) the minimum green time; 2) the passage time; and 3) the maximum green time.

The minimum green time is set to allow queuing vehicles to pass the intersection and also guarantee an adequate pedestrian crossing time. It should be determined by the distance between signal stop line and installed traffic detector. For instance, when the distance is below 40 feet, Kell and Fullerton (1998) recommended a minimum green time of 8 seconds. If the distance exceeds 40 feet, the minimum green time should increase at the rate of 2 seconds per 20 feet. Highway Capacity Manual (2010) also provides an equation to compute the minimum green time using crosswalk length L, pedestrian moving speed S_p, effective crosswalk width W_E, number of crossing pedestrians N_{ped}:

$$MinG = \begin{cases} 3.2 + \dfrac{L}{S_p} + 2.7\dfrac{N_{ped}}{W_E}, & if \;\; W_E > 10 \;\; ft \\ 3.2 + \dfrac{L}{S_p} + 2.7 N_{ped}, & if \;\; W_E \le 10 \;\; f \end{cases} \tag{2.31}$$

The passage time, also called extension interval, serves as both the minimum allowable gap to retain a green signal, and the amount of green time extension when a coming vehicle is detected. The passage time should be sufficiently long such that a subsequent vehicle, operating in dense traffic at a safe headway, will be able to retain a green signal, but not be over-extended to account for randomly arriving traffic. The value of passage time is typically based on detection zone length, detection zone location, number of lanes served, and vehicle speed. Highway Capacity Manual (2010) provides the following equation to compute the passage time:

$$PT = MAH - \frac{L_{ds} + L_v}{1.47 S_a} \tag{2.32}$$

with

$$L_v = L_{pc}(1 - 0.01 P_{HV}) + 0.01 L_{HV} P_{HV} - D_{sv} \tag{2.33}$$

where PT denotes the passage time; MAH is the maximum allowable headway; L_{ds} is the length of the stop-line detection zone; L_v is the detected length of vehicle; Sa is the average speed on the intersection approach; L_{pc} is the stored passenger car lane length; P_{HV} is the percent heavy vehicles; L_{HV} is the stored heavey vehicle lane length; and D_{sv} is the distance between stored vehicles.

The maximum green time is set to limit the maximum time that a phase can hold the green after receiving a service call from a conflicting approach. For both safety and efficiency needs, Lin (1985) investigated the relations between average delay and maximum green time in an actuated signal control system and produced different levels of peak hour factors to determine the optimal maximum green time. Courage et al. (1989) indicated that the value of maximum green time has little impact on intersection performance under the low volume condition. However, such impact will become significant when the intersection becomes congested. Orcutt (1993) suggested that the maximum green time should be long enough to serve 1.3 times the average queue length to accommodate arriving vehicles during the phase service time. He also suggested that at least a gap of 5 seconds should be held between the minimum and maximum green times.

Kell and Fullerton (1998) suggested a value of maximum green time falling within the range from 30 seconds to 60 seconds, which shall also be 1.25-1.5 times longer than the optimal green time in the pre-timed signal plan. Based on the average green time of actuated phases, Kim and Courage (2003) developed a computer-based model, named Enhanced Value Iteration process Actuated Signals (EVIPAS), to optimize the control parameters in the actuated control system. Due to the complexity of the EVIPAS model, a hybrid genetic algorithm was adopted for model solving. By taking advantage of the real-time queue information, Zhang and Wang (2011) proposed a stochastic model to dynamically optimize the minimum and maximum green times to adapt to the traffic fluctuations in practice.

2.3.2 Adaptive Signal Control

To improve the operational efficiency and reduce the traffic delay at local arterials, adaptive signal control is to coordinate the adjacent intersections to progress the through traffic movements. Some of such systems also include the transit signal priority functions.

Based on the pre-timed optimization strategy (i.e. TRANSYT), the Transportation Research Laboratory (Hunt et al. 1982) developed an adaptive control system, named SCOOT (Spite Cycle and Offset Optimization Technique), which stores the detected data in the form of Cyclic Flow Profiles (CFP). The CFPs will be used to estimate queue length at each link, which is a key indicator to constitute the Performance Index (PI). On the real-time operation, the integrated optimizer will continuously search the optimal signal settings, including green splits, offsets and cycle length, to minimize the PI (Day et al., 1998, Robertson et al., 1991, Hansen et al, 2000). Also, to limit the variation of signal settings between consecutive time intervals, the system only adjusts the signal timings by a small increment (e.g., 4 seconds for cycle length). To deal with the congestion, SCOOT can also implement the gating control to limit the traffic flows to the sensitive area (Bretherton et al., 2005).

The Sydney Coordinated Adaptive Control System (SCATS), developed by the Department of Main Roads NSW, is a centralized hierarchical signal control system (Sims, 1984, Cornwell, 1986; Lowrie, 1990). For operational efficiency, SCATS divides the road network into systems and subsystems. Signal systems are

separated but uncoordinated with each other due to the geographical constraints. As shown in Figure 2.2, the SCATS system architecture includes three primary levels: central management system, regional computers, and local traffic controllers.

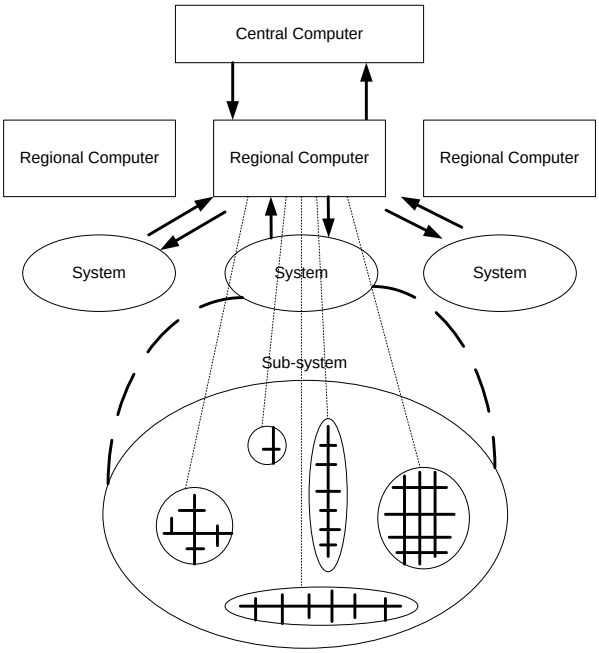


Figure 2.2 SCATS system architecture

Typically, the central computer is connected with up to 32 regional computers. And each regional computer is connected with about 250 local controllers. Based on the detector data and the intersection's degree of saturation, SCATS adjusts the cycle length, splits, and offsets to minimize the total delay and stops.

Optimized Policies for Adaptive Control (OPAC) is a distributed real-time traffic signal control system that continuously adapts signal timings to minimize a performance function, based on the total intersection delay and vehicle stops over a pre-specified horizon (Gartner, 1983; Gartner et al., 1995; Gartner et al., 2001). The

system can operate as an independent smart controller, or as part of a coordinated system. It has been developed and updated since 1979 in the following versions:

- OPAC I (1979): dynamic programming optimization with infinite projection horizon;
- OPAC II (1980): OSCO search procedure with an finite projection horizon length;
- OPAC III (1981): first use of the rolling horizon approach;
- OPAC-RT (1986): real-time implementation;
- OPAC IV (VFC-OPAC, 1995): network model for real-time traffic adaptive control using Virtual Fixed Cycle (VFC) principle;
- OPAC V (2000+): pro-active control, integration with DTA for combined control-assignment.

The most recent version of OPAC uses the rolling horizon approach to adjust the signal-timing plan and to optimize the pre-defined performance index. OPAC divides the network into sub-networks, which can be linked based on the level of congestion. OPAC can dynamically adjust the splits, offsets and cycle length, but not the phase sequence.

The Real-time Hierarchical Optimizing Distributed Effective System (RHODES) was developed by a research team at the University of Arizona (Mirchandani et al., 2000, 2001, 2004). The system uses a three-level hierarchy to characterize and manage traffic, which can explicitly predict traffic at these levels based on detector and other sensor information. As shown in Figure 2.3, the highest

level is a “dynamic network loading” model that captures the time-varying traffic conditions, which pertain to the network geometry and the route selection of travelers. Based on the traffic patterns loaded onto each particular link, RHODES allocates green time for each demand pattern and each phase. These decisions are made at the middle level of the hierarchy, referred as “Network flow control”. Given the approximate green times, the “intersection control” at the third level selects the appropriate phase change epochs, based on observed and predicted arrivals of individual vehicles at each intersection.

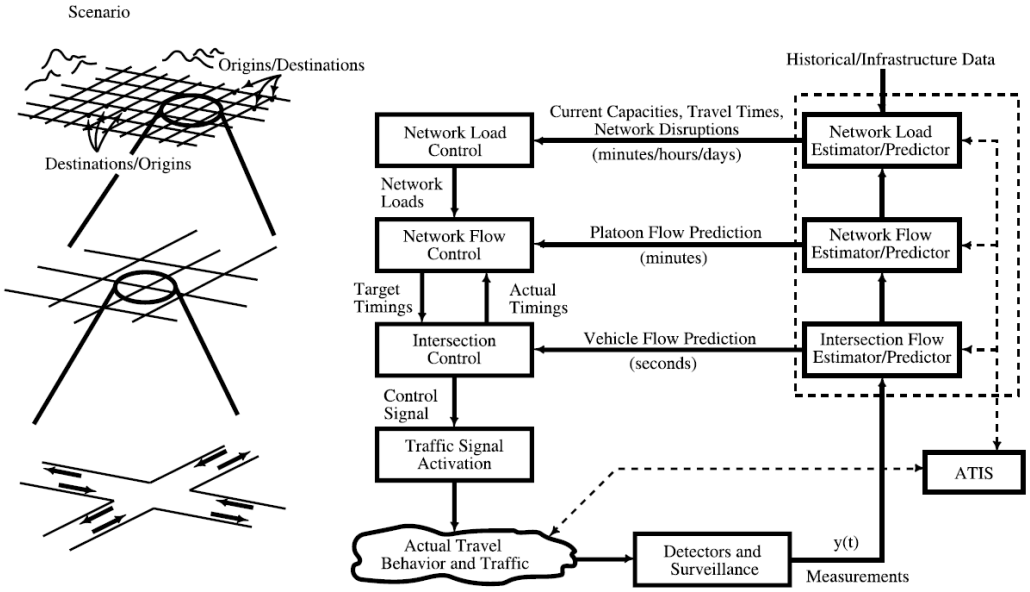


Figure 2.3 The RHODES system architecture (Mirchandani, 2004)

Other well-developed adaptive control systems include PRODYN (Henry et al., 1983) and UTOPIA (Mauro and Ditaranto, 1989). A summary of the existing adaptive signal control system is given in Table 2.1 (Cai et al., 2009):

Table 2.1 Summary of Adaptive Control Systems

System	Detection Strategy	Decision on Signal Settings	Origin Country	System Architecture
SCOOT	Upstream detectors	Adjustment of whole signal plan	UK	Centralized
SCATS	Downstream detectors	Pre-calculated signal plan	Australia	Centralized
OPAC	Upstream detectors	Change of current signal settings using rolling horizon approach	USA	Decentralized
RHODES	Fully actuated detectors	Change signal timings using three-level hierarchy	USA	Decentralized
PRODYN	Upstream detectors	Change of current signal settings	France	Decentralized
MOVA	Upstream detectors	Green extension or not	UK	Decentralized

2.4 Integrated Control Strategies

2.4.1 Integrated Corridor Control

Due to the interdependent nature of freeways and arterials at the interchanged area, some researchers attempt to address integrated controls at the corridor level. The pioneer work, done by Cremer and Schoof (1989), first proposed an integrated control framework that consists of off-ramp traffic diversion, on-ramp metering, mainline speed limit, and signal timings controls. For the freeway segment, this study

developed a traffic flow model with a set of control variables to represent the freeway speed limit, off-ramp flow diversion portion and on-ramp metering rate. On the urban streets, the platoon dispersion model for TRANSYT, taking the signal timings as control variables, was implemented to model the traffic dynamics. Then, based on these dynamic models, a mixed integer non-linear programming model was further formulated to minimize the total delay time over the entire network. For traffic controls under incident conditions, Zhang and Hobeika (1997) proposed a nonlinear programming model to optimize the traffic diversion routes, ramp metering rates, and arterial signal timings. Also, this model is capable of mitigating traffic congestions by limiting the queue lengths with constraints and penalizing long queues in the control objective.

To overcome the difficulty of solving these nonlinear programming models, Wu and Chang (1999) developed a linear programming model to optimize the corridor control under non-recurrent congestion situations. The control strategies include ramp metering, off-ramp diversion, and arterial traffic signals. On the urban streets, the traffic dynamics are modeled with three sets of linear formulations, including flow conservation, flow transitions, and flow discharging. Also, by simplifying the speed-density relation with a two-segment linear function, this study proposed a linear traffic model to capture the flow evolution on the freeway mainline and ramp links. Then, the global optimality of control variables can be achieved with such an integrated linear programming model.

In addition to those programming models, Chang et al. (1993) proposed a dynamic system-optimal model for the traffic control in commuting corridors,

including both the freeway segment and parallel local arterial. Ramp metering and intersection signal timing variables were incorporated in this optimization model, where traffic diversion and route choice were treated as predicable with on-line data. Based on the Store-and-forward approach, Papageorgiou (1995) developed a linear optimization model to design integrated control strategies for freeway corridors. However, some unrealistic assumptions embedded in this model, such as constant link travel time and controllable discharging flow rates may diminish its potential for real-world applications.

Some other studies also use an embedded macroscopic traffic flow model or cell transmission models (CTM) to optimize an integrated corridor traffic control. Van den Berg et al. (2004) proposed a model predictive control (MPC) methodology for mixed urban and freeway networks. They implemented the METANET model (Payne, 1971) to describe the traffic dynamics on the freeway segment and proposed a set of queue models to capture traffic evaluation in the urban arterial and ramp links. With the CTM concept, Li (2012) developed an integrated control model to concurrently capture the off-ramp spillback, freeway mainline spillback, and arterial link blockage. Haddad et al. (2013) introduced a cooperative traffic control model for a mixed network with two urban regions and a freeway. Perimeter controllers on the border of urban regions and ramp metering controllers at on-ramps are implemented to regulate traffic flows, where the freeway was regarded as one alternative connecting route that has one on-ramp and off-ramp within each urban region. Then, the urban and freeway flow dynamics were formulated with the theory of

Macroscopic Fundamental Diagram (MFD) and asymmetric CTM. However, control variables for signal timings in urban regions are not taken into account in this study.

Using the information from the queue detectors, Zhang et al. (2009) proposed a local synchronization control scheme for congested freeway interchange areas. They proposed the following three operational strategies:

1) On-ramp priority: when the queue spillback is detected at the upstream of an on-ramp, the system will turn off the ramp metering or reduce the maximum green time to discharge traffic at the ramp;

2) Off-ramp priority: when the queue detector placed at the upstream end of an off-ramp is triggered, the maximum green time of the phase that discharges the off-ramp traffic will be increased by the adjustment factor;

3) Intersection gating control: once the queue detector positioned on the upstream of a critical link is triggered, the flow-feeding phases will be reduced and the discharging phases will be increased to clear the queue blockage.

2.4.2 Off-ramp control strategies

Different from those integrated control models that aim to improve the operational efficiency of the entire corridor network, another set of studies focus on coordinate ramp flows with neighboring intersections. However, most of these studies focused on on-ramp metering control. A comprehensive review of such studies can be found in Papageorgiou et al. (2000). Similar studies, intending to coordinate the ramp

metering with arterial traffic signals, can be found in the studies by Pooran (1994), Tian et al. (2002) and Lu et al. (2013).

In contrast, the equally critical issue of off-ramp control has not yet received adequate attention. To mitigate the freeway congestion caused by the excessive off-ramp queue, some studies focus on regulating drivers' behaviors such as eliminating the lane changing maneuvers near the off-ramp. For example, Daganzo et al. (2002) presented a dynamic lane assignment strategy to reduce the frequency of lane-changing maneuvers at the congested off-ramp areas. With variable message signs (VMS) and associated technologies for monitoring driver behavior, the strategy was proposed to segregate drivers by destination and force them to change lanes before reaching the bottleneck area.

Based on the field observations, Rudjanakanoknad (2012) proposed two traffic control strategies to increase the off-ramp capacity in the congested area: off-ramp control and prohibiting lane change maneuvers near the off-ramp. The off-ramp control strategy was designed to increase the exiting flows by blocking its competing traffic streams once the speeds of freeway through traffic flows drop below the critical value.

To account for drivers' queue-jump behavior at the entry of an off-ramp, Di et al. (2013) proposed a cellular automata-based simulation model to evaluate different configurations of pavement markings around off-ramps. Based on the simulation results, they concluded that prohibiting lane-changing maneuvers near an off-ramp can effectively enhance the freeway capacity when the cycle length of the

downstream signal is sufficient long. However, such strategy may increase the vehicle delays for arterials with a short signal cycle length.

To avoid the freeway congestion caused by off-ramp queue spillback, another category of studies aims to detour the flows to other non-congested areas. For example, Günther et al. (2012) proposed a model to detour some vehicles on the surface streets, and offer the control priority to the off-ramp flows. Hence, their model intends to benefit the off-ramp flows at the expense of surface street users. To prevent the negative impacts to the traffic flows on surface streets, Spiliopoulou et al. (2013) developed a real-time route diversion model from the user-optimum perspective. Given a detected off-ramp queue spillback, the control module will be executed to detour some off-ramp flows to an alternative route, aiming to prevent queue spillback at off-ramps. Since drivers may ignore the detour instructions, Spiliopoulow et al. (2014) further proposed another control model that calls for temporary off-ramp closure to force the route diversion.

Aside from the aforementioned strategies, a more efficient way to mitigate such off-ramp queue spillover is to control the traffic signals at its connecting local arterial. In review of the literature, it is noticeable that considerable studies have been done on optimizing the corridor control for freeway off-ramp and local arterials. Along the same line, Messer (1998) provided a control strategy and simulation study to solve traffic congestion at a closely-spaced signalized arterial which has a short distance between its intersections and the interchange exit. Tian et al. (2002) developed an integrated control algorithm, including ramp metering and local signal timings, to improve the performance of a freeway diamond interchange and its

neighboring surface street. Li et al. (2009) presented a mixed-integer model for an integrated control between the off-ramp and arterial traffic flows, intending to minimize the queue spillback from an off-ramp to its the freeway mainline. Similarly, Lim et al. (2011) proposed a signal control model to minimize the total delay for off-ramps and their connected arterials. Pei (2013) developed a control model to optimize the green time and cycle length at a surface road, based on the off-ramp traffic conditions.

2.5 Discussions

In summary, this chapter has provided a comprehensive review of research efforts on the subject of traffic signal optimization and control in urban networks. Existing control models has been classed into three groups: pre-timed arterial signal optimization, real-time arterial signal control, and integrated control. Those studies focus on both pre-timed and real-time arterial control have demonstrated their effectiveness in reducing travel delays and improving signal operational efficiency at local networks. However, such intersection-level and arterial-level models may fall short of efficiency at the off-ramp interchanged area due to the heavy entering flows from freeway segment.

Compared with those intersection-level and arterial-level control methods, only a few integrated control models, which accounts for the operational benefit of both the freeway and local arterial, could be found in the literature. The first category of studies proposed multiple strategies such as off-ramp traffic diversion, on-ramp metering, mainline speed limit, and signal timings controls. Also, to coordinate those

control strategies and to optimize the overall network performance, these studies developed a set of linear/non-linear programming and simulation-based optimization models. However, due to the large size of formulations, some models may not be able to find the optimal solution for their system control variables. In addition, the complex flow interacting nature at some off-ramp interchanged areas is not well analyzed in those studies.

The second category of studies focused on exploration of integrated off-ramp control strategies. A part of them aims to reduce the entering flow in the interchanged areas by detouring flows to neighboring networks. However, such strategies may not be implemented in practice due to the difficulty in detouring vehicles during peak-hours. Aside from these studies, another part of researchers developed optimization models to prevent off-ramp queue spillover and minimize total delay of both freeway segment and local arterial. Despite the progress of existing literature on traffic control, some critical issues remain to be addressed:

- How to properly design signal timings at off-ramp connected intersections so as to prevent the occurrence of off-ramp queue spillover and best utilize the intersection capacity;

- How to deal with the conflicts between off-ramp flows and local traffic, and to facilitate both types of traffic flows to reach their destinations;

- How to accurately estimate the evolution of off-ramp queue length to face the uncertainty of freeway exiting flows;

- How to design real-time responsive strategies to support the signal control system when the off-ramp queue spillover has been detected.

To solve these critical issues, this dissertation aims to develop a promising integrated signal control system to mitigate traffic congestion in a corridor network due to off-ramp queue spillover.

Chapter 3 : System Framework of Integrated Signal Control

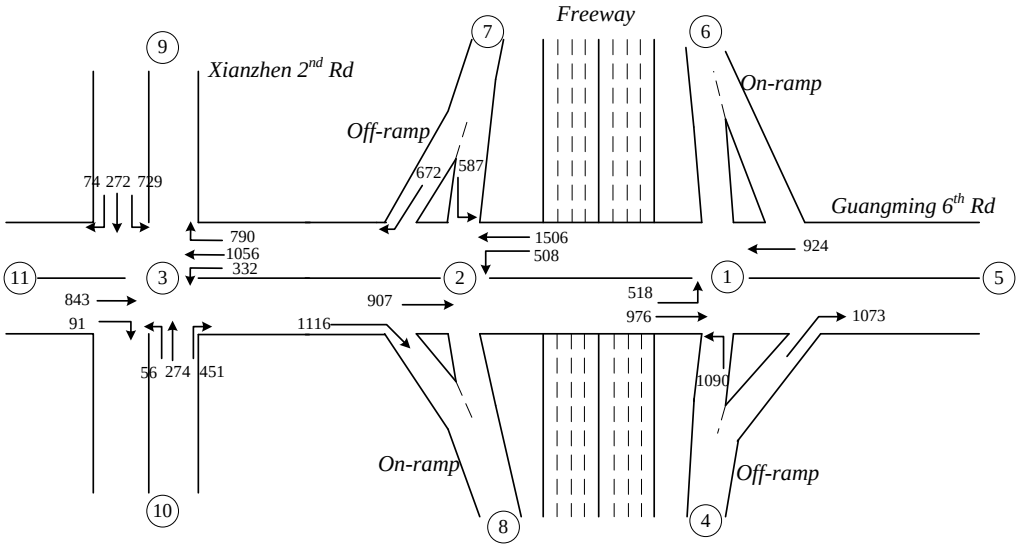
3.1 Introduction

This chapter illustrates the framework of the proposed integrated signal control system. The rest of this chapter is organized as follows: Section 3.2 introduces the research background and identifies a set of key research issues associated with the development of control systems in the off-ramp interchange area; Section 3.3 decomposes the system into three primary stages and specifies the corresponding control models at each stage.

3.2 Research Background and Key Issues

As discussed in the previous chapters, the presence of off-ramp queue spillover may bring negative impacts to the freeway mainline flows and result in operational bottlenecks. Hence, to provide an effective signal control to overcome this issue, the proposed system shall first optimize the signal timings at the neighboring intersections and provide sufficient green time to discharge off-ramp queuing vehicles. In addition, a signal progression model is essential to best coordinate the off-ramp flows with local arterial traffic, and to facilitate those traffic flows to reach their destinations. Figure 3.1 presents an example of interchange network in Chupei, Taiwan, where the arterial segment comprises three intersections to connect a congested commuting freeway and local arterial. The heavy turning volumes from the on-ramp and to the off-ramps are in conflict with through traffic.

Thus, the design of conventional two-way progression often yields overflow at off-ramps and consequently a gridlock for the entire network.



Volume Data: 18:00-19:00 (veh/h), April 24, 2013

Figure 3.1 An arterial segment and volume distribution in Chupei, Taiwan

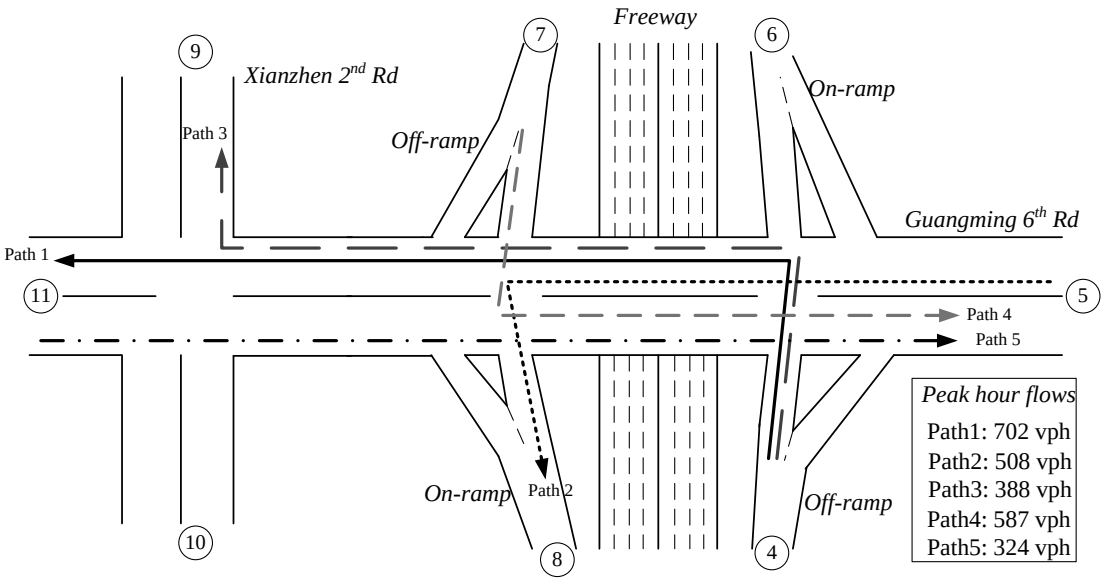


Figure 3.2 Critical traffic paths at the interchanged area

A further field survey and analysis has revealed that the traffic patterns along the arterial segment are the collective manifestation of five congested path-flows. As shown in Figure 3.2, Path-1 flows from Node 4 to Node 11 exhibit the highest volume (702 vph), and all vehicles along this path need to first manipulate a turning movement and then join the through traffic on the main arterial. Other primary path-flows, including Path-2 (Node 5 to Node 8), Path-3 (Node 4 to Node 9), and Path-4 (Node 7 to Node 5) also share the common features of having heavy turning volume to merge into the through traffic flows. Moreover, traffic flows in those outbound paths (Paths 1, 2, 3) are in conflict with those inbound paths (Paths 4, 5), inevitably causing the conventional design of two-way progression ineffective. Hence, an effective control system shall be capable of optimizing the signal plans and smooth traffic movements on such a congested arterial that serves as a connector between a commuting freeway and primary trip-destination streets.

In summary, to design a pre-timed control system at the commonly-observed networks shown in Figure 3.1 and 3.2, this dissertation has identified several research issues to be addressed:

- Estimation of the origin-destination flow patterns at the interchanged area, which yields the critical paths of flows as shown in Figure 3.2;
- Optimization of signal timings at each impacted intersection to prevent the potential off-ramp queue from spilling over and to best utilize each intersection's capacity; and

- Formulation of a multi-path progression model to coordinate those neighboring intersections and to progress the primary path-flows to reach their destinations.

However, due to the freeway traffic fluctuation and the time-varying arrival rate at the off-ramp, an integrated corridor control with pre-timed strategies may not be sufficiently responsive. Using the real-time traffic information obtained from detectors, this dissertation further proposes a real-time control module to prevent the queue spillback at freeway off-ramps. By installing detectors at the off-ramp and nearby local intersections, such a control module shall effectively address the following key issues:

- Detection of potential queue spillover at the off-ramp based on a reliable queue estimation model;
- Designing of adaptive priority strategies in response to various traffic conditions and congestion levels at the freeway and local arterials; and
- Construction of real-time priority control functions to allow the off-ramp flows to quickly pass the downstream intersections.

3.3 System Framework

In response to those research issues listed in the last subsection, Figure 3.3 depicts the framework of the proposed integrated control system, which includes a traffic detection system and three key control functions.

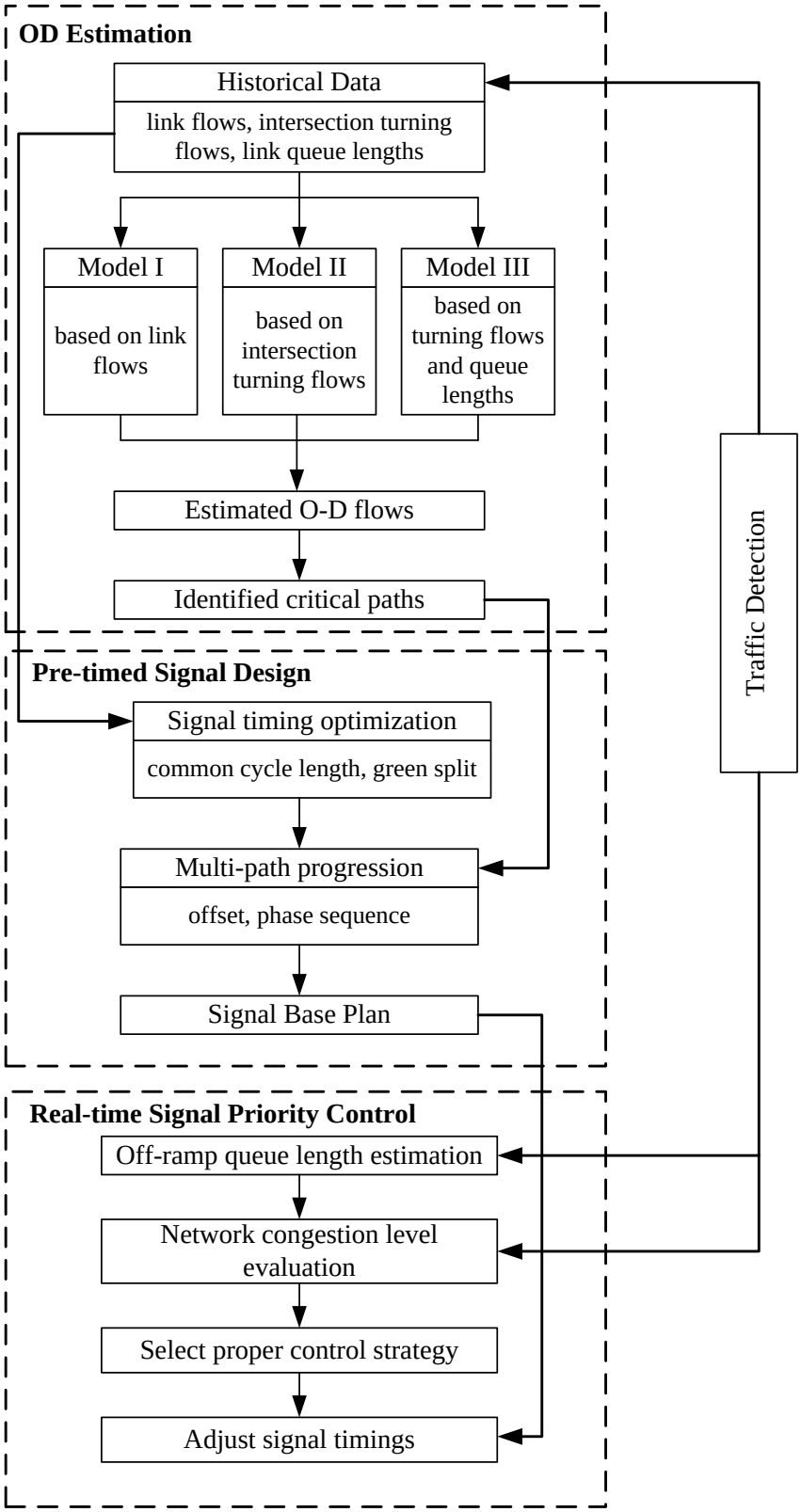


Figure 3.3 A Modeling Framework of the Proposed System

A brief description of each key system function is presented below:

- **O-D matrix estimation**: this function aims to estimate the O-D flows at the off-ramp interchanged area and identify the critical path-flows which need to be coordinated. Based on the available data, including link flows, intersection turning flows, and queue length evolution at critical links, three estimation models are proposed for comparison. Chapter 4 will discuss the details of this component, in which a set of formulations based on Kalman Filter are proposed. These three models will also be tested with a field case, and their effectiveness of identifying critical paths will be evaluated.

- **Pre-timed signal optimization**: this function includes two sequential models to optimize the signal plans at the off-ramp connected arterial. The first model uses the intersection turning flows to optimize signal cycle length and green split at each intersection with the objectives of maximizing intersection capacity and preventing off-ramp queue spillover. Based on the identify critical paths, the second model, a multi-path progression model, optimizes the offset and phase sequence at each intersection so as to facilitate those heavy path-flows to reach their destinations. Chapter 5 will present the formulations of those optimization models. To guarantee the global optimum, all proposed models are formulated with linear or mixed-integer-linear programming techniques.

- **Real-time signal priority control:** this function integrates three control modules to deal with the traffic fluctuation in practice, based on the real-time traffic information obtained from detectors. The first module is designed to

dynamically estimate the off-ramp queue evolution and to identify whether or not an off-ramp queue spillover will occur in the following signal cycles. If no spillover is predicted, the second module will be implemented to adaptively control the signal plan on the target arterial. Otherwise the third module will be used to provide signal priority control, such as green extension and signal progression, to the off-ramp flows. Chapter 6 will introduce the formulations of these three modules, and the integrated real-time control system will be evaluated with the field case for demonstrating its effectiveness.

Chapter 4 : Origin-Destination Matrix Estimations

4.1 Introduction

As is well recognized, the main purpose of most O-D estimation models is to provide essential information for traffic assignment or network simulation. However, as discussion in the last chapter, design of signal plan at the off-ramp interchanged area have also raised the need of using O-D estimation for identifying critical traffic paths and designing signal progression.

In review of the related literature, it is noticeable that there exist very limited studies on estimating O-D patterns for signalized arterials. Among those, Lou and Yin (2010) used the detected link accounts to develop a decomposition framework for estimating time-varying dynamic O-D flows on an arterial. Their proposed framework is to first decompose the entire arterial into a set of individual intersections, and then to perform the estimation of turning flows with link counts, which in turn serves as the measurements for the arterial O-D estimation. Solution algorithms integrated with extended Kalman filter function are developed for parameter estimation at both the intersection and arterial levels. Note that this pioneering work, however, does not address the impacts of implemented signal plans on observable time-varying link flows, and consequently on the estimation accuracy of the resulting O-D patterns. For estimating time-dependent turning fractions at intersections, Chang and Tao (1998) incorporated additional constraints from signal timing information in their enhanced model. The results of extensive simulation

experiments indicate that such methods can yield more accurate estimation, compared with the base model which ignores the impact of signal timings.

Note that despite the scanty of literature on arterial O-D estimation, there is a large body of models for estimating O-D demands on general networks, and most of those fall in either of the following two main categories: assignment and non-assignment-based models. The former category of studies is grounded on the prerequisites that a reliable prior time-varying O-D set and a dynamic traffic assignment model are available for model construction and estimation (e.g., Yang et al. 1992, 2001; Willumsen, 1984; Cascetta 1984; Keller and Ploss, 1987; Cascetta et al., 1993; Ashok and Ben-Akiva, 1993, 2002; Hazelton, 2000; Sherali and Park, 2001; Lu et al. 2013). To circumvent those prerequisites, some researchers proposed the non-assignment-based models, intending to utilize only the time series of available link volume counts for estimation, and thus reduce the dependency of a dynamic traffic assignment model (Cremer and Keller, 1981; 1987; Cremer, 1984; Bell, 1991; Chang and Wu, 1994; Wu and Chang, 1996; Chang and Tao, 1996, 1999; Wu, 1997; Lin and Chang, 2005, 2006, 2007; Lou and Yin, 2010). However, none of those studies has addressed the critical impacts of signal presence and different timing plans on the time-varying distributions of network O-D flows.

Aside from those two categories of studies, some researchers have taken advantage of additional information, measured from emerging sensing technologies, to increase the developed model's observability. Examples of studies along this line include the use of AVI systems (Dixon and Rilett, 2002; Zhou and Mahmassani, 2006; Chen et al., 2011), vehicle plate scanning (Castillo et al., 2008), sporadic

routing data (Parry and Hazelton, 2012), GPS probe vehicles (Eisenman and List, 2004; Cao et al., 2013), and cell phone data (Sohn and Kim, 2008; Calabrese et al., 2011; Iqbal et al., 2014).

Another dimension of research on network O-D estimation is to develop reliable and efficient computing algorithms for the large number of unknown model parameters. On this regard, transportation scholars have proposed both on-line and off-line algorithms in the literature. For off-line applications, existing estimation methods could be categorized into three groups: 1) generalized least squares models (GLS) by Bell (1983, 1991), Cascetta (1984), Cascetta et al. (1993, 2013); 2) maximum likelihood models (ML) by Spiess (1987) and Cascetta and Nguyen (1988); and 3) Bayesian inference models (BI) by Maher (1983). In all aforementioned methods, the reference time period is divided into a sequence of sliced intervals and the O-D flows of all intervals are estimated concurrently. In contrast, for potential on-line operations, most researchers proposed to solve the O-D estimation models with the Kalman filter technique. Examples of dynamic O-D estimation models developed along this line include Okutani and Stephanedes (1984), Nihan and Davis (1987), Chang and Wu (1994), Ashok and Ben-Akiva (2000, 2002), and Antoniou et al. (2007).

This study follows the research line of non-assignment–based methods, and intends to estimate the time-varying O-D matrix on a signalized arterial. Depending on the information availability in practice, this study has proposed three different models to capture the relations between time-varying O-D flows and the observable measurements. Model-I mainly employs the link flows for arterial time-varying O-D

estimation, while Model-II directly takes the turning flows at each intersection as the primary model input. In view of the emerge of queue detectors for intersection signal control, this study has further taken the measurable queue information in the Model-III formulations, because the evolution of time-varying queue patterns offers valuable information to reflect the interrelations between the observable arterial's O-D flows and its signal plans. More specifically, under the same level of traffic demand, the queue length at a given link is expected to be relatively short, if most of its arriving flows are coordinated with neighboring traffic signals. Hence, by constructing the interrelations between the traffic accounts from detectors, the signal plan, and the observed queue length evolution, one can better estimate the origin of arriving flows (e.g., the turning traffic from side streets or through flows from upstream intersections) and their destinations with proper algorithms.

This chapter presents the formulations and estimation algorithms for the three proposed models, and is organized as follows: Section 4.2 and 4.3 illustrate two basic models using the interrelations between the time-dependent O-D flows and time series measurements of traffic accounts at either links or intersections. An enhanced model with additional queue length measurements is presented in Section 4.4. To evaluate the effectiveness of the proposed model and algorithm, this study has performed extensive simulation experiments along with field data, and reported key results in Section 4.5. The last section summarizes the main contributions of this study and its future extensions.

4.2 Model-I Formulations

Considering a signalized arterial of N intersections as shown in Figure 4.1, this model is proposed to take its link accounts as the primary input for its O-D estimation. Also note that all detectors are assumed to place at the upstream location of each link.

With the available link counts data, the information ready for use includes the time series of entering flows, $q_i(k)$, leaving flows, $y_i(k)$, and link flows, $u_i(k)$. And the objective of this model is to estimate the dynamic OD matrix which consists of $(2N+2)^2$ components.

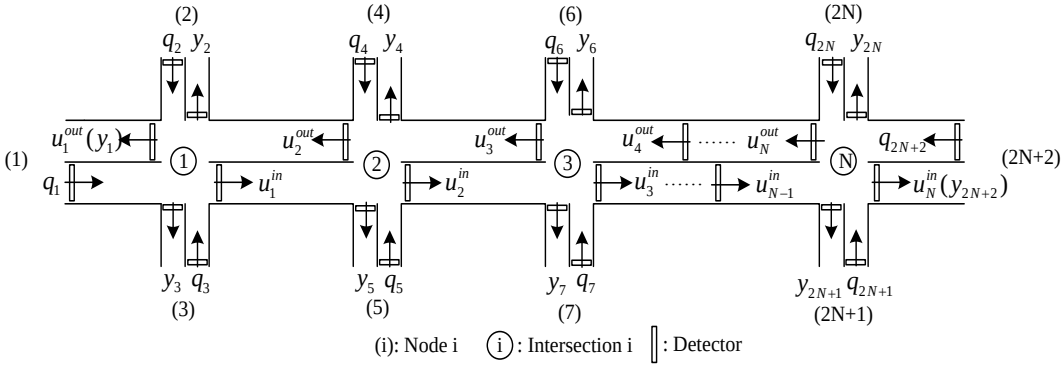


Figure 4.1 A typical local arterial segment

Also, for convenience of discussion, the key notations used hereafter are listed in Table 4.1.

Table 4.1 Key notations for O-D Estimations

$q_i(k)$:	Number of vehicles entering the arterial from node *i* during time interval *k*;
$y_i(k)$:	Number of vehicles leaving the arterial via node *i* during time interval *k*;
$u_l^{in}(k)$:	Number of inbound vehicles passing the arterial detector at intersection *l* during time interval k;
$u_l^{out}(k)$	Number of outbound vehicles passing the arterial detector at intersection *l* during time interval k;
$f_{ij}(k)$:	Number of vehicles entering the arterial from origin node *i* to destination node *j* during time interval *k*;
$\rho_{ij}^m(k)$:	Fraction of $f_{ij}(k)$ vehicles that takes *m* time intervals to reach a detector at node *j* during time interval *k*;
$\theta_{il}^m(k)$:	Fraction of $f_{ij}(k)$ vehicles that takes *m* time intervals to arrive at intersection *l* during time interval *k*;
$b_{ij}(k)$:	Fraction of vehicles departing from node *i* which toward node *j* during time interval *k*;
$\mu_{ij}(k)$:	Average travel time from node *i* to node *j* during time interval *k*;
ΔT:	Length of one unit time interval.
M:	The maximum number of intervals for a vehicle to travel from origin to destination;
N:	Number of intersections at the target signalized arterial;

With the definitions shown in Table 4.1, one can construct the following relations, based on the flow conservations at each intersection:

$$q_i(k) = \sum_{j=1}^{2N+2} f_{ij}(k) \tag{4.1}$$

$$f_{ij}(k) = q_i(k) b_{ij}(k) \tag{4.2}$$

$$\sum_{j=1}^{2N+2} b_{ij}(k) = 1 \tag{4.3}$$

In view of the speed variation among drivers, those concurrently entering node *i* may take a different number of time intervals to arrive at node *j*. Assuming that the travel times of those drivers are distributed among multiple consecutive time intervals (e.g., *M* intervals), one can then formulate the exiting traffic volumes as follows:

$$y_j(k) = \sum_{m=0}^{M} \sum_{i=1}^{2N+2} q_i(k-m) \rho_{ij}^m(k-m) b_{ij}(k-m) \tag{4.4}$$

$$\sum_{m=0}^{M} \rho_{ij}^m(k-m) = 1 \tag{4.5}$$

Also note that the flow measurements between neighboring intersections, $u_i(k)$, can provide additional valuable information to construct the following equations:

$$u_l^{in}(k) = \sum_{m=0}^{M} \sum_{i=1}^{2l+1} \sum_{j=2l+2}^{2N+2} q_i(k-m) \theta_{il}^m(k-m) b_{ij}(k-m) \tag{4.6}$$

$$u_l^{out}(k) = \sum_{m=0}^{M} \sum_{i=2l}^{2N+2} \sum_{j=1}^{2l-1} q_i(k-m) \theta_{il}^m(k-m) b_{ij}(k-m) \tag{4.7}$$

where *l* denotes the index of link flow detectors shown in Figure 4.1.

Hence, one can implement Eqs. (4.1)-(4.7) to capture the dynamic relationships between the O-D patterns and link flows. However, it is noticeable that the above system formulations contain a large number of state variables, i.e., $b_{ij}(k)$, $\rho_{ij}^{m}(k)$, and $\theta_{il}^{m}(k)$. The number of these unknown parameters increases with the required value of *M* (i.e., the maximum time intervals that may take to travel from the origin to the destination on the target arterial). As such, some more information and refinement steps are necessary to ensure the computing efficiency and tractability of this proposed model.

To deal with the large number of unknown parameters, Chang and Wu (1994) simplified the formulations by assuming that the speeds of vehicles entering the network at the same time interval are likely to be distributed in a small range, due to the small size of the target network. Following the same logic, this study assumes that the variation of vehicle travel times between nodes *i* and *j* is within an interval, $[\tau_{ij}^{k-}, \tau_{ij}^{k+}]$. Then, Eqs. (4.1)-(4.7) can be rewritten as follows:

$$y_j(k) = \sum_{i=1}^{2N+2} \left[q_i(k-\tau_{ij}^{k+}) \rho_{ij}^{+}(k-\tau_{ij}^{k+}) b_{ij}(k-\tau_{ij}^{k+}) + q_i(k-\tau_{ij}^{k-}) \rho_{ij}^{-}(k-\tau_{ij}^{k-}) b_{ij}(k-\tau_{ij}^{k-}) \right] \tag{4.8}$$

$$\rho_{ij}^{+}(k) + \rho_{ij}^{-}(k) = 1 \tag{4.9}$$

$$u_l^{in}(k) = \sum_{i=1}^{2l+1} \sum_{j=2l+2}^{2N+2} \left[q_i(k-\tau_{il}^{k+}) \theta_{il}^{+}(k-\tau_{il}^{k+}) b_{ij}(k-\tau_{il}^{k+}) + q_i(k-\tau_{il}^{k-}) \theta_{il}^{-}(k-\tau_{il}^{k-}) b_{ij}(k-\tau_{il}^{k-}) \right] \tag{4.10}$$

$$u_l^{out}(k) = \sum_{i=2l}^{2N+2} \sum_{j=1}^{2l-1} \left[q_i(k-\tau_{il}^{k-}) \theta_{il}^{-}(k-\tau_{il}^{k-}) b_{ij}(k-\tau_{il}^{k-}) + q_i(k-\tau_{il}^{k-}) \theta_{il}^{-}(k-\tau_{il}^{k-}) b_{ij}(k-\tau_{il}^{k-}) \right] \tag{4.11}$$

$$\theta_{il}^{+}(k)+\theta_{il}^{-}(k)=1 \tag{4.12}$$

where, $\tau_{ij}^{k-}=\text{int}[\mu_{ij}(k)/\Delta t]$ and $\tau_{ij}^{k+}=\tau_{ij}^{k-}+1$;

$\rho_{ij}^{+}(k)=\rho_{ij}^{\tau_{ij}^{k+}}(k)$ and $\rho_{ij}^{-}(k)=\rho_{ij}^{\tau_{ij}^{k-}}(k)$;

$\theta_{ij}^{+}(k)=\theta_{ij}^{\tau_{ij}^{k+}}(k)$ and $\rho_{ij}^{-}(k)=\rho_{ij}^{\tau_{ij}^{k-}}(k)$.

4.3 Model-II Formulations

Due to the underdetermine nature of the O-D estimation system, increasing the number of observable measurements can greatly enhance the estimation accuracy. In the literature, Lou and Yin (2010) derived a decomposition model to estimate O-D flows on signalized arterials, where the turning flows, estimated at the first stage, serve as the measurement for estimation at the second stage. However, the potential estimation errors from the estimation by the first stage are likely to propagate to the computations in the second stage.

To overcome this problem, Model-II proposed in this study focuses on modeling the interrelations between the O-D patterns and intersection turning flows since the technology for detecting intersection turning flows is now available in practice.

For convenience of discussion, this study re-defines the notations for turning flows at each intersection. As shown in Figure 4.2, η_{il}^{L}, η_{il}^{T} and η_{il}^{R} denote the flows of left-turn, through, and right-turn movements for leg *i* at intersection *l*.

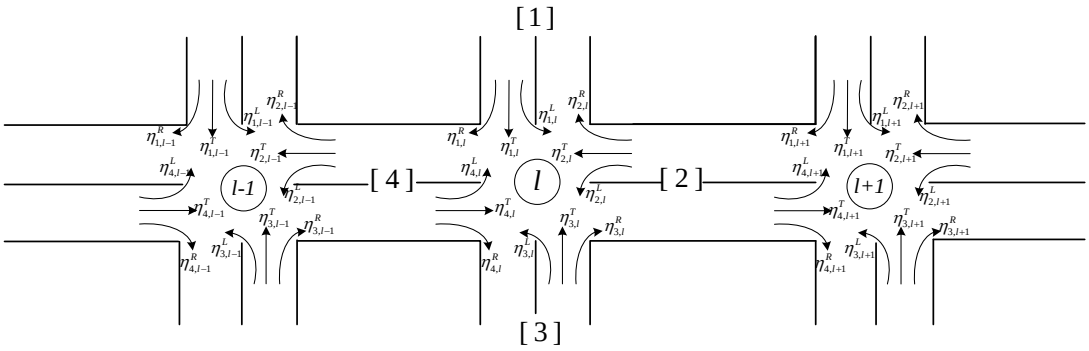


Figure 4.2 Numbering scheme and notations for intersection and turning flows

Then, one can construct the following relations between the O-D flows and turning flows for approaches [1] and [3] of intersection *l* with the notations shown in Figure 4.1:

$$\eta_{1,l}^{L}(k)=\sum_{m=\tau_{2l,l}^{-}}^{\tau_{2l,l}^{-}+1}\sum_{j=2l+2}^{2N+2}q_{2l}(k-m)\theta_{2l,l}^{m}(k-m)b_{2l,j}(k-m) \tag{4.13}$$

$$\eta_{1,l}^{T}(k)=\sum_{m=\tau_{2l,l}^{-}}^{\tau_{2l,l}^{-}+1}q_{2l}(k-m)\theta_{2l,l}^{m}(k-m)b_{2l,2l+1}(k-m) \tag{4.14}$$

$$\eta_{1,l}^{R}(k)=\sum_{m=\tau_{2l,l}^{-}}^{\tau_{2l,l}^{-}+1}\sum_{j=1}^{2l-1}q_{2l}(k-m)\theta_{2l,l}^{m}(k-m)b_{2l,j}(k-m) \tag{4.15}$$

$$\eta_{3,l}^{L}(k)=\sum_{m=\tau_{2l+1,l}^{-}}^{\tau_{2l+1,l}^{-}+1}\sum_{j=1}^{2l-1}q_{2l+1}(k-m)\theta_{2l+1,l}^{m}(k-m)b_{2l+1,j}(k-m) \tag{4.16}$$

$$\eta_{3,l}^{T}(k)=\sum_{m=\tau_{2l+1,l}^{-}}^{\tau_{2l+1,l}^{-}+1}q_{2l+1}(k-m)\theta_{2l+1,l}^{m}(k-m)b_{2l+1,2l}(k-m) \tag{4.17}$$

$$\eta_{3,l}^{R}(k)=\sum_{m=\tau_{2l+1,l}^{-}}^{\tau_{2l+1,l}^{-}+1}\sum_{j=2l+2}^{2N+2}q_{2l+1}(k-m)\theta_{2l+1,l}^{m}(k-m)b_{2l+1,j}(k-m) \tag{4.18}$$

By the same token, one can construct the relations between the O-D flows and turning flows for approaches [2] and [4] of intersection l with the following equations:

$$\eta_{2l}^{L}(k) = \sum_{i=2l+2}^{2N+2} \sum_{m=\tau_{il}^{-}}^{\tau_{il}^{-}+1} q_i(k-m)\theta_{i,l}^{m}(k-m)b_{i,2l+1}(k-m) \tag{4.19}$$

$$\eta_{2l}^{T}(k) = \sum_{i=2l+2}^{2N+2} \sum_{m=\tau_{il}^{-}}^{\tau_{il}^{-}+1} \sum_{j=1}^{2l-1} q_i(k-m)\theta_{i,l}^{m}(k-m)b_{ij}(k-m) \tag{4.20}$$

$$\eta_{2l}^{R}(k) = \sum_{i=2l+2}^{2N+2} \sum_{m=\tau_{il}^{-}}^{\tau_{il}^{-}+1} q_i(k-m)\theta_{i,l}^{m}(k-m)b_{i,2l}(k-m) \tag{4.21}$$

$$\eta_{4l}^{L}(k) = \sum_{i=1}^{2l-1} \sum_{m=\tau_{il}^{-}}^{\tau_{il}^{-}+1} q_i(k-m)\theta_{i,l}^{m}(k-m)b_{i,2l}(k-m) \tag{4.22}$$

$$\eta_{4l}^{T}(k) = \sum_{i=1}^{2l-1} \sum_{m=\tau_{il}^{-}}^{\tau_{il}^{-}+1} \sum_{j=2l+2}^{2N+2} q_i(k-m)\theta_{i,l}^{m}(k-m)b_{ij}(k-m) \tag{4.23}$$

$$\eta_{4l}^{R}(k) = \sum_{i=1}^{2l-1} \sum_{m=\tau_{il}^{-}}^{\tau_{il}^{-}+1} q_i(k-m)\theta_{i,l}^{m}(k-m)b_{i,2l+1}(k-m) \tag{4.24}$$

4.4 Model-III Formulations

Note that the number of measurements in both Model-I and Model-II are always less than the number of parameters to be estimated ($(2N+2)(5N+4)$). Hence, creatively identifying additional observable relations to augment the system's observability is essential to increase the accuracy of such dynamic O-D estimation models. One new set of system relations proposed in Model-III is to formulate the interrelations between the observable intersection and the arterial signal plans, which may vary with the time-varying O-D distributions.

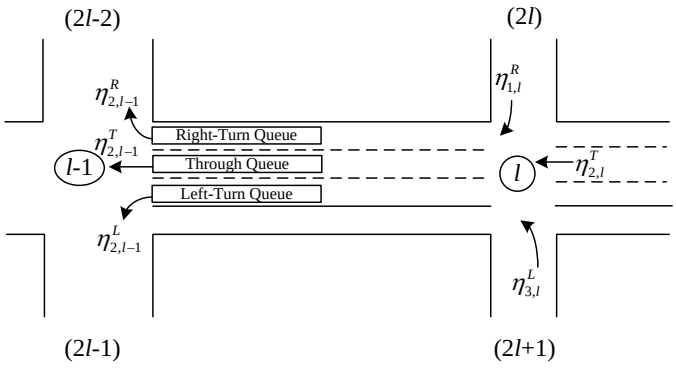


Figure 4.3 Flow diverging and conservation at two adjacent intersections *l* and *l-1*

Figure 4.3 shows a typical arterial segment, consisting of two adjacent intersections. Noticeably, the target link, connecting these two intersections, has three source flows, $\eta_{1,l}^{R}$, $\eta_{2,l}^{T}$, and $\eta_{3,l}^{L}$, and three out-flows, $\eta_{2,l-1}^{L}$, $\eta_{2,l-1}^{T}$, and $\eta_{2,l-1}^{R}$. To hold the flow conservations using traffic counts, one can reach the following relations:

$$\eta_{2,l-1}^{L}(k)+\eta_{2,l-1}^{T}(k)+\eta_{2,l-1}^{R}(k)=\eta_{1,l}^{R}(k-m')+\eta_{2,l}^{T}(k-m')+\eta_{3,l}^{L}(k-m') \tag{4.25}$$

where *m* is the travel time from intersection *l* to intersection *l*-1.

Conceivably, any O-D flow patterns, consistent with Eq. (4.25), could be the same set of feasible solutions yielded by Model-I and Model-II. However, if queue detectors have been deployed, the relations between the queue variation and signals offer another set of constraints to improve model estimation. This is due to the fact that different O-D flow patterns may result in different queue lengths on the connecting links due to the impact of traffic signals. Taking the through queue in Figure 4.3 as an example and assuming that the signal progression is provided to the through traffic along the arterial, Figure 4.4 shows the discrepancy on traffic queue

patterns between two scenarios where the same numbers of vehicles per cycle arrive at the intersection, but come from different origins.

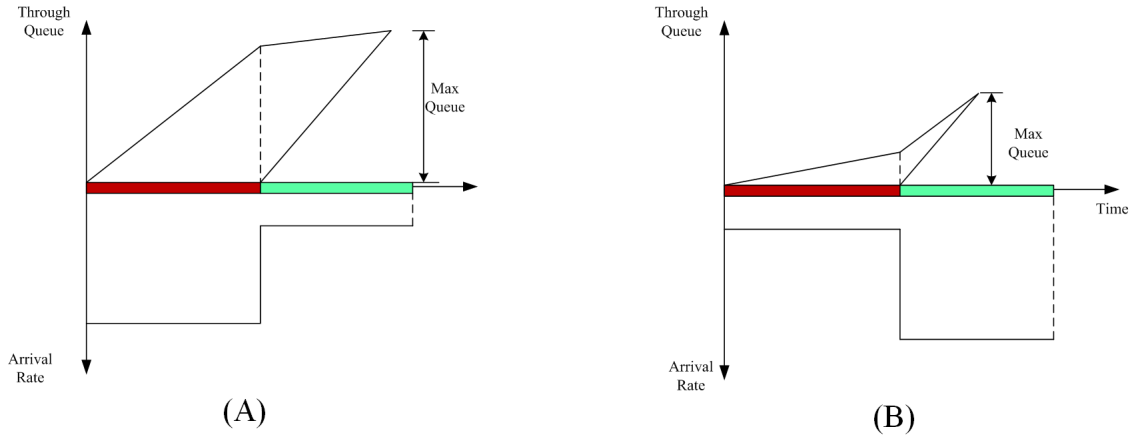


Figure 4.4 The formation of through queue with different arrival patterns

Figure 4.4 (A) illustrates the queue formation patterns when most queuing vehicles are from the turning source flows, such as $\eta_{1,l}^{R}$ and $\eta_{3,l}^{L}$. In contrast, Figure 4.4 (B) shows the resulting queue patterns when most of those vehicles are from the upstream through flows $\eta_{2,l}^{T}$. Note that since the progression is for through movements on the arterial traffic, the through-flows from intersection *l* are most likely to encounter a green phase at intersection *l*-1, whereas the turning flows will experience mainly the red phase. Hence, depending on the contributing sources of vehicles that constitute the arriving flow patterns, the resulting queue pattern may differ significantly even with the same number of arrivals. More specifically, the time-varying queue patterns at an intersection are highly correlated with the O-D flows and intersection signal timings. As such, one may consider incorporating such additional information in formulating the intersection O-D estimation model.

Taking the link shown in Fig. 3 as an example, the enhanced model formulations, using the additional time-varying queue information, are presented below:

$$\delta_{l,l-1}^{T}(k) = \sigma_{l,l-1}^{T}(k) + \varphi_T(\eta_{1,l}^{R}\xi_{1,l}^{T}r_{1,l}^{T} + \eta_{2,l}^{T}\xi_{2,l}^{T}r_{2,l}^{T} + \eta_{3,l}^{L}\xi_{3,l}^{T}r_{3,l}^{T}) \tag{4.26}$$

$$\delta_{l,l-1}^{L}(k) = \sigma_{l,l-1}^{L}(k) + \varphi_L(\eta_{1,l}^{R}\xi_{1,l}^{L}r_{1,l}^{L} + \eta_{2,l}^{T}\xi_{2,l}^{L}r_{2,l}^{L} + \eta_{3,l}^{L}\xi_{3,l}^{L}r_{3,l}^{L}) \tag{4.27}$$

$$\delta_{l,l-1}^{R}(k) = \sigma_{l,l-1}^{R}(k) + \varphi_R(\eta_{1,l}^{R}\xi_{1,l}^{R}r_{1,l}^{R} + \eta_{2,l}^{T}\xi_{2,l}^{R}r_{2,l}^{R} + \eta_{3,l}^{L}\xi_{3,l}^{R}r_{3,l}^{R}) \tag{4.28}$$

where, $\delta_{l,l-1}^{i}(k)$ is the queue length at the end of a red phase on lane group *i*; $\sigma_{l,l-1}^{i}(k)$ is the queue length at the start of a red phase on lane group *i*; φ_i is the lane use factor for lane group *i* which accounts for the uneven distribution of flows on different lanes; $\xi_{i,l}^{j}$ is the ratio of flows, $\eta_{i,l}^{m}$, that will join the downstream flows, $\eta_{2,l-1}^{j}$, where:

$$\xi_{i,l}^{L} + \xi_{i,l}^{T} + \xi_{i,l}^{R} = 1; \ \forall i = 1,2,3 \tag{4.29}$$

And $r_{i,l}^{j}$ is the ratio of uncoordinated flows which can be approximated as follows:

$$r_{i,l}^{j} = \frac{\phi_{i,l}^{j}}{\Delta T} \tag{4.30}$$

where, $\phi_{i,l}^{j}$ is the progression duration within one time interval which can be computed with the signal timings.

For convenience of computation, this study sets the common cycle length to equal the length of estimation time interval, and thus let the value of $r_{i,l}^{j}$ be fixed in a pre-timed signal control system.

Based on Eqs. (4.13)-(4.24), one can further reformulate Eqs. (4.26)-(4.28) as follows:

$$\begin{aligned}\delta_{l,l-1}^{p}(k) = \sigma_{l,l-1}^{p}(k) + \varphi_p[&(\sum_{j=1}^{2l-1}\sum_{m=\tau_{2l,l}^{-}}^{\tau_{2l,l}^{-}+1} q_{2l}(k-m)\theta_{2l,l}^{m}(k-m)b_{2l,j}(k-m))\xi_{1,l}^{p}r_{1}^{p} \\ &+(\sum_{i=2l+2}^{2N+2}\sum_{m=\tau_{il}^{-}}^{\tau_{il}^{-}+1}\sum_{j=1}^{2l-1} q_i(k-m)\theta_{i,l}^{m}(k-m)b_{ij}(k-m))\xi_{2,l}^{p}r_{2,l}^{p} \\ &+(\sum_{j=1}^{2l-1}\sum_{m=\tau_{2l+1,l}^{-}}^{\tau_{2l+1,l}^{-}+1} q_{2l+1}(k-m)\theta_{2l+1,l}^{m}(k-m)b_{2l+1,j}(k-m))\xi_{3,l}^{p}r_{3,l}^{p}];\ \forall p\in\{L,T,R\}\end{aligned} \tag{4.31}$$

Note that Eq.(4.31) is derived for the outbound flows from intersection *l* to *l*-1. For the inbound flows from intersection *l* to *l*+1, one can derive the same relations between $\delta_{l,l+1}^{i}(k)$ and O-D parameters as follows:

$$\begin{aligned}\delta_{l,l+1}^{p}(k) = \sigma_{l,l+1}^{p}(k) + \varphi_T[&(\sum_{j=2l+2}^{2N+2}\sum_{m=\tau_{2l,l}^{-}}^{\tau_{2l,l}^{-}+1} q_{2l}(k-m)\theta_{2l,l}^{m}(k-m)b_{2l,j}(k-m))\xi_{1,l}^{p}r_{1}^{p} \\ &+(\sum_{i=1}^{2l-1}\sum_{m=\tau_{il}^{-}}^{\tau_{il}^{-}+1}\sum_{j=2l+2}^{2N+2} q_i(k-m)\theta_{i,l}^{m}(k-m)b_{ij}(k-m))\xi_{2,l}^{p}r_{2,l}^{p} \\ &+(\sum_{j=2l+2}^{2N+2}\sum_{m=\tau_{2l+1,l}^{-}}^{\tau_{2l+1,l}^{-}+1} q_{2l+1}(k-m)\theta_{2l+1,l}^{m}(k-m)b_{2l+1,j}(k-m))\xi_{3,l}^{p}r_{3,l}^{p}];\ \forall p\in\{L,T,R\}\end{aligned} \tag{4.32}$$

4.5 Estimation Algorithm

4.5.1 Estimation Algorithm for Model-I

As used in most existing approaches, the dynamic O-D parameters, $b_{ij}(k)$, $\rho_{ij}^{-}(k)$, and $\theta_{il}^{-}(k)$, are assumed to follow the random walk process between successive time intervals as shown in the following equations:

$$b_{ij}(k+1) = b_{ij}(k) + w_{ij}^{b}(k), \quad 1 \le i, j \le 2N+2 \tag{4.33}$$

$$\rho_{ij}^{-}(k+1) = \rho_{ij}^{-}(k) + w_{ij}^{\rho}(k), \quad 1 \le i, j \le 2N+2 \tag{4.34}$$

$$\theta_{il}^{-}(k+1) = \theta_{il}^{-}(k) + w_{il}^{\theta}(k), \quad 1 \le i \le 2N+2; 1 \le l \le N \tag{4.35}$$

where the random terms, $w_{ij}^{b}(k)$, $w_{ij}^{\rho}(k)$ and $w_{il}^{\theta}(k)$, are independent Gaussian white noises with zero means. Note that these error terms of O-D parameters do not typically follow a Gaussian distribution because a Gaussian distribution extends infinitely in both directions. However, one can employ a transformation to convert the non-Gaussian data to a Gaussian distribution (MacDougall, 1987).

To facilitate the presentation, let those model parameters be defined as follows:

$$\mathbf{X}(k) = \begin{bmatrix} \mathbf{B}(k) \\ \mathbf{P}(k) \\ \mathbf{\Theta}(k) \end{bmatrix} \tag{4.36}$$

$$\mathbf{W}(k)=\begin{bmatrix}\mathbf{W}^{\mathbf{b}}(k)\\ \mathbf{W}^{\boldsymbol{\rho}}(k)\\ \mathbf{W}^{\boldsymbol{\theta}}(k)\end{bmatrix} \tag{4.37}$$

where, $\mathbf{B}(k)=[b_{ij}(k)]_{(2N+2)^2\times 1}$; $\mathbf{P}(k)=[\rho_{ij}^{-}(k)]_{(2N+2)^2\times 1}$; $\boldsymbol{\Theta}(k)=[\theta_{il}(k)]_{N(2N+2)\times 1}$; and $\mathbf{W}^{\mathbf{b}}(k)$, $\mathbf{W}^{\boldsymbol{\rho}}(k)$, and $\mathbf{W}^{\boldsymbol{\theta}}(k)$ are the matrices of the error terms.

Hence, one can obtain the matrix form of Eqs. (4.13)-(4.15) as follows:

$$\mathbf{X}(k+1)=\mathbf{X}(k)+\mathbf{W}(k) \tag{4.38}$$

With the above transformation for **B**(k), the observation system Eqs. (4.8)-(4.12) can be restructured into the following matrix form:

$$\mathbf{Z}(k+1)=\mathbf{H}(k)\cdot\mathbf{X}(K)+\mathbf{e}(k) \tag{4.39}$$

where, $\mathbf{Z}(k)=[q_1(k),q_2(k),...,q_{2N+2}(k),y_1(k),y_2(k),...,y_{2N+2}(k),u_2(k),u_3(k),...,u_{2N-1}(k)]^T$

Note that **Z**(k) is a column vector of dimension 2(2N+2)+(2N-2), and **e**(k) is an observation noise vector of the same dimension, which can be taken as Gaussian white noises with zero mean and covariance matrix **R**.

Denote *h*(.) as the measurement functions shown by Eqs. (4.1), (4.8), (4.10), and (4.11), then **H**(x) is the Jacobian matrix of partial derivatives of function *h*(.) with respect to the estimation state x:

$$\mathbf{H}_{[rs]}(k)=H_{rs}^{k}=\frac{\partial h_{[r]}(\mathbf{X}(k))}{\partial \mathbf{X}_{[s]}} \tag{4.40}$$

Specifically, Eq. (40) is given by the following equations:

$$\frac{\partial q_i(k)}{\partial b_{ij}(k)} = q_i(k) \tag{4.41}$$

$$\frac{\partial y_j(k)}{\partial b_{ij}(k-\tau_{ij}^{k-})} = q_i(k-\tau_{ij}^{k-})\rho_{ij}^{-}(k-\tau_{ij}^{k-}) \tag{4.42}$$

$$\frac{\partial y_j(k)}{\partial \rho_{ij}^{-}(k-\tau_{ij}^{k-})} = q_i(k-\tau_{ij}^{k-})b_{ij}(k-\tau_{ij}^{k-}) \tag{4.43}$$

$$\frac{\partial u_l^{in}(k)}{\partial b_{ij}(k-\tau_{il}^{k-})} = q_i(k-\tau_{il}^{k-})\theta_{i,l}^{-}(k-\tau_{il}^{k-}), \quad i \le l+1;\ j \ge l+2 \tag{4.44}$$

$$\frac{\partial u_l^{out}(k)}{\partial b_{ij}(k-\tau_{il}^{k-})} = q_i(k-\tau_{il}^{k-})\theta_{i,l}^{-}(k-\tau_{il}^{k-}), \quad i \ge l+1;\ j \le l \tag{4.45}$$

$$\frac{\partial u_l^{in}(k)}{\partial \theta_{i,l}^{-}(k-\tau_{il}^{k-})} = q_i(k-\tau_{il}^{k-})\sum_{j=2l+2}^{2N+2} b_{ij}(k-\tau_{il}^{k-}) \tag{4.46}$$

$$\frac{\partial u_l^{out}(k)}{\partial \theta_{i,l}^{-}(k-\tau_{il}^{k-})} = q_i(k-\tau_{il}^{k-})\sum_{j=1}^{2l-1} b_{ij}(k-\tau_{il}^{k-}) \tag{4.47}$$

So far, a canonical state-space system model, Eqs. (4.36)-(4.40), has been established. Due to the nonlinear nature of the formulations and the concern of computing efficiency, this study has employed the extended Kalman filter algorithm to perform the sequential estimation. A step-by-step illustration of the estimation algorithm is presented in Table 4.2:

Table 4.2 Estimation Algorithm using extended Kalman filter

Step 0: Initialization

- Set the length of time interval Δt;
- Initial the parameters $b_{ij}(0)$, $\rho_{ij}^{-}(0)$, and $\theta_{il}^{-}(0)$.

Step 1: Compute the mean travel time

$$\mu_{ij}(k) = \frac{L_{ij}}{v_{ij}(k)} + \sum_{l \in \delta_{ij}} d_l(k)$$

where, L_{ij} is the travel distance between nodes i and j; $v_{ij}(k)$ is the average travel speed; δ_{ij} is the set of intersections included in the path i to j; $d_l(k)$ is the average delay at intersection l.

Step 2: Compute the linearized transformation matrix **H**(k) using Eqs. (4.40)-(4.47).

Step 3: Initialization of the extended Kalman filtering.

- Set $\mathbf{X}(0)=[\mathbf{B}(0)\ \mathbf{P}(0)\ \boldsymbol{\Theta}(0)]$;
- Set $\mathbf{P}^{-}(0)$ and Q; where $\mathbf{P}^{-}(k)$ is the covariance matrix of $\mathbf{X}(k)$ and $\mathbf{Q}$ is the variance matrix of $\mathbf{W}(k)$.

Step 4: Extended Kalman filtering iteration

- $\mathbf{X}^{-}(k) = \mathbf{X}(k-1)$; $\mathbf{P}^{-}(k) = \mathbf{P}(k-1) + \mathbf{Q}$
- $\mathbf{K}(k) = \mathbf{P}^{-}(k)\mathbf{H}^{T}(k)(\mathbf{H}(k)\mathbf{P}^{-}(k)\mathbf{H}^{T}(k) + \mathbf{R})^{-1}$
- $\mathbf{X}(k) = \mathbf{X}^{-}(k) + \mathbf{K}(k)\left[\mathbf{z}(k) - h(\mathbf{X}^{-}(k))\right]$
- $\mathbf{P}(k) = \left[\mathbf{I} - \mathbf{K}(k)\mathbf{H}(k)\right]\mathbf{P}^{-}(k)$

Step 5: Go back to *Step* 1 for next interval.

4.5.2 Estimation Algorithm for Model-II

Similar to Model-I, one can also implement the extended Kalman filter to estimate the following dynamic O-D parameters in in Model-II:

$$\mathbf{X}(k)=\begin{bmatrix}\mathbf{B}(k)\\ \mathbf{\Theta}(k)\end{bmatrix} \tag{4.48}$$

$$\mathbf{W}(k)=\begin{bmatrix}\mathbf{W^b}(k)\\ \mathbf{W^\theta}(k)\end{bmatrix} \tag{4.49}$$

where, $\mathbf{B}(k)=[b_{ij}(k)]_{(2N+2)^2\times 1}$; $\mathbf{\Theta}(k)=[\theta_{il}(k)]_{N(2N+2)\times 1}$.

Also, the observation systems can be restructured as a column vector of dimension 12N:

$$\mathbf{Z}(k)=[\eta_{11}^{L}(k),\eta_{11}^{T}(k),\eta_{11}^{R}(k),\cdots,\eta_{4N}^{L}(k),\eta_{4N}^{T}(k),\eta_{4N}^{R}(k)]^{T} \tag{4.50}$$

Then **H**(x), the Jacobian matrix of partial derivative of measurement function *h*(.) with respect to the estimation state **X,** could be computed with the following equations:

$$\frac{\partial \eta_{1l}^{L}(k)}{\partial \theta_{2l,l}^{-}(k-\tau_{2l,l}^{-})}=\sum_{j=2l+2}^{2N+2} q_{2l}(k-\tau_{2l,l}^{-})b_{2l,j}(k-\tau_{2l,l}^{-}) \tag{4.51}$$

$$\frac{\partial \eta_{1l}^{T}(k)}{\partial \theta_{2l,l}^{-}(k-\tau_{2l,l}^{-})}=q_{2l}(k-\tau_{2l,l}^{-})b_{2l,2l+1}(k-\tau_{2l,l}^{-}) \tag{4.52}$$

$$\frac{\partial \eta_{1l}^{R}(k)}{\partial \theta_{2l,l}^{-}(k-\tau_{2l,l}^{-})}=\sum_{j=1}^{2l-1} q_{2l}(k-\tau_{2l,l}^{-})b_{2l,j}(k-\tau_{2l,l}^{-}) \tag{4.53}$$

$$\frac{\partial \eta_{1l}^{L}(k)}{\partial b_{2l,\mathrm{j}}(k-\tau_{2l,l}^{-})}=\frac{\partial \eta_{1l}^{T}(k)}{\partial b_{2l,2l+1}(k-\tau_{2l,l}^{-})}=\frac{\partial \eta_{1l}^{R}(k)}{\partial b_{2l,j}(k-\tau_{2l,l}^{-})}=q_{2l}(k-\tau_{2l,l}^{-})\theta_{2l,l}^{-}(k-\tau_{2l,l}^{-}) \quad (4.54)$$

$$\frac{\partial \eta_{2,l}^{L}(k)}{\partial \theta_{i,l}^{-}(k-\tau_{i,l}^{-})}=q_i(k-\tau_{i,l}^{-})b_{i,2l+1}(k-\tau_{i,l}^{-}) \quad (4.55)$$

$$\frac{\partial \eta_{2,l}^{T}(k)}{\partial \theta_{i,l}^{-}(k-\tau_{i,l}^{-})}=\sum_{j=1}^{2l-1}q_i(k-\tau_{i,l}^{-})b_{ij}(k-\tau_{i,l}^{-}) \quad (4.56)$$

$$\frac{\partial \eta_{2,l}^{R}(k)}{\partial \theta_{i,l}^{-}(k-\tau_{i,l}^{-})}=q_i(k-\tau_{i,l}^{-})b_{i,2l}(k-\tau_{i,l}^{-}) \quad (4.57)$$

$$\frac{\partial \eta_{2,l}^{L}(k)}{\partial b_{i,2l+1}(k-\tau_{i,l}^{-})}=\frac{\partial \eta_{2,l}^{T}(k)}{\partial b_{ij}(k-\tau_{i,l}^{-})}=\frac{\partial \eta_{2,l}^{R}(k)}{\partial b_{i,2l}(k-\tau_{i,l}^{-})}=q_i(k-\tau_{i,l}^{-})\theta_{i,l}^{-}(k-\tau_{i,l}^{-}) \quad (4.58)$$

$$\frac{\partial \eta_{3,l}^{L}(k)}{\partial \theta_{2l+1,l}^{-}(k-\tau_{2l+1,l}^{-})}=\sum_{j=1}^{2l-1}q_{2l+1}(k-\tau_{2l+1,l}^{-})b_{2l+1,l}(k-\tau_{2l+1,l}^{-}) \quad (4.59)$$

$$\frac{\partial \eta_{3l}^{T}(k)}{\partial \theta_{2l+1,l}^{-}(k-\tau_{2l+1,l}^{-})}=q_{2l+1}(k-\tau_{2l+1,l}^{-})b_{2l,2l+1}(k-\tau_{2l+1,l}^{-}) \quad (4.60)$$

$$\frac{\partial \eta_{3,l}^{R}(k)}{\partial \theta_{2l+1,l}^{-}(k-\tau_{2l+1,l}^{-})}=\sum_{j=2l+2}^{2N+2}q_{2l+1}(k-\tau_{2l+1,l}^{-})b_{2l+1,l}(k-\tau_{2l+1,l}^{-}) \quad (4.61)$$

$$\frac{\partial \eta_{3,l}^{L}(k)}{\partial b_{2l+1,\mathrm{j}}(k-\tau_{2l+1,l}^{-})}=\frac{\partial \eta_{3,l}^{T}(k)}{\partial b_{2l,2l+1}(k-\tau_{2l+1,l}^{-})}=\frac{\partial \eta_{3,l}^{R}(k)}{\partial b_{2l,j}(k-\tau_{2l+1,l}^{-})}=q_{2l+1}(k-\tau_{2l,l}^{-})\theta_{2l+1,l}^{-}(k-\tau_{2l+1,l}^{-}) \quad (4.62)$$

$$\frac{\partial \eta_{4,l}^{L}(k)}{\partial \theta_{i,l}^{-}(k-\tau_{i,l}^{-})}=q_i(k-\tau_{i,l}^{-})b_{i,2l}(k-\tau_{i,l}^{-}) \quad (4.63)$$

$$\frac{\partial \eta_{4,l}^{T}(k)}{\partial \theta_{i,l}^{-}(k-\tau_{i,l}^{-})} = \sum_{j=2l+2}^{2N+2} q_i(k-\tau_{i,l}^{-}) b_{ij}(k-\tau_{i,l}^{-}) \tag{4.64}$$

$$\frac{\partial \eta_{4,l}^{R}(k)}{\partial \theta_{i,l}^{-}(k-\tau_{i,l}^{-})} = q_i(k-\tau_{i,l}^{-}) b_{i,2l+1}(k-\tau_{i,l}^{-}) \tag{4.65}$$

$$\frac{\partial \eta_{4,l}^{L}(k)}{\partial b_{i,2l}(k-\tau_{i,l}^{-})} = \frac{\partial \eta_{4,l}^{T}(k)}{\partial b_{ij}(k-\tau_{i,l}^{-})} = \frac{\partial \eta_{4,l}^{R}(k)}{\partial b_{i,2l+1}(k-\tau_{i,l}^{-})} = q_i(k-\tau_{i,l}^{-}) \theta_{i,l}^{-}(k-\tau_{i,l}^{-}) \tag{4.66}$$

Then, one can implement the estimation algorithm presented in Table 4.2 to estimate the O-D flows with Model-II.

4.5.3 Estimation Algorithm for Model-III

By the same token, one can also implement the extended Kalman filter to estimate the time-varying O-D parameters in Model-III. Notably, the measurement vector, **Z**(k), shall be expanded due to the availability of additional time-varying queue information on each monitoring link:

$$\mathbf{Z}(k) = [\mathbf{H}^{T}(k), \mathbf{\Delta}_{o}^{T}(k), \mathbf{\Delta}_{i}^{T}(k)]^{T} \tag{4.67}$$

$$\mathbf{H}(k) = [\eta_{11}^{L}(k), \eta_{11}^{T}(k), \eta_{11}^{R}(k), \cdots, \eta_{4N}^{L}(k), \eta_{4N}^{T}(k), \eta_{4N}^{R}(k)]^{T} \tag{4.68}$$

$$\mathbf{\Delta}_{o}^{T}(k) = [\delta_{2,1}^{L}(k)\delta_{2,1}^{T}(k)\delta_{2,1}^{R}(k), \cdots, \delta_{N,N-1}^{L}(k)\delta_{N,N-1}^{T}(k)\delta_{N,N-1}^{R}(k)]^{T} \tag{4.69}$$

$$\mathbf{\Delta}_{i}^{T}(k) = [\delta_{1,2}^{L}(k)\delta_{1,2}^{T}(k)\delta_{1,2}^{R}(k), \cdots, \delta_{N-1,N}^{L}(k)\delta_{N-1,N}^{T}(k)\delta_{N-1,N}^{R}(k)]^{T} \tag{4.70}$$

Then, the additional elements in **H**(x) matrix could be computed as follows:

$$\frac{\partial \delta_{l,l-1}^{p}(k)}{\partial \theta_{i,l}^{-}(k-\tau_{i,l}^{-})}=\begin{cases}\varphi_p \xi_{1,l}^{p} r_{1,l}^{p} \sum_{j=1}^{2l-1} q_{2l}(k-\tau_{i,l}^{-}) b_{2l,j}(k-\tau_{i,l}^{-}); & i=2l \\ \varphi_p \xi_{2,l}^{p} r_{2,l}^{p} \sum_{j=1}^{2l-1} q_{i}(k-\tau_{i,l}^{-}) b_{ij}(k-\tau_{i,l}^{-}); & 2l+2 \le i \le 2N+2 \\ \varphi_p \xi_{3,l}^{p} r_{3,l}^{p} \sum_{j=1}^{2l-1} q_{2l+1}(k-\tau_{i,l}^{-}) b_{2l+1,j}(k-\tau_{i,l}^{-}); & i=2l+1\end{cases} \tag{4.71}$$

$$\frac{\partial \delta_{l,l-1}^{p}(k)}{\partial b_{ij}(k-\tau_{i,l}^{-})}=\begin{cases}\varphi_p \xi_{1,l}^{p} r_{1,l}^{p} q_i(k-\tau_{i,l}^{-}) \theta_{i,l}^{m}(k-\tau_{i,l}^{-}); & i=2l \\ \varphi_p \xi_{2,l}^{p} r_{2,l}^{p} q_i(k-\tau_{i,l}^{-}) \theta_{i,l}^{m}(k-\tau_{i,l}^{-}); & 2l+2 \le i \le 2N+2 \\ \varphi_p \xi_{3,l}^{p} r_{3,l}^{p} q_i(k-\tau_{i,l}^{-}) \theta_{i,l}^{m}(k-\tau_{i,l}^{-}); & i=2l+1\end{cases} \tag{4.72}$$

Then, one can implement the estimation algorithm presented in Table 4.2 to estimate the O-D flows with Model-III.

4.6 Numerical Test

4.6.1 Experimental Design

To assess the proposed models' potential for field applications, this dissertation has further selected an arterial segment in Guangming 6th Rd, Chupei, Taiwan for case study. The geometric layout of the target arterial and its topology are shown in Figure 4.5 and Figure 4.6. Note that the arterial consists of 5 intersections, 12 nodes, and 34 links. Also, both intersections 3 and 4 contain one freeway on-ramp and off-ramp. Hence, the number of O-D flows to be estimated will be reduced to 48 per time interval.

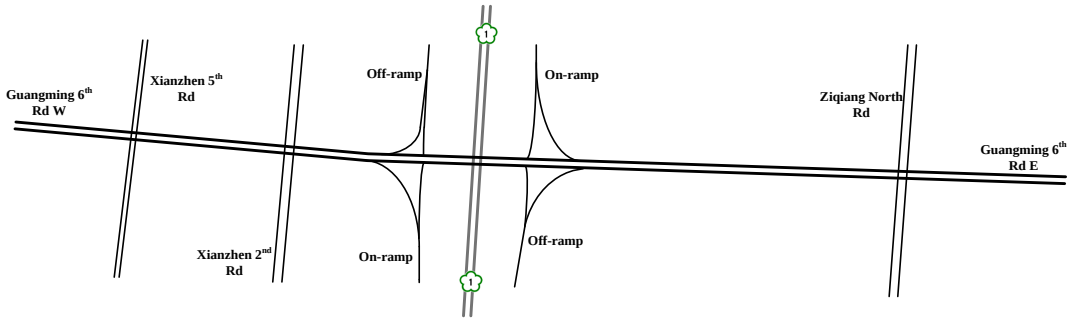


Figure 4.5 The geometric layout of the study site

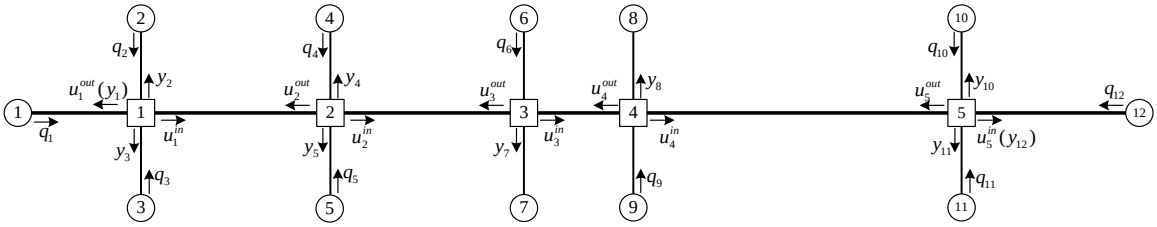


Figure 4.6 Arterial topology of the study site

To replicate the field traffic conditions, this study has selected VISSIM as the simulation platform, and used the observed intersection queues and link volumes from 16:30 to 21:30 on April 24, 2013 for calibration. The calibrated simulation is then used to generate simulated scenarios for model evaluation. Table 4.3 summarizes the demand patterns from each source generated with the simulated system, where the simulation period is set to be 2 hours and 4 time periods (30 minutes each).

The phasing plan, signal timings, and original phase sequence at each intersection are presented in Figure 4.7. The common cycle length is set to be 150 seconds.

Table 4.3 Time-varying in-flow at each source node

Time period (seconds)	Node Number									
	1	2	3	4	5	6	9	10	11	12
0-1800	560	210	210	753	547	882	1470	420	490	840
1800-3600	640	240	240	860	625	1008	1680	480	560	960
3600-5400	720	270	270	968	703	1134	1890	540	630	1080
5400-7200	640	240	240	860	625	1008	1680	480	560	960

*Demand Unit: vehicles/hour

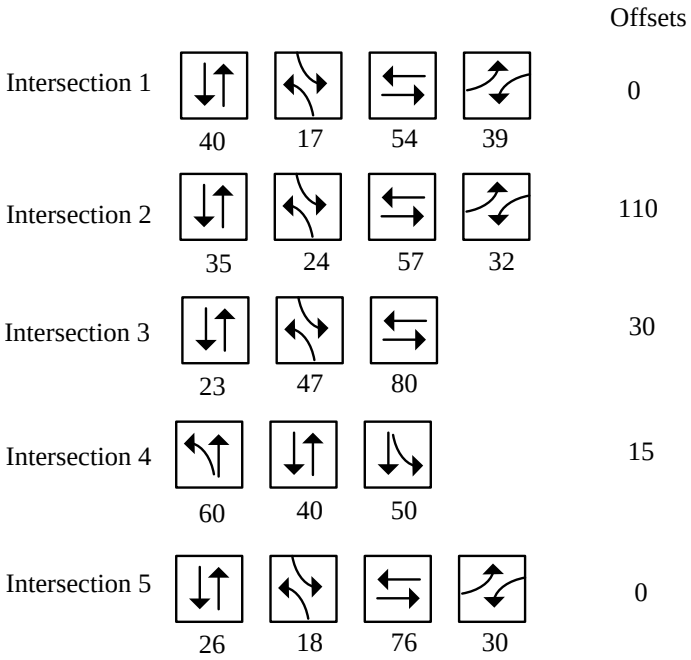


Figure 4.7 Signal timings and initial phase sequence

Using the well-calibrated VISSIM network, this study has further taken advantage of the following information for model assessment: 1) GPS data with 'one second' time interval for each vehicle; and 2) loop detector and queue detector data.

Then, one can use the GPS data to track the trajectory of each individual vehicle and subsequently identify the ground-true O-D flow patterns. Also note that the three proposed models require different sets of detectors. For Model-I, each link upstream is installed with one loop detector for flow data collection, while loop detectors are installed at the downstream of each link in Model-II so as to capture the turning flows at each intersection. For Model III, four additional queue detectors are needed at those four links between neighboring intersections (i.e., intersections 2-3, intersections 4-5).

4.6.2 Model Evaluation and Results

To evaluate the effectiveness of the proposed three models, the dynamic O-D flows estimated with different models are to compare with the ground-true O-D flows obtained with GPS data. Also, Model-I is treated as a benchmark to measure the benefits of other two enhanced models. Comparisons of estimation accuracy are based on the following three key outputs: link flow counts, turning flows at all intersections, and dynamic O-D flows. Figures 4.8-4.16 show the comparisons between those estimated and ground truth patterns, displayed in the format of scatter plots.

The quality of accuracy is measured with the following indicators: MAE (Mean Absolute Error), MAPE (Mean Absolute Percentage Error) and RMSE (Root Mean Square Error). All performance indicators used hereafter are defined below:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|x_i - \hat{x}_i\right|$$

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{x_i - \hat{x}_i}{\hat{x}_i}\right|$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(x_i - \hat{x}_i\right)^2}$$

where, x_i and $\hat{x}_i$ are estimated and ground-true values, respectively; N is the number of estimated states.

Table 4.4 summarizes all the performance indicators for O-D flows, link flows, and turning flows given by different models.

Table 4.4 Estimation Accuracy of O-D Flows

Models	Base Model-I			Base Model-II			Enhanced Model-III		
	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE
Link flows	4.54	18.56%	5.48	4.10	16.31%	5.21	3.99	15.92%	4.99
Turning flows	4.02	42.39%	5.54	2.75	18.27%	4.07	2.70	17.46%	3.92
OD flows	1.885	42.02%	3.075	1.473	33.20%	2.512	1.251	28.11%	1.979

*Note: the unit of MAE and RMSE is “vehicle”

As shown in Figure 4.8, Model I can produce quite accurate estimation of link flows, and its resulting MAPE is around 15.86%. This is due to the fact that link flows are the major measurements in Model-I, and the estimation results are corrected by Kalman filter over subsequent time intervals. However, Model-I cannot yield

sufficiently reliable estimation for both turning flows and O-D flows, as shown in Table 4.4, where the percentage errors can reach as high as 42% in MAPE. Notably, the target arterial with Model-I has 28 available measurements for estimation, but has over 300 unknown parameters.

Taking turning flows at each intersection as the measurements, Model-II can clearly outperform Model-I in estimating O-D patterns. As shown in Table 4.4, the MAPE of turning flows and O-D flows with Model-II have been reduced to 18.27% and 33.20%, respectively. Also note that the sum of turning flows at each link will equal the collected link flows when neglecting the travel time impact. However, comparing the results between the two models, it is evident that increasing the number of observable measurements can clearly improve the estimation accuracy.

Evidenced by the results in Table 3 and Figures 4.8-4.16, Model-III can produce the most accurate estimates over the other two models. The estimated accuracy with respect to link flows, turning flows, and OD flows, based on MAPE, are 15.92%, 17.46% and 28.11%, respectively. As stated in the last section, Model-III has taken real-time queue information at several critical links as the additional measurement over Model-II. To confirm the effectiveness contributed by incorporating the queue information, one can further compare the results between Model-II and Model-III. For instance, based on the data in Table 3, it is evident that the improvements for O-D flow estimation by Model III over Model II with respect to MAE, MAPE, and RMSE, are calculated as 15.07%, 15.33% and 21.21%, respectively.

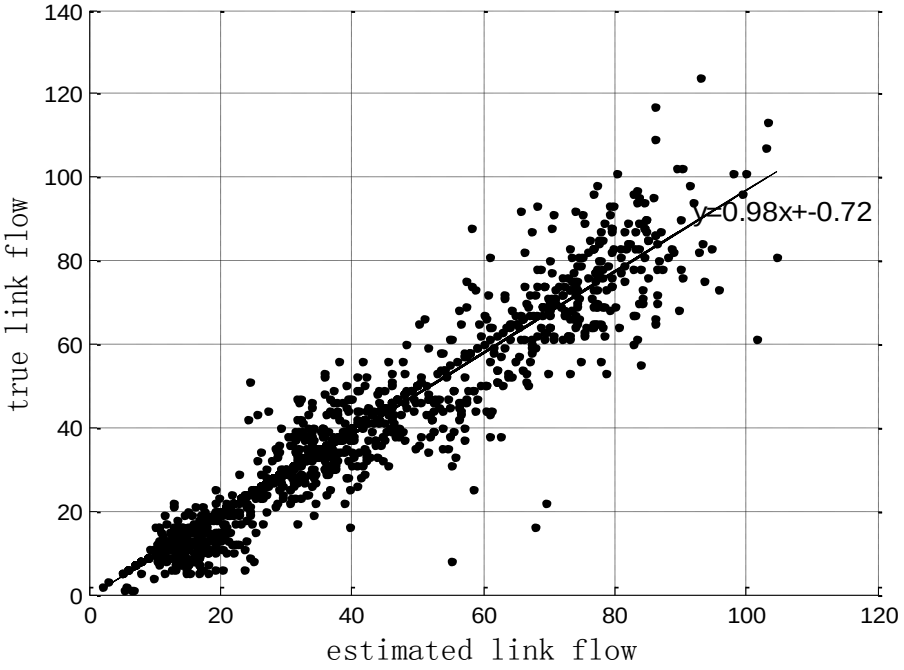


Figure 4.8 Estimation accuracy for link flow by Model-I

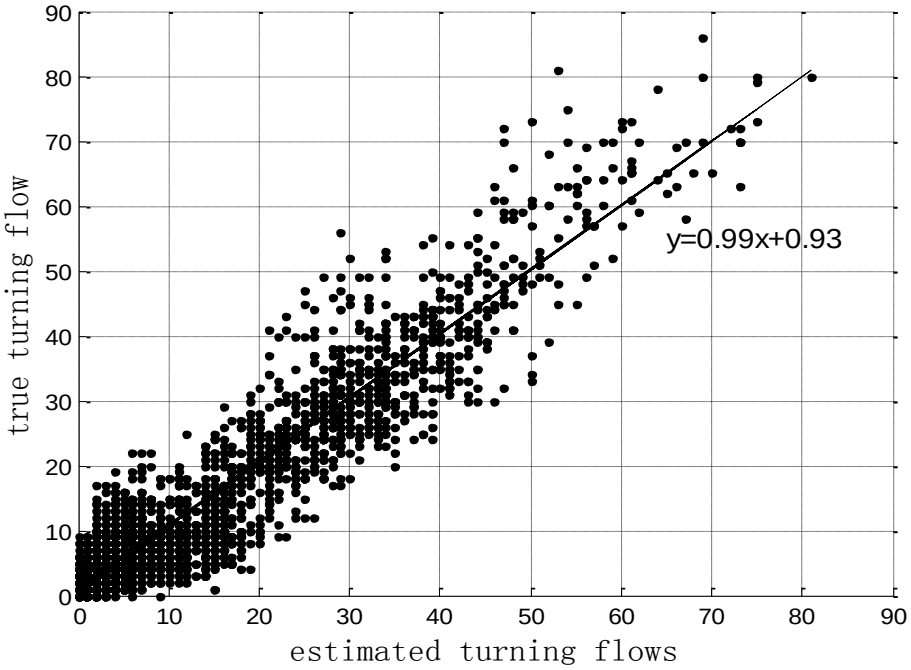


Figure 4.9 Estimation accuracy for turning flow by Model-I

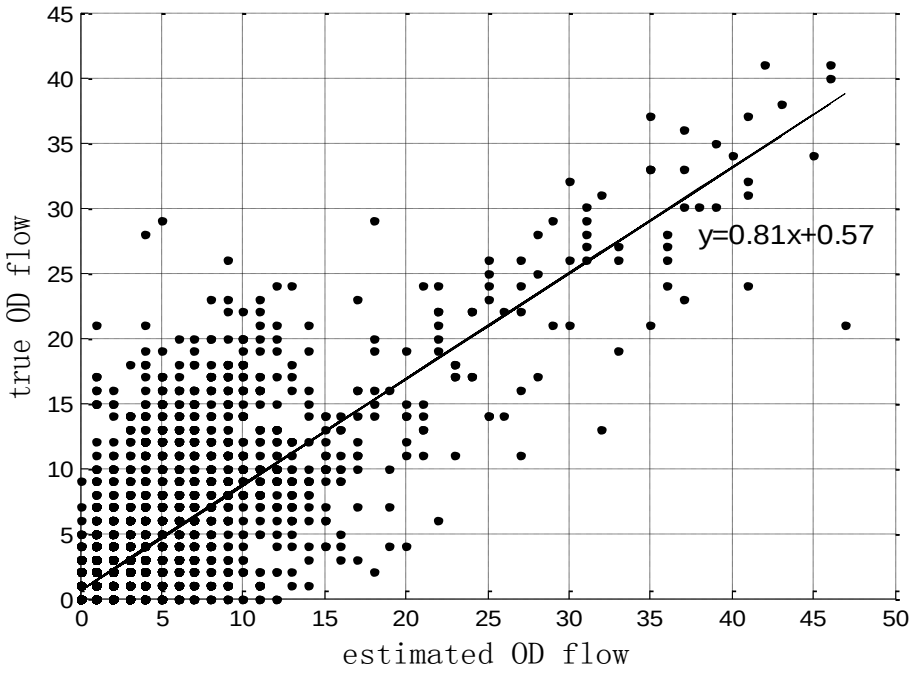


Figure 4.10 Estimation accuracy for OD flow by Model-I

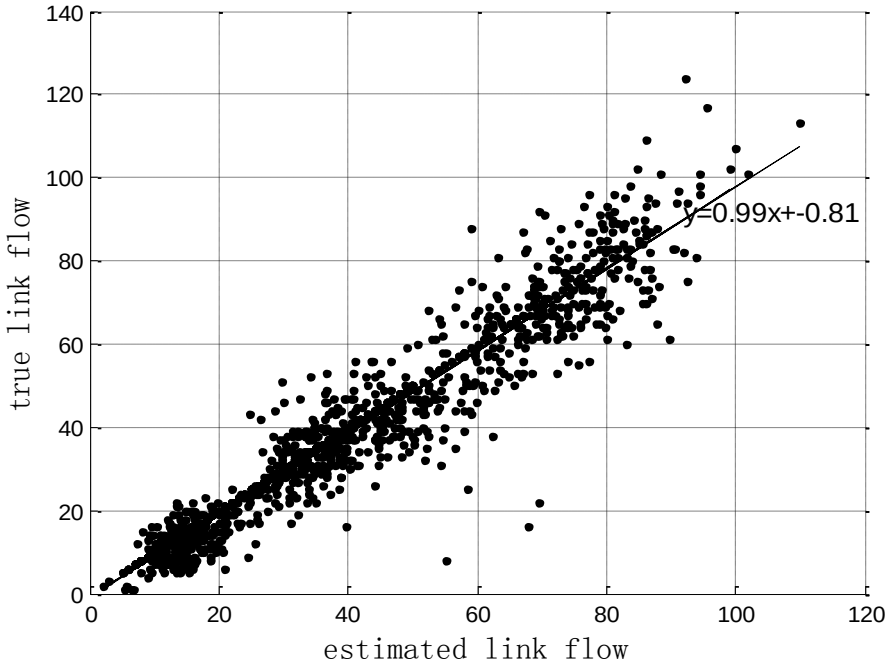


Figure 4.11 Estimation accuracy for link flow by Model-II

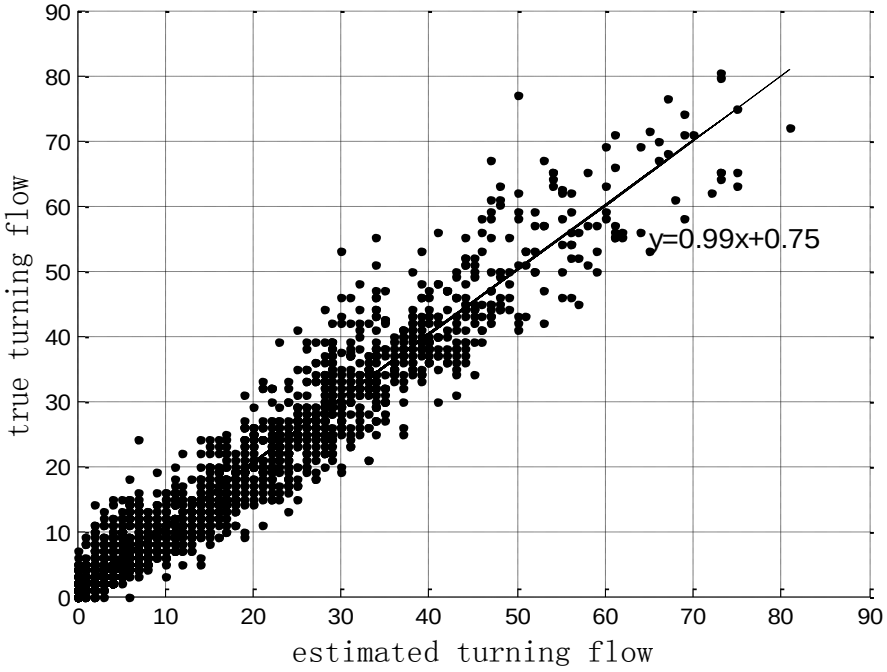


Figure 4.12 Estimation accuracy for turning flow by Model-II

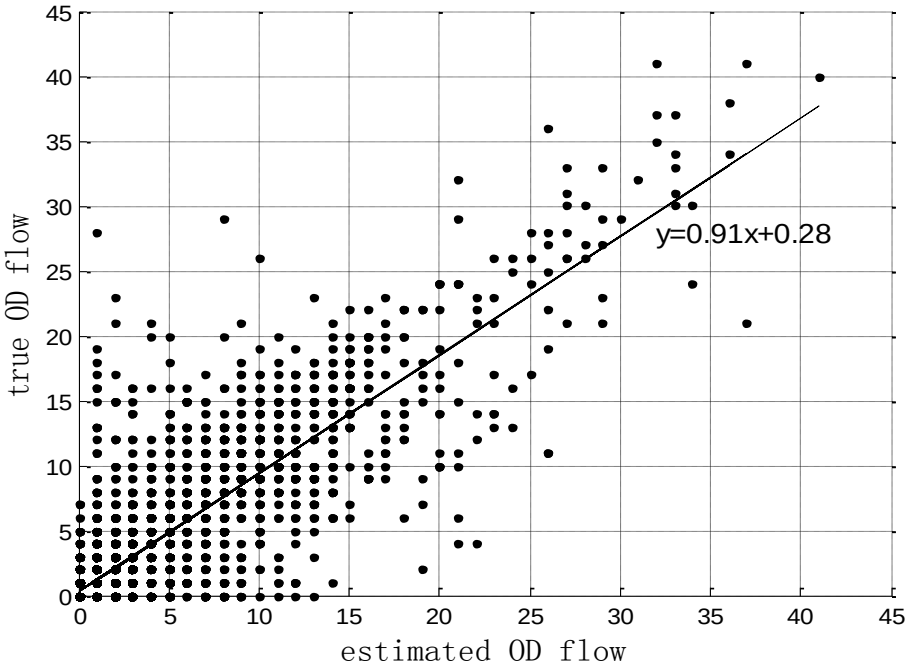


Figure 4.13 Estimation accuracy for OD flow by Model-II

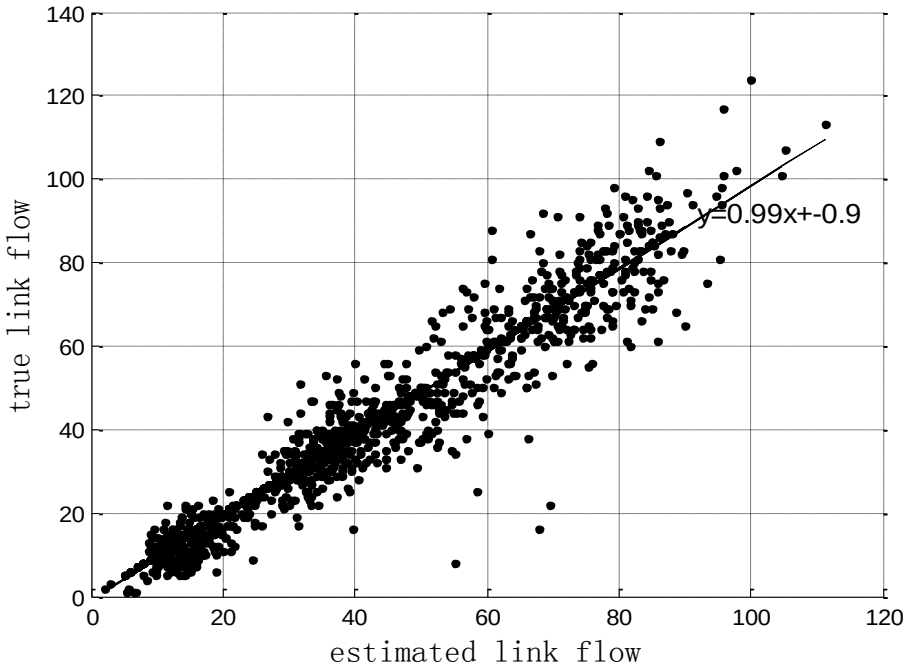


Figure 4.14 Estimation accuracy for link flow by Model-III

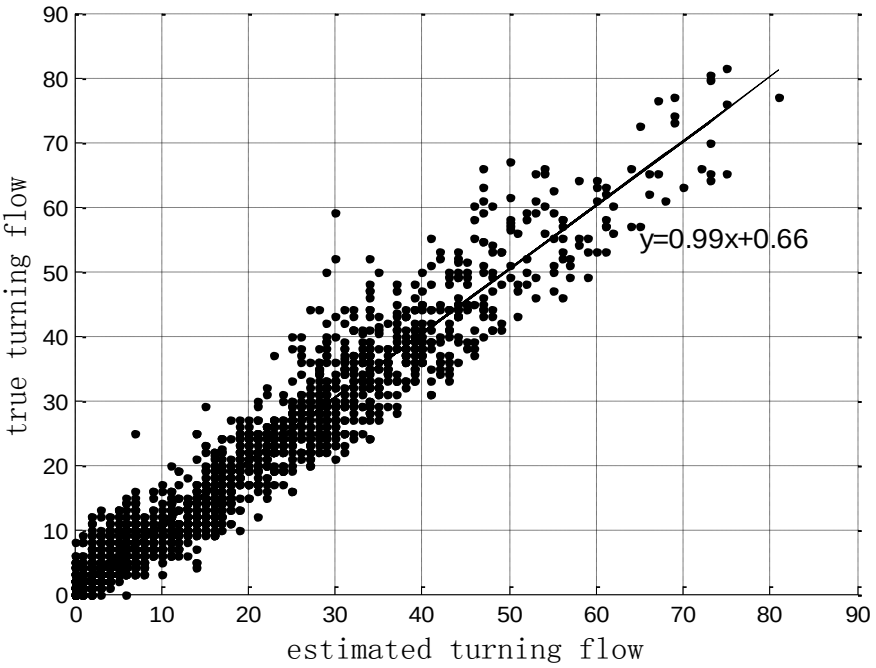


Figure 4.15 Estimation accuracy for turning flow by Model-III

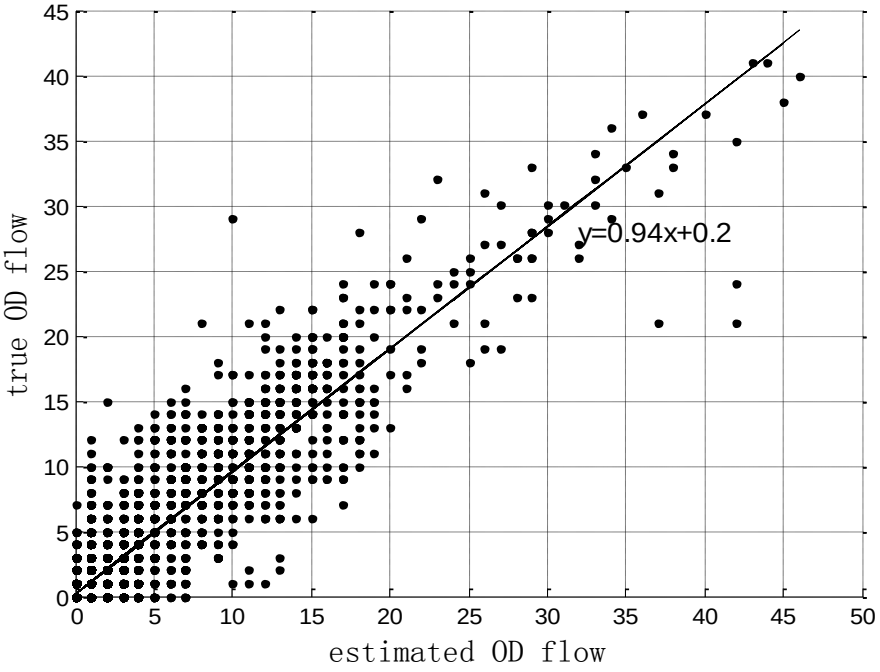


Figure 4.16 Estimation accuracy for OD flow by Model-III

Table 4.5 The ground-true and identified critical paths by proposed models

Ground Truth		Base Model-I		Base Model-II		Enhanced Model-III	
OD Pair	Total Flows	OD Pair	Total Flows	OD Pair	Total Flows	OD Pair	Total Flows
9→12	1390	**9→12**	1658	**9→12**	1372	**9→12**	1480
6→12	765	**6→12**	985	**6→12**	860	**6→12**	784
9→1	756	9→4	649	9→4	727	**9→1**	722
6→4	729	4→7	497	4→7	571	**6→4**	642
12→7	553	4→8	465	12→8	544	**12→7**	540
12→1	472	**9→1**	427	**9→1**	531	**12→1**	452

As stated in the Introduction Section, a reliable estimate of arterial O-D flows offers the essential information to identify the critical traffic paths (O-D pairs) for

design of signal progression. Therefore, the critical paths identified with above three O-D models are also compared below. Note that those critical paths are identified by ranking the O-D volumes and eliminating those pairs span only one intersection.

As shown in Table 4.4, both Model-I and Model-II can identify only three out of six actual critical paths. However, Model-III can successfully identify all six critical paths with the correct rankings. Therefore, one can argue that the enhanced model is more reliable in terms of identifying major path-flow patterns to design a multi-path signal progression system. To further verify the effectiveness of incorporating real-time queue information in O-D flow estimation, Figure 4.17-4.20 show the comparisons between estimated and actual O-D flows.

As shown in Figures 4.17-4.18, all three proposed models are able to track the trend of the first two critical OD flows (9→12; 6→12) and produce estimates sufficiently close to the ground-true values. Figures 4.19-4.20 indicate that the two base models seem to produce biased estimation for the O-D flows (9→1) and (6→4). In contrast, the enhanced model can still capture the changes of flows via these two O-D paths.

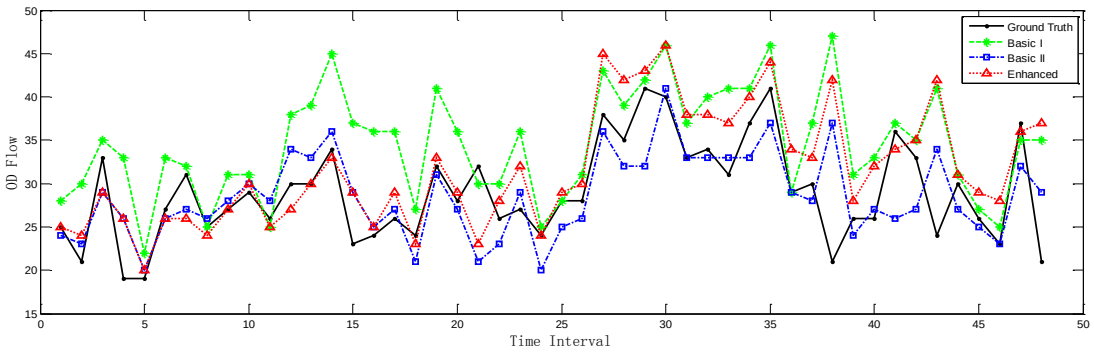


Figure 4.17 time-dependent flows for O-D 9→12

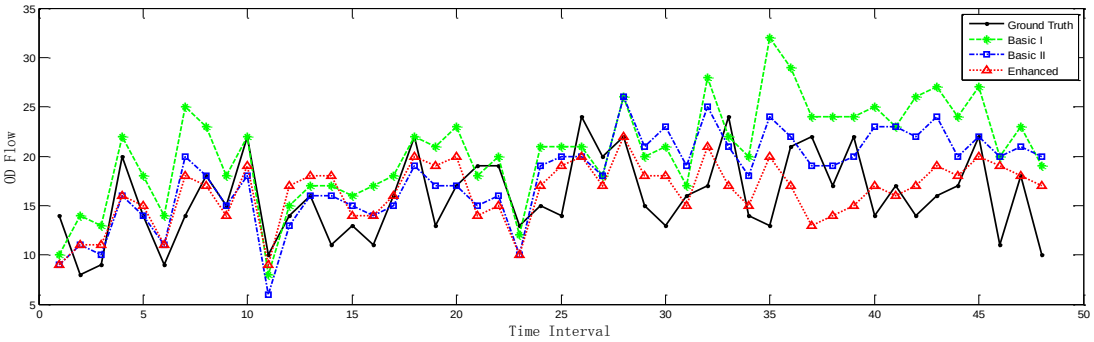


Figure 4.18 time-dependent flows for O-D 6→12

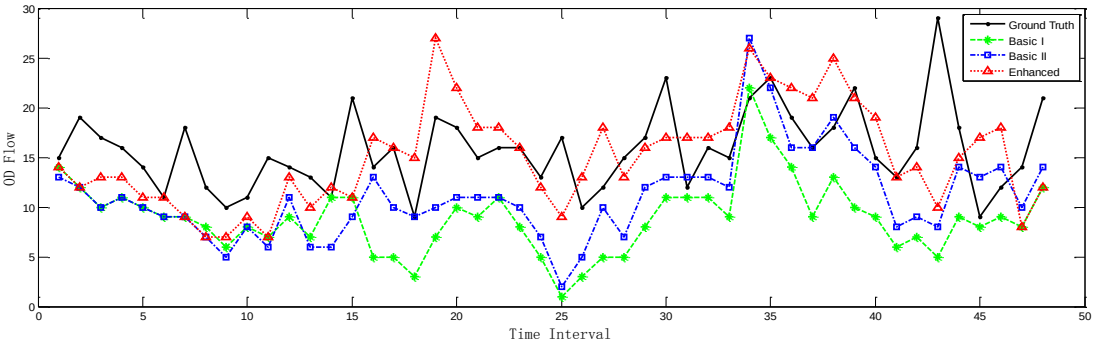


Figure 4.19 time-dependent flows for O-D 9→1

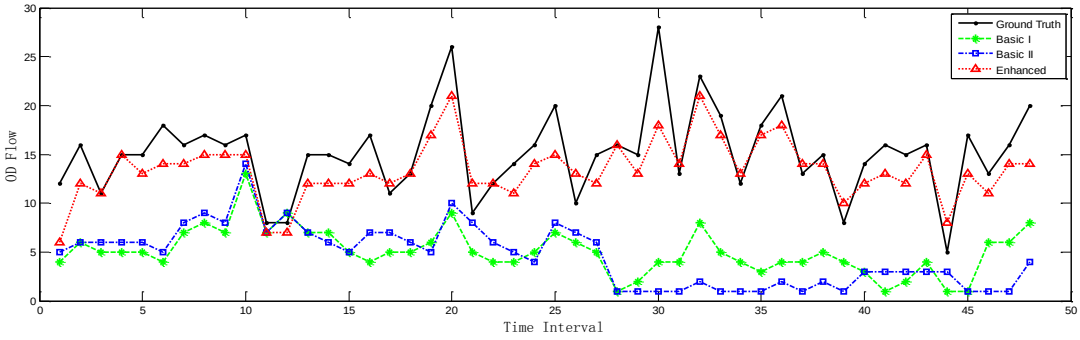


Figure 4.20 time-dependent flows for O-D 6→4

To investigate the resulting biases by these two base models, this study has further analyzed the collected link flow patterns, and found out that both node 1 and

node 4 have extremely high exit flows. Also, there are two major source flows (from node 6 and node 9) may contribute to the exit flows. The ground-true values indicate that most exit flows via node 4 are coming from the source node 6. However, the two basic models fail to capture such relations and result in underestimation of the O-D flows (9→1) but overestimation of the O-D flows (6→4). Since the signal plan at this study site is designed to coordinate the through traffic along the arterial (e.g, the link from intersection 3 to 1), the high O-D flows (6→4) and (9→1) have caused a long right-turn queue but relevant short through queue. Hence, taking real-time queue information at such links as the additional measurements, the enhanced model can produce much more reliable estimations.

Note that the queue detectors, which are more expensive than the loop detectors, are not commonly used at signalized intersections. Hence, how to reduce the number of queue detectors and select critical links for installation have emerged as vital issues. The preliminary analysis reveals that those links with high downstream turning outflows and upstream turning inflows shall have the priority in installing queue detectors.

4.6 Closure

This chapter has presented three models for estimating the dynamic O-D flows on a signalized arterial. Model-I features the use of link accounts as the main measurements, whereas Model-II directly takes intersection turning flows as its primary input. Recognizing the impacts of O-D patterns and signal plans on the intersection queue distribution, Model-III further incorporates the real-time queue patterns as additional measurements to improve the estimation accuracy. All three

models can be solved with the sequential algorithm based on the extended Kalman filter.

Based on the actual arterial flows and O-D information, this study has further conducted extensive numerical experiments to assess the performance of these proposed models, especially with respect to their capability in identifying critical arterial path flows, based on their respective volumes. The comparison results convincingly evidence that Model-III that incorporates time-varying intersection queue information can yield the best estimation accuracy, and can also correctly identify and rank all critical path-flows based on their respective volumes. Such promising results offer the valuable information for traffic engineers to design a multi-path signal progression system, rather than the conventional through flow-based progression.

Although increasing the number of queue detectors on arterial links would increase the number of measurements and may consequently help improve the estimation accuracy, it will also require much higher installation and maintained costs. Hence, further extensions along this study will be focused on advancing the proposed models to select the most critical links that need queue detectors, whereas the estimation accuracy could still be guaranteed.

Chapter 5 : A Pre-timed Corridor Signal Control Model

5.1 Introduction

Based on the detected intersection turning flows and identified critical path-flows for the O-D estimation model, this chapter introduces two models to optimize the signal plans at the off-ramp connected arterial. Taking intersection turning flow and the storage space of the target off-ramp as the input, the first model aims to maximize intersection capacities and prevent the formation of off-ramp queue spillover, by optimizing the signal cycle length and green split at each intersection. Then, based on the cycle length, green splits, and identified critical paths, the second model offers an innovative multi-path progression solution to facilitate the progression need of both the off-ramp path-flow and local arterial traffic flows.

5.2 Signal Timing Optimization

Objective Function

Note that a well-designed signal plan needs to be able to maximize the capacity of an intersection under the given geometric layout (Allsop, 1972, Wong and Wong, 2003). Based on the assumption that the traffic demand matrix can be multiplied with a common flow multiplier μ to represent the maximum amount of the increased volume that would still allow the intersection to perform reasonably well (Silcock, 1997), the signal optimization problem can be converted to the issue of determining the maximal multiplier μ:

$$Maxmize \sum_i \mu_i \tag{5.1}$$

where, μ_i is the multiplier applied to the demand pattern at intersection i.

Constraints

Given the phasing plan and traffic demand pattern at each intersection, the following constraints should be satisfied to ensure that the degree of saturation in each lane group is below the acceptable limit.

$$\mu_i \alpha_{k,i} q_{k,i} \le s_{k,i} (\sum_m \beta_{k,m,i} \Phi_{m,i} - \delta \times \xi) \quad \forall i,k \tag{5.2}$$

where $\alpha_{k,i}$ denotes the lane use factor at lane group k of intersection i, which is a function of number of lanes (eg., 0.55 for two lanes); $q_{k,i}$ is the traffic volume at lane group k of intersection i and $s_{k,i}$ is the corresponding saturation flow rate (unit: veh/hour/lane); ξ is the reciprocal of common cycle length; $\Phi_{m,i}$ is the duration of phase m at intersection i; δ denotes the duration of lost time due to the transition between consecutive signal phases; and $\beta_{k,m,i}$ is a parameter representing the phase designs, and is defined as follows:

$$\beta_{k,m,i} = \begin{cases} 1; & \text{if lane group } k \text{ can recieve green in phase } m \text{ at intersection } i; \\ 0; & o.w. \end{cases} \tag{5.3}$$

The common cycle lengths and phase durations shall be subjected to the following constraints:

$$\frac{1}{C_{\max}} \le \xi \le \frac{1}{C_{\min}} \tag{5.4}$$

$$\xi \times g_{\min} \le \Phi_{m,i} \le \xi \times g_{\max} \qquad \forall m,i \tag{5.5}$$

where $C_{\min}$ and $C_{\max}$ are the minimum and maximum cycle lengths; and $g_{\min}$ and $g_{\max}$ are the minimum and maximum phase durations. Also, at each intersection i, the sum of all phase durations should equal to the cycle length:

$$\sum_{m} \Phi_{m,i} = 1 \quad \forall i \tag{5.6}$$

When the freeway mainline flow exceeds the critical threshold, queue spillover at off-ramps can result in a significant capacity reduction. Hence, under such conditions, one shall control the traffic queue at the off-ramp to be below its storage capacity. Figure 5.1 shows a queuing process with the assumption that the arriving flows follow a uniform distribution.

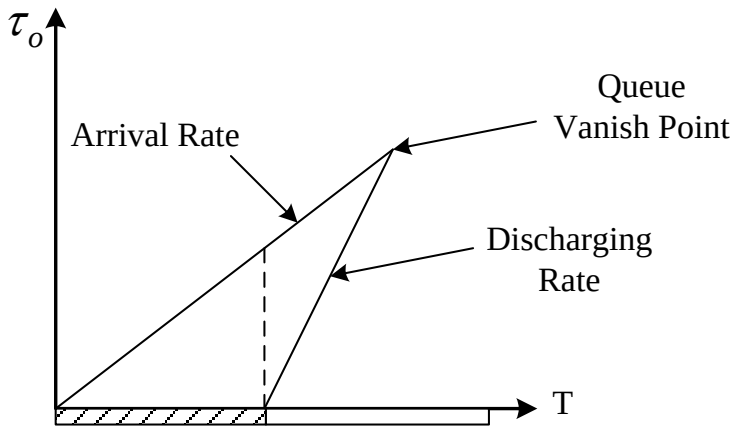


Figure 5.1 the deterministic queuing process at the off-ramp

Based on signal timings and coming flows, one can estimate the queue length at the off-ramp as follows:

$$\tau_{o,i} = \frac{(1-\sum_{m}\beta_{o,m,i}\Phi_{m,i} + \delta \times \xi) \cdot q_{o,i} \cdot s_{o,i}}{(s_{o,i} - q_{o,i})\xi} \tag{5.7}$$

Then, the queue length constraint to prevent off-ramp queue spillback could be established as follows:

$$(1-\sum_{m}\beta_{o,m,i}\Phi_{m,i} + \delta \times \xi) \cdot q_{o,i} \cdot s_{o,i} \le \tau_{o,i}^{\max}(s_{o,i} - q_{o,i})\xi \tag{5.8}$$

where $\tau_{o,i}^{\max}$ is the maximal allowable queue length at the off-ramp. Note that Eq. (5.7) may underestimate the queue length due to the stochastic nature of arriving flows in practice. Hence, to ensure that the corresponding constraint (5.8) can function efficiently in preventing the formation of queue spillbacks, the parameter $\tau_{o,i}^{\max}$ shall be slightly smaller than the off-ramp storage capacity to reserve some space to overcome the uncertainty of arriving flow patterns.

In brief, the signal timing optimization model could be summarized as follows:

$$Maxmize \ \sum_{i}\mu_i$$

s.t.

$$\mu_i \alpha_{k,i} q_{k,i} \le s_{k,i}\sum_{m}\beta_{k,m,i}\Phi_{m,i} - \delta \times \xi \quad \forall i,k$$

$$\sum_{m}\Phi_{m,i} = 1 \quad \forall i$$

$$(1-\sum_{m}\beta_{o,m,i}\Phi_{m,i} + \delta \times \xi) \cdot q_{o,i} \cdot s_{o,i} \le \tau_{o,i}^{\max}(s_{o,i} - q_{o,i})\xi$$

$$\frac{1}{C_{\max}} \le \xi \le \frac{1}{C_{\min}}$$

$$\xi \times g_{\min} \le \Phi_{m,i} \le \xi \times g_{\max} \qquad \forall m,i$$

Note that the above optimization model is formulated with the linear-programming formulations, which can be easily solved with existing algorithms.

5.3 Multi-path progression model

5.3.1 Critical issues in a multi-path progression model

Compared with these methods mainly for the two-way progression, any models designed to offer the progression for multi-path flows such as those comprising both through and turning movements will certainly need a more complex phase diagram to yield progression bands for different path-flows. As shown in Figure 5.2, aside from the through path (Path 1), turning flows along Path 2 will join the arterial at intersection-3 via a west-eastbound left-turn phase. Similarly, vehicles along path 3 will leave the arterial at intersection-2 via a south-northbound left-turn phase. Hence, one needs to use a multiple phase plan to accommodate the progression need for those flows traveling on different paths.

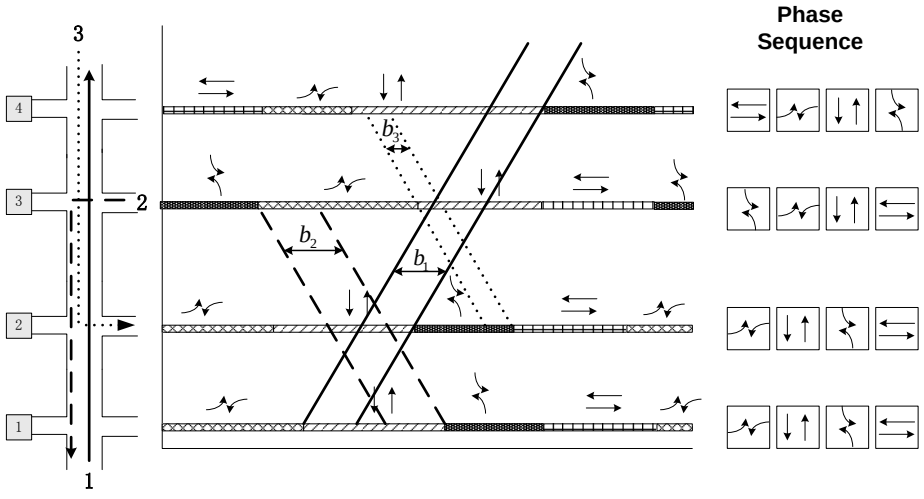


Figure 5.2 An illustrative example for the multi-path progression (A)

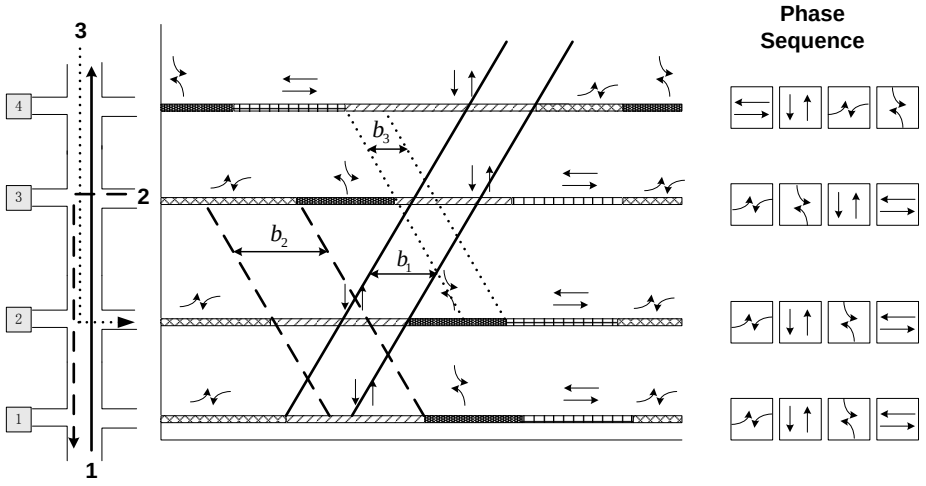


Figure 5.3 An illustrative example for the multi-path progression (B)

Given the need for using a multiple phase plan, the optimization of its phase sequence at each intersection will thus become a vital issue in maximizing the progression efficiency. For example, Figure 5.3 shows a multi-path progression plan with an optimized phase sequence. Compared with the case in Figure 5.2, it is noticeable that after changing the phase sequence at intersections 3 and 4, the green bandwidths along these three paths (b_1, b_2, and b_3) increase significantly.

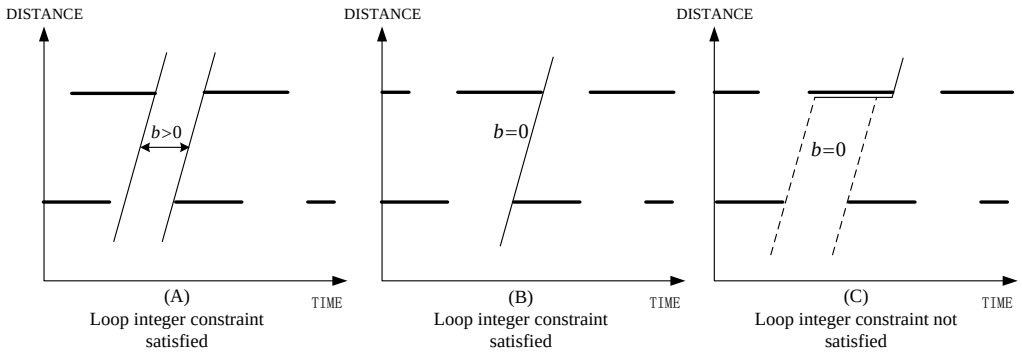


Figure 5.4 Three possible cases for one particular progress path

In addition to the phase sequence optimization, the loop integer constraint in the MAXBAND which ensures that both through directions have non-zero bands (see Figure 5.4(A)) also needs to be extended. This is due to the fact that such constraints in a multi-path progression system may result in near-zero or non-productive bandwidths along some paths (see Figure 5.4(B)). Furthermore, due to the competing nature of green bands, some path flows may encounter the case shown in Figure 5.4(C) and their loop integer constraints will not be satisfied. Under such conditions, the formulations for the entire multi-path progression may yield an infeasible solution.

In brief, an effective arterial signal model, grounded on the logic of MAXBAND, for multi-path progression needs to explicitly account for the following critical issues: 1) use a proper multi-phase plan for the path-based progression; 2) concurrently optimize the signal phase sequence and offsets; and 3) effectively eliminate some infeasible paths flows so as to maximize the sum of weighted bandwidths for all paths, and yield the maximal benefit for the total flows in the target arterial.

5.3.2 Model Formulation

This section presents two sets of formulations, where the first is a direct-extension of MAXBAND for the multi-path scenario, and the second highlights the feature of concurrently optimizing phase sequences and offsets. An enhancement of the second model, designed to eliminate the likelihood of generating some non-productive progression bands, will also be discussed. Note that the non-productive

band is defined in this study as the bandwidth that is not sufficient to accommodate a prespecified number of vehicles (e.g., 3 vehicles).

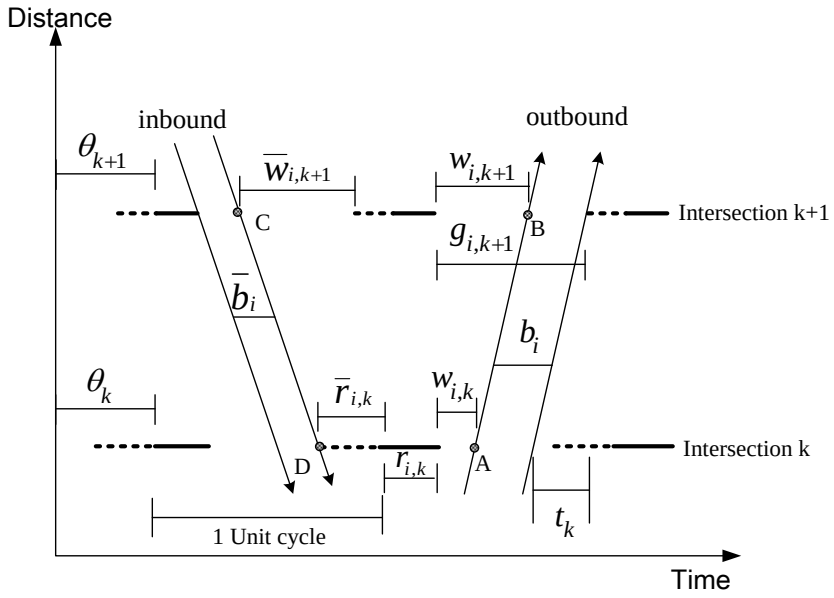


Figure 5.5 Key notations in the proposed models

Model-I: optimize offsets for all target path flows

As discussed previously, with the identified multi-path traffic flow patterns, phase plans, and sequences, one can directly extend MAXBAND with the following objective function for offset optimization:

$$Max \sum_i (\varphi_i b_i) + \sum_i (\overline{\varphi}_i \overline{b}_i) \tag{5.9}$$

where, $\varphi_i(\overline{\varphi}_i)$ and $b_i(\overline{b}_i)$ denote the weighting factor and green bandwidth for inbound (outbound) path *i*, respectively.

Similar to MAXBAND, the interference constraints for a multi-path progression system are given as follows:

$$0 \le w_{i,k} + b_i \le g_{i,k} \quad \forall i \in \Omega; \forall k \in \sigma_i \tag{5.10}$$

$$0 \le \overline{w}_{i,k} + \overline{b}_i \le \overline{g}_{i,k} \quad \forall i \in \overline{\Omega}; \forall k \in \sigma_i \tag{5.11}$$

where, $g_{i,k}(\overline{g}_{i,k})$ is the maximum green duration that the inbound (outbound) path *i* can be obtained at intersection *k*; $w_{i,k}(\overline{w}_{i,k})$ is the part of a green duration before (after) the green band for inbound (outbound) path *i* at intersection *k*; $\Omega(\overline{\Omega})$ is the set of outbound (inbound) paths; σ_i is the set of intersections included in path i.

Note that the formulations for MAXBAND are based on a two-phase signal plan for two-way progression. Hence, the loop integer constraint (Eq. 2.7) can be obtained by substituting those two progression constraints for inbound and outbound though paths. However, due to the complexity of the multi-path progression nature, one shall explore new constraints to represent the progress of green bands for all path flows, rather than directly using the same loop integer constraints in MAXBAND. In this study, a set of alternative progression constraints are derived to represent the progress of an outbound green band from point A to point B (see Figure 5.4):

$$\theta_k + r_{i,k} + w_{i,k} + t_k + n_{i,k} = \theta_{k+1} + r_{i,k+1} + w_{i,k+1} + \tau_{i,k+1} + n_{i,k+1} \quad \forall i \in \Omega; \forall k \in \sigma_i \tag{5.12}$$

where, θ_k is the offset of intersection k; $r_{i,k}$ is the total red duration at the left side of the green band for path i; t_k is the travel time between intersection k and k+1; $\tau_{i,k}$ is the initial queue clearance time along path i at intersection k; $n_{i,k}$ is an integer variable to represent the number of signal cycles.

Similarly, for traffic along the inbound paths, one can use the following equation to represent the progress of the green band from point C to point D as shown in Figure 5.4:

$$-\theta_k + \bar{r}_{i,k} + \overline{w}_{i,k} - \bar{\tau}_{i,k} + \bar{t}_k + \bar{n}_{i,k} = -\theta_{k+1} + \bar{r}_{i,k+1} + \overline{w}_{i,k+1} + \bar{n}_{i,k+1} \quad \forall i \in \overline{\Omega}; \forall k \in \sigma_i \tag{5.13}$$

where, $\bar{r}_{i,k}$ is the total red duration at the right side of the green band for path i; $\bar{n}_{i,k}$ is an integer variable.

Given the green durations for all intersections, one can then compute the values of $r_{i,k}(\bar{r}_{i,k})$ and $g_{i,k}(\bar{g}_{i,k})$ with the pre-determined phase sequence. Hence, the Model-I (M1) could be summarized as follows:

$$M1: Max \sum_i (\varphi_i b_i) + \sum_i (\bar{\varphi}_i \bar{b}_i)$$

s.t.

$$0 \le w_{i,k} + b_i \le g_{i,k} \quad \forall i \in \Omega; \forall k \in \sigma_i$$

$$0 \le \overline{w}_{i,k} + \bar{b}_i \le \bar{g}_{i,k} \quad \forall i \in \overline{\Omega}; \forall k \in \sigma_i$$

$$\theta_k + r_{i,k} + w_{i,k} + t_k + n_{i,k} = \theta_{k+1} + r_{i,k+1} + w_{i,k+1} + \tau_{i,k+1} + n_{i,k+1} \quad \forall i \in \Omega; \forall k \in \sigma_i$$

$$-\theta_k + \bar{r}_{i,k} + \bar{w}_{i,k} - \bar{\tau}_{i,k} + \bar{t}_k + \bar{n}_{i,k} = -\theta_{k+1} + \bar{r}_{i,k+1} + \bar{w}_{i,k+1} + \bar{n}_{i,k+1} \quad \forall i \in \bar{\Omega}; \forall k \in \sigma_i$$

$$b_i, w_{i,k}, \bar{b}_i, \bar{w}_{i,k} > 0 \qquad \forall i \in \Omega + \bar{\Omega}; \forall k \in \sigma_i$$

$\bar{n}_{i,k}, \bar{n}_{i,k}$ are integer variables

Note that the M1 model assumes that the travel speeds on all links are predetermined. To concurrently optimize the progression speeds, one can simply add additional constraints shown in Eqs. (2.8)-(2.11).

Model-II: Concurrently optimizes phase sequences and offsets

As shown in Figure 5.3 and Figure 5.4, changing the phase sequence at each intersection can minimize the green band competition between different path flows. To optimize such sequences, one can first define a set of binary variables as follows:

$$x_{l,m,k} = \begin{cases} 1, & \text{if phase } l \text{ is before phase } m \text{ within the same cycle of intersection } k; \\ 0, & \text{o.w.} \end{cases}$$

(5.14)

To ensure the operational feasibility of each produced phase sequence, this study has further specified the following constraints related to $x_{l,m,k}$:

$$x_{l,l,k} = 0 \quad \forall l; \forall k \tag{5.15}$$

$$x_{l,m,k} + x_{m,l,k} = 1 \quad \forall l \neq m; \forall k \tag{5.16}$$

$$x_{l,n,k} >= x_{l,m,k} + x_{m,n,k} - 1 \quad \forall l \neq m \neq n; \forall k \tag{5.17}$$

$$x_{l,m,k} + x_{m,l,k} = 1 \quad l \neq m \tag{5.18}$$

$$x_{l,n,k} + x_{n,m,k} = 1 \quad l \neq m \neq n \tag{5.19}$$

Note that Eq. (5.15) is based on the definition of $x_{l,m,k}$ and Eq.(5.16) indicates that phase l is either before or after phase m. To prevent a "sub-loop" in a phase sequence, the model must ensure that "if phase l is before phase m ($x_{l,m,k}$ =1), and phase m is prior to phase n ($x_{m,n,k}$ =1), then phase l is before phase n ($x_{l,n,k}$ =1)". The mathematical formulations are shown in Eq. (5.17) for such relations. Eqs. (5.18)-(5.19) are two optional constraints, where Eq. (5.18) ensures that phase l and m in a sequential order and Eq. (5.19) indicates that phase l is ahead of phase m.

By denoting $\phi_{l,k}$ as the duration of phase l at intersection k, one can further define a set of binary parameters as follows:

$$\beta_{i,l,k} = \begin{cases} 1, & \text{if path } i \text{ obtains the green duration in phase } l \text{ at intersection } k; \\ 0, & \text{o.w.} \end{cases} \tag{5.20}$$

Hence, one can compute the available green duration for the progress of an outbound (inbound) path i, $g_{i,k}$ ($\bar{g}_{i,k}$) with the following equations:

$$g_{i,k} = \beta_{i,l,k}\phi_{l,k} \quad \forall i \in \Omega; \forall k \in \sigma_i \tag{5.21}$$

$$\bar{g}_{i,k} = \beta_{i,l,k}\phi_{l,k} \quad \forall i \in \Omega; \forall k \in \sigma_i \tag{5.22}$$

Then, the interference constraints can be re-written as follows:

$$0 \le w_{i,k} + b_i \le \sum_l \beta_{i,l,k}\phi_{l,k} \quad \forall i \in \Omega; \forall k \in \sigma_i \tag{5.23}$$

$$0 \le \bar{w}_{i,k} + \bar{b}_i \le \sum_l \beta_{i,l,k}\phi_{l,k} \quad \forall i \in \bar{\Omega}; \forall k \in \sigma_i \tag{5.24}$$

Note that in the M1 model, the red duration of path *i* before (after) its available green time, $r_{i,k}(\bar{r}_{i,k})$, can be directly computed since the phase sequences are provided. However, by relaxing the phase sequences as decision variables, the values of $r_{i,k}(\bar{r}_{i,k})$ will be varied with the selected phase sequence. Hence, to ensure the progression constraints to function properly, one shall establish the relations between phase sequences, $x_{l,m}$, and red durations $r_{i,k}(\bar{r}_{i,k})$. A set of constraints for such a need are given below:

$$r_{i,k} \le \sum_l \beta_{i,m,k} x_{l,m} \cdot \phi_{l,k} + M(1-\beta_{i,m,k}) \quad \forall i \in \Omega + \bar{\Omega}; \forall k \in \sigma_i; \forall m \tag{5.25}$$

$$\bar{r}_{i,k} \le \sum_l \beta_{i,m,k} x_{m,l} \cdot \phi_{l,k} + M(1-\beta_{i,m,k}) \quad \forall i \in \Omega + \bar{\Omega}; \forall k \in \sigma_i; \forall m \tag{5.26}$$

$$r_{i,k} + \bar{r}_{i,k} + \sum_l \beta_{i,l,k} \cdot \phi_{k,n} = 1 \quad \forall i \in \Omega + \bar{\Omega}; \forall k \in \sigma_i \tag{5.27}$$

where, *M* is a large positive number. The second term at the right-hand side of Eq. (5.25) is used to dominate other variables when path *i* cannot receive the green duration in phase *m* ($\beta_{i,m,k}$ = 0). If path *i* can receive green time in phase *m* ($\beta_{i,m,k}$ = 1), the first term in Eq. (5.25) will function to compute the time duration before phase *m*. Following the similar logic, the first term in the right-hand side of Eq. (5.26) is used to compute the duration after phase *m* if $\beta_{i,m,k}$ = 1. Since path *i* may receive green in

multiple consecutive phases, Eqs. (5.25)-(5.26) will ensure that $r_{i,k}(\bar{r}_{i,k})$ is taking place prior to (after) the first (last) phase that is given green to path *i*. Also note that the third term in Eq. (31) is the total green duration that path *i* can obtain within one cycle. Hence, Eqs. (5.25)-(5.27) together can force $r_{i,k}(\bar{r}_{i,k})$ to equal the red duration before (after) the green time of path *i*, which corresponds to its definition.

In brief, the M2 model could be summarized as follows:

$$M2: Max \sum_i (\varphi_i b_i) + \sum_i (\bar{\varphi}_i \bar{b}_i)$$

s.t.

$$\theta_k + r_{i,k} + w_{i,k} + t_k + n_{i,k} = \theta_{k+1} + r_{i,k+1} + w_{i,k+1} + \tau_{i,k+1} + n_{i,k+1} \quad \forall i \in \Omega; \forall k \in \sigma_i$$

$$-\theta_k + \bar{r}_{i,k} + \bar{w}_{i,k} - \bar{\tau}_{i,k} + \bar{t}_k + \bar{n}_{i,k} = -\theta_{k+1} + \bar{r}_{i,k+1} + \bar{w}_{i,k+1} + \bar{n}_{i,k+1} \quad \forall i \in \bar{\Omega}; \forall k \in \sigma_i$$

$$x_{l,l,k} = 0 \quad \forall l; \forall k$$

$$x_{l,m,k} + x_{m,l,k} = 1 \quad \forall l \neq m; \forall k$$

$$x_{l,n,k} >= x_{l,m,k} + x_{m,n,k} - 1 \quad \forall l \neq m \neq n; \forall k$$

$$x_{l,m,k} + x_{m,l,k} = 1 \quad l \neq m$$

$$x_{l,n,k} + x_{n,m,k} = 1 \quad l \neq m \neq n$$

$$0 \leq w_{i,k} + b_i \leq \sum_l \beta_{i,l,k} \phi_{l,k} \quad \forall i \in \Omega; \forall k \in \sigma_i$$

$$0 \leq \bar{w}_{i,k} + \bar{b}_i \leq \sum_l \beta_{i,l,k} \phi_{l,k} \quad \forall i \in \bar{\Omega}; \forall k \in \sigma_i$$

$$r_{i,k} \leq \sum_l \beta_{i,m,k} x_{l,m} \cdot \phi_{l,k} + M(1 - \beta_{i,m,k}) \quad \forall i \in \Omega + \bar{\Omega}; \forall k \in \sigma_i; \forall m$$

$$\bar{r}_{i,k} \le \sum_{l} \beta_{i,m,k} x_{m,l} \cdot \phi_{l,k} + M(1-\beta_{i,m,k}) \quad \forall i \in \Omega + \bar{\Omega}; \forall k \in \sigma_i; \forall m$$

$$r_{i,k} + \bar{r}_{i,k} + \sum_{l} \beta_{i,l,k} \cdot \phi_{k,n} = 1 \quad \forall i \in \Omega + \bar{\Omega}; \forall k \in \sigma_i$$

$$b_i, w_{i,k}, \bar{b}_i, \bar{w}_{i,k} > 0 \quad \forall i \in \Omega + \bar{\Omega}; \forall k \in \sigma_i$$

$\bar{n}_{i,k}, \bar{n}_{i,k}, x_{i,j,k}$ are integer variables

Model-III: Select part of paths for progression and prevent infeasible solutions

Note that both the M1 and M2 models assume that every selected critical path can receive a green band on the arterial, regardless of their resulting bandwidths. In other words, the progression constraints (5.12)-(5.13) will ensure that every selected path can obtain a bandwidth under the conditions in either Figure 5.3(A) or (B). Obviously, providing a near-zero green band is practically non-productive, and removing the constraints of those paths may offer a larger bandwidth to other primary paths. Also, with an increase in the number of paths and intersections, some paths may inevitably encounter the case in Figure 5.3(C), due to their competing for progression. Consequently, both the M1 and M2 models will yield no feasible solutions under such condition.

To tackle such issues, one shall further define a set of new decision variables to automatically select proper paths for green-band maximization:

$$y_i(\bar{y}_i) = \begin{cases} 1, & \text{if path } i \text{ obtains signal progression with non-zero green band} \\ 0, & \text{o.w.} \end{cases}$$

(5.28)

Then, the following two additional constraints are derived to remove those paths without progression from the optimization process:

$$b_i \le y_i \tag{5.29}$$

$$\bar{b}_i \le \bar{y}_i \tag{5.30}$$

Note that Eqs. (5.29)-(5.30) are used to force the bandwidth of path *i* to zero if it is removed from progression need ($y_i = 0$).

For those outbound paths without progression, one needs to relax their corresponding progression constraints to ensure the feasibility of the optimization model. To do so, one can rewrite the set of progression constraints in Eq. (5.12) as follows:

$$\theta_k + r_{i,k} + w_{i,k} + t_k + n_{i,k} \ge \theta_{k+1} + r_{i,k+1} + w_{i,k+1} + \tau_{i,k+1} + n_{i,k+1} - M(1-y_i) \quad \forall i \in \Omega; \forall k \in \sigma_i \tag{5.31}$$

$$\theta_k + r_{i,k} + w_{i,k} + t_k + n_{i,k} \le \theta_{k+1} + r_{i,k+1} + w_{i,k+1} + \tau_{i,k+1} + n_{i,k+1} + M(1-y_i) \quad \forall i \in \Omega; \forall k \in \sigma_i \tag{5.32}$$

where, M is a large positive number.

Hence, when y_i equals to zero, Eq. (5.31)-(5.32) will become as follows:

$$\theta_k + r_{i,k} + w_{i,k} + t_k + n_{i,k} \ge \theta_{k+1} + r_{i,k+1} + w_{i,k+1} + \tau_{i,k+1} + n_{i,k+1} - M \quad \forall i \in \Omega; \forall k \in \sigma_i \tag{5.33}$$

$$\theta_k + r_{i,k} + w_{i,k} + t_k + n_{i,k} \le \theta_{k+1} + r_{i,k+1} + w_{i,k+1} + \tau_{i,k+1} + n_{i,k+1} + M \quad \forall i \in \Omega; \forall k \in \sigma_i \tag{5.34}$$

Since M can dominate any variable in the constraints, Eq. (5.31)-(5.32) will always be held for those decision variables. In other words, constraints (5.31)-(5.32) are ineffective when y_i equals zero.

Grounded on the same logic, one can derive the following constraints for the inbound paths by modifying Eq. (5.13) as follows:

$$-\theta_k + \bar{r}_{i,k} + \bar{w}_{i,k} - \bar{\tau}_{i,k} + \bar{t}_k + \bar{n}_{i,k} \geq -\theta_{k+1} + \bar{r}_{i,k+1} + \bar{w}_{i,k+1} + \bar{n}_{i,k+1} - M(1-\bar{y}_i) \quad \forall i \in \bar{\Omega}; \forall k \in \sigma_i \tag{5.35}$$

$$-\theta_k + \bar{r}_{i,k} + \bar{w}_{i,k} - \bar{\tau}_{i,k} + \bar{t}_k + \bar{n}_{i,k} \leq -\theta_{k+1} + \bar{r}_{i,k+1} + \bar{w}_{i,k+1} + \bar{n}_{i,k+1} + M(1-\bar{y}_i) \quad \forall i \in \bar{\Omega}; \forall k \in \sigma_i \tag{5.36}$$

Note that under some extreme scenarios, the enhanced model may sacrifice all paths from one direction (e.g. inbound) and offers only a one-direction progression if they are specified with small weight factors. Hence, similar to the constraint (2.4) in MAXBAND, the following constraint should be satisfied:

$$(1-k)\sum_{i\in\bar{\Omega}} \bar{b}_i \geq (1-k)k\sum_{i\in\Omega} b_i \tag{5.37}$$

In brief, one can summarize the M3 model as follows:

$$M3: Max \sum_i (\varphi_i b_i) + \sum_i (\bar{\varphi}_i \bar{b}_i)$$

s.t.

$$x_{l,l,k} = 0 \quad \forall l; \forall k$$

$$x_{l,m,k} + x_{m,l,k} = 1 \quad \forall l \neq m; \forall k$$

$$x_{l,n,k} >= x_{l,m,k} + x_{m,n,k} - 1 \quad \forall l \neq m \neq n; \forall k$$

$$x_{l,m,k} + x_{m,l,k} = 1 \quad l \neq m$$

$$x_{l,n,k} + x_{n,m,k} = 1 \quad l \neq m \neq n$$

$$0 \leq w_{i,k} + b_i \leq \sum_l \beta_{i,l,k} \phi_{l,k} \quad \forall i \in \Omega; \forall k \in \sigma_i$$

$$0 \leq \bar{w}_{i,k} + \bar{b}_i \leq \sum_l \beta_{i,l,k} \phi_{l,k} \quad \forall i \in \bar{\Omega}; \forall k \in \sigma_i$$

$$r_{i,k} \leq \sum_l \beta_{i,m,k} x_{l,m} \cdot \phi_{l,k} + M(1 - \beta_{i,m,k}) \quad \forall i \in \Omega + \bar{\Omega}; \forall k \in \sigma_i; \forall m$$

$$\bar{r}_{i,k} \leq \sum_l \beta_{i,m,k} x_{m,l} \cdot \phi_{l,k} + M(1 - \beta_{i,m,k}) \quad \forall i \in \Omega + \bar{\Omega}; \forall k \in \sigma_i; \forall m$$

$$r_{i,k} + \bar{r}_{i,k} + \sum_l \beta_{i,l,k} \cdot \phi_{k,n} = 1 \quad \forall i \in \Omega + \bar{\Omega}; \forall k \in \sigma_i$$

$$b_i \leq y_i$$

$$\bar{b}_i \leq \bar{y}_i$$

$$\theta_k + r_{i,k} + w_{i,k} + t_k + n_{i,k} \geq \theta_{k+1} + r_{i,k+1} + w_{i,k+1} + \tau_{i,k+1} + n_{i,k+1} - M(1 - y_i)$$
$$\forall i \in \Omega; \forall k \in \sigma_i$$

$$\theta_k + r_{i,k} + w_{i,k} + t_k + n_{i,k} \leq \theta_{k+1} + r_{i,k+1} + w_{i,k+1} + \tau_{i,k+1} + n_{i,k+1} + M(1 - y_i)$$
$$\forall i \in \Omega; \forall k \in \sigma_i$$

$$-\theta_k + \bar{r}_{i,k} + \bar{w}_{i,k} - \bar{\tau}_{i,k} + \bar{t}_k + \bar{n}_{i,k} \geq -\theta_{k+1} + \bar{r}_{i,k+1} + \bar{w}_{i,k+1} + \bar{n}_{i,k+1} - M(1 - \bar{y}_i)$$
$$\forall i \in \bar{\Omega}; \forall k \in \sigma_i$$

$$-\theta_k + \bar{r}_{i,k} + \bar{w}_{i,k} - \bar{\tau}_{i,k} + \bar{t}_k + \bar{n}_{i,k} \leq -\theta_{k+1} + \bar{r}_{i,k+1} + \bar{w}_{i,k+1} + \bar{n}_{i,k+1} + M(1 - \bar{y}_i)$$
$$\forall i \in \bar{\Omega}; \forall k \in \sigma_i$$

$$b_i, w_{i,k}, \bar{b}_i, \bar{w}_{i,k} > 0 \quad \forall i \in \Omega + \bar{\Omega}; \forall k \in \sigma_i$$

$\bar{n}_{i,k}, \bar{n}_{i,k}, x_{i,j,k}, y_i, \bar{y}_i$ are integer variables

Note that all above three models, M1, M2 and M3, are formulated with the mixed-integer-linear-programming formulations, which can be solved with existing algorithms such as Branch-and-Band technique due to their limited number of decision variables.

5.4 Numerical Test-I

The experimental system used to evaluate the proposed models consists of an arterial of six intersections. As shown in Figure 5.6, six critical paths have been identified along the arterial, where Paths 1, 2, and 3 are for outbound flows and the remaining paths are for inbound flows.

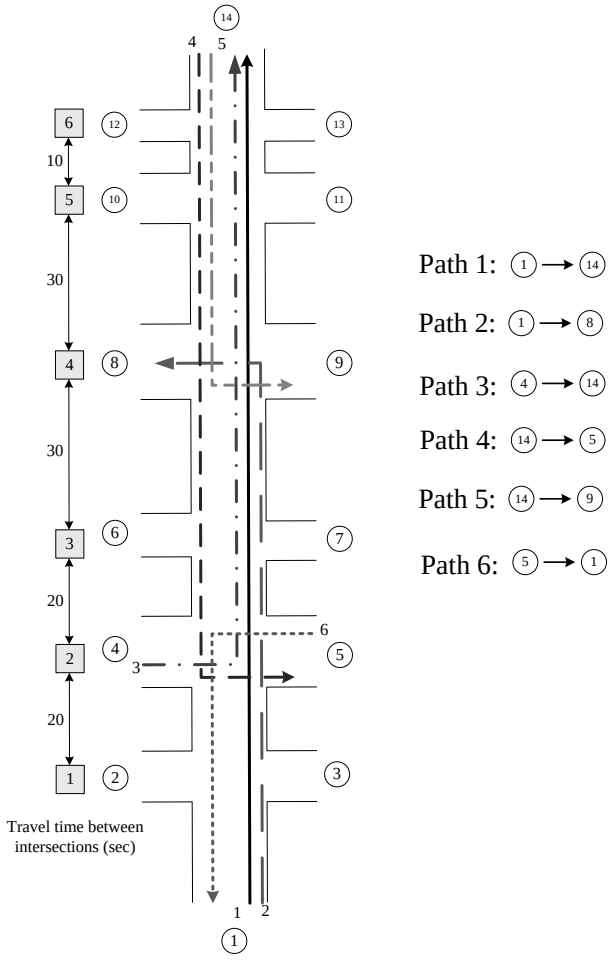


Figure 5.6 Illustration of the arterial for experimental analysis

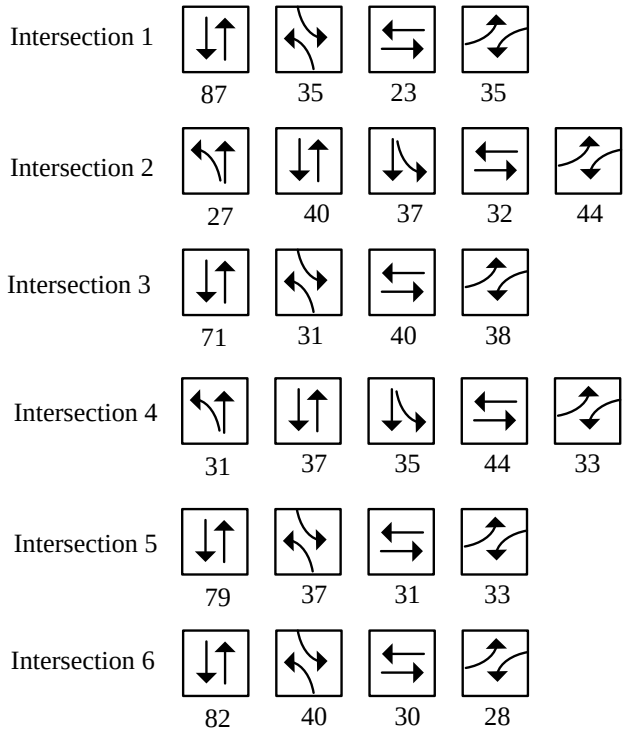


Figure 5.7 Signal timings and the initial phase sequences

The phasing plan, signal timings, and original phase sequence at each intersection are presented in Figure 5.7. The common cycle length is set to be 180 seconds. The weighting factors for the six paths are assumed as 0.2, 0.4, 0.1, 0.5, 0.1, and 0.3, respectively.

Using the signal plans shown in Figure 5.7 and assuming no initial queues on all links, the first model, M1, is implemented to optimize the offset at each intersection. The resulting green bandwidth for each path is presented in Figure 5.8, where paths 5 and 6 have relatively wider green bands than those for paths 1-4. Notably, the bandwidth for path 4 is zero and its green band is actually converged

into a single line. The sum of all weighted bandwidths for this signal plan is 20.1 seconds.

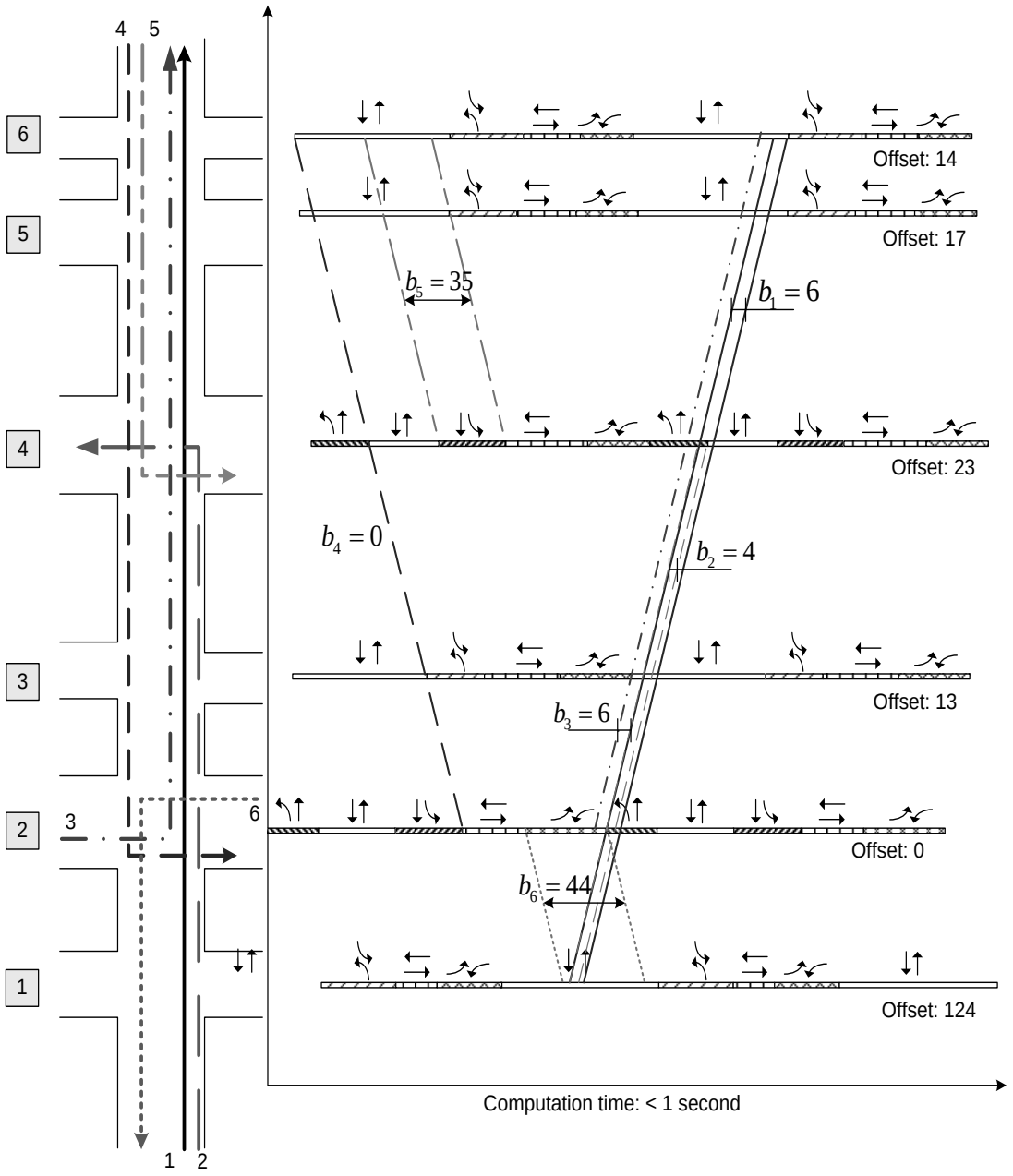


Figure 5.8 The resulting green bands obtained by Model-I

By specifying the phase sequence at each intersection as a decision variable, Figure 5.9 shows the resulting signal plan obtained by the M2 model. Comparing

with the results from M1 model, the bandwidth for paths 1, 3 and 5 are nearly reduced to zeroes and the bandwidth of path 6 remains unchanged. The bandwidths for paths 1 and 4 have been increased. The sum of weighted bandwidths for this signal plan is 28.8 seconds, which is larger than the one produced by the M1 model.

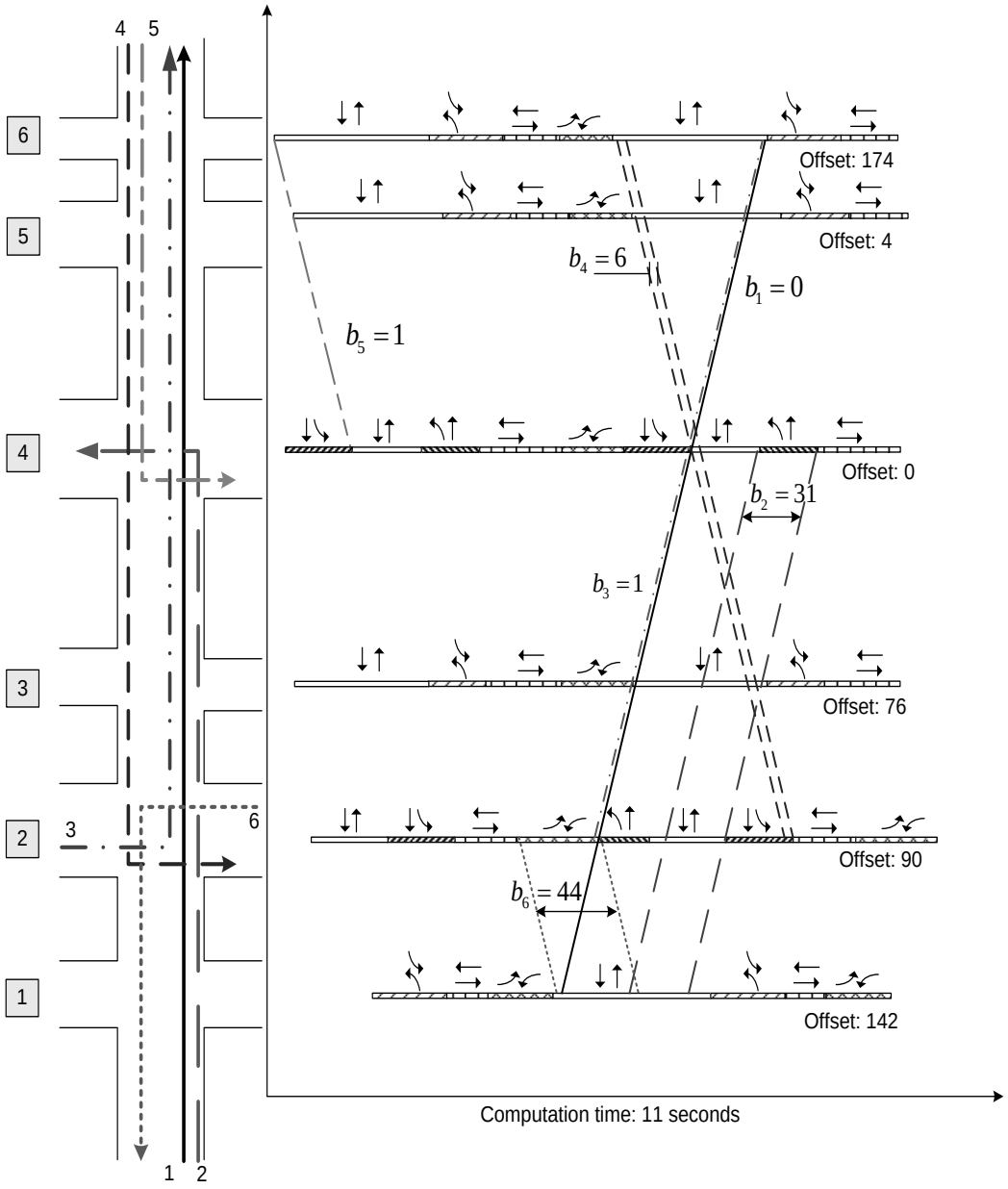


Figure 5.9 The resulting green bands obtained by Model-II

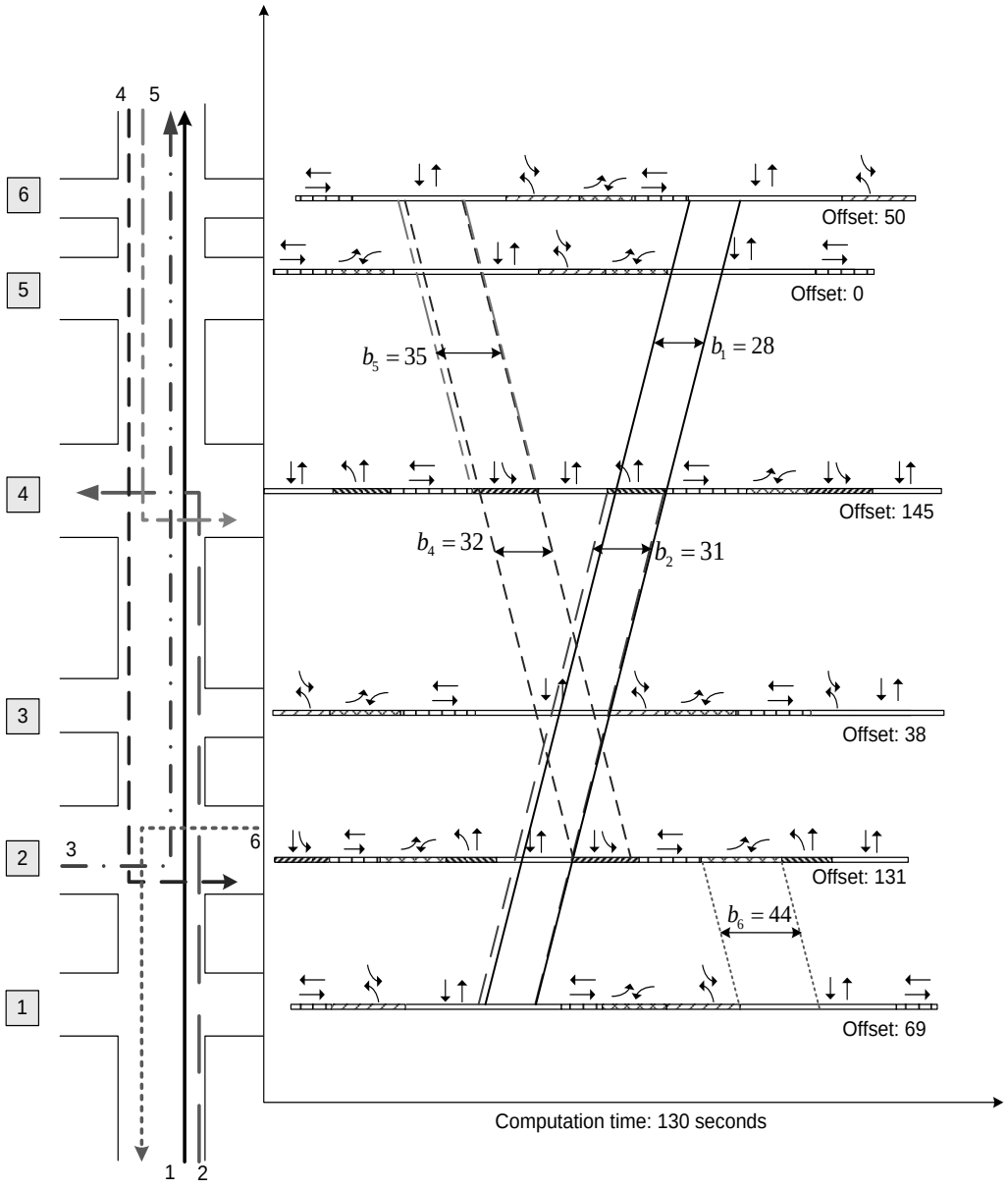


Figure 5.10 The resulting green bands obtained by Model-III

From both Figure 5.8 and Figure 5.9, one can observe that the bandwidths of several paths are close to zeroes, which are not useful in practice. Moreover, incorporating the progression constraints of those paths in the optimization models

will reduce the feasible solution set which will in turn impact the progression results of the remaining paths.

The results from the M3 model, designed to circumvent the limitations of the M1 and M2 models, are shown in Figure 5.10, whose five out of six paths are concurrently accounted for progression and no near-zero green bands are found in this case. In the path-flows competing process, path 3 has been automatically removed from the progression design, and its corresponding constraints have become ineffective during the solution process. The sum of weighted bandwidths for this signal plan is 50.7, which is much larger than the results from the M1 and M2 models. Hence, one can conclude that the M3 model with additional critical constraints can indeed outperform the other two models for arterial experiencing multiple paths of heavy traffic patterns.

5.5 Numerical Test-II

To evaluate the proposed control system, one freeway segment in Chupei, Taiwan along with its nearby intersections is selected as the study site. As shown in Figure 5.11, during PM peak hours, heavy traffic volumes will take the off-ramp via Node 4 to the Guanming 6th Rd W. Since the current pre-timed signal timings are designed to coordinate the through traffic on the arterial, the NB off-ramp flows often form a long queue and spill back to the freeway mainline.

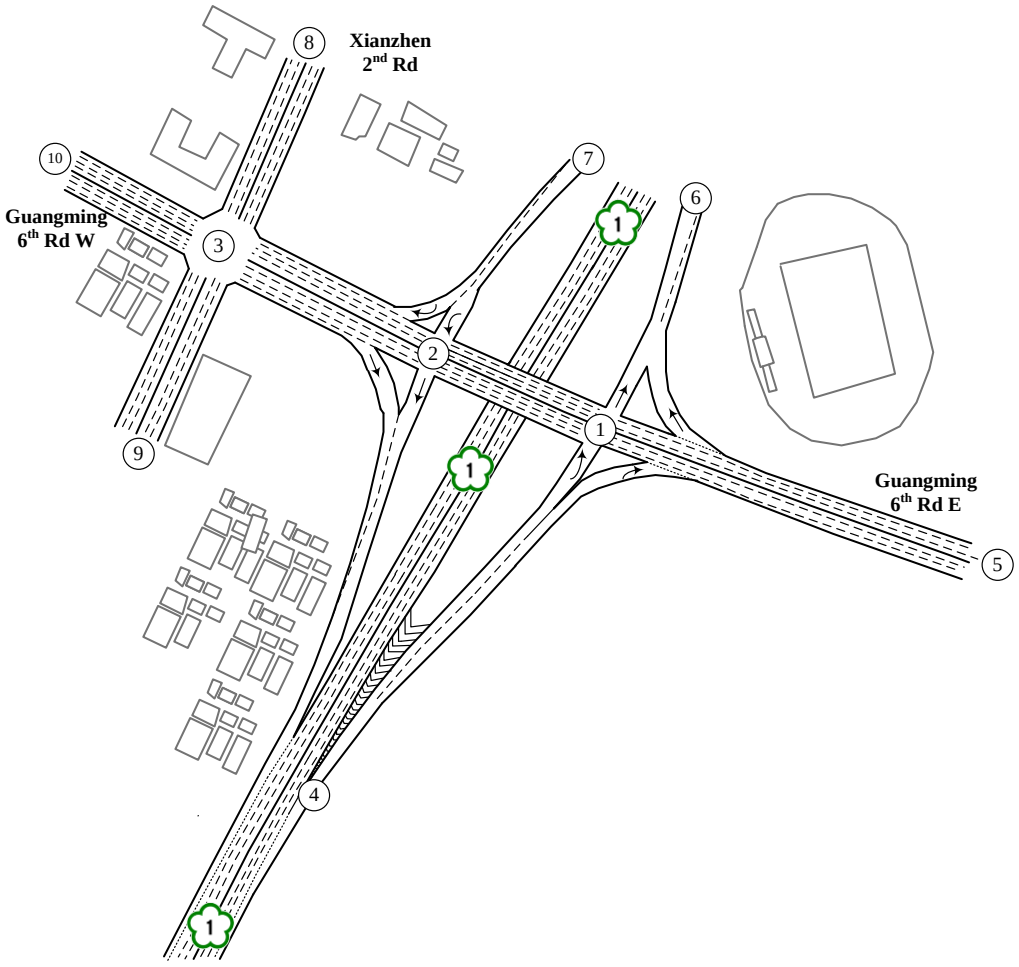


Figure 5.11 The geometric layout of the study site

To analyze the reduced freeway capacity due to the queue spillback, the data collection group in National Chiao Tung University (NCTU), sponsored by Taiwan Department of Transportation, has completed a field survey from 16:30PM to 21:30 PM on April 24-25, 2013. The collected data includes:

1) Freeway northbound flow rate along with its turning ratios at the off-ramp (Node 4);

2) Traffic volume in each lane group at each intersection shown in Table 6.1;

3) Maximum Queue length per cycle at critical arterial links. Using the node number in Figure 8, these critical links are: 1 –> 2, 2 –> 3, 5 –> 1, 8 –> 3, 9 –> 3, 10 –> 3, 3 –> 2 and 2 –> 1.

4) Current signal timings, including cycle length, green splits and offsets.

5) The major path of the off-ramp movement and local traffic flows, as shown in Figure 5.12.

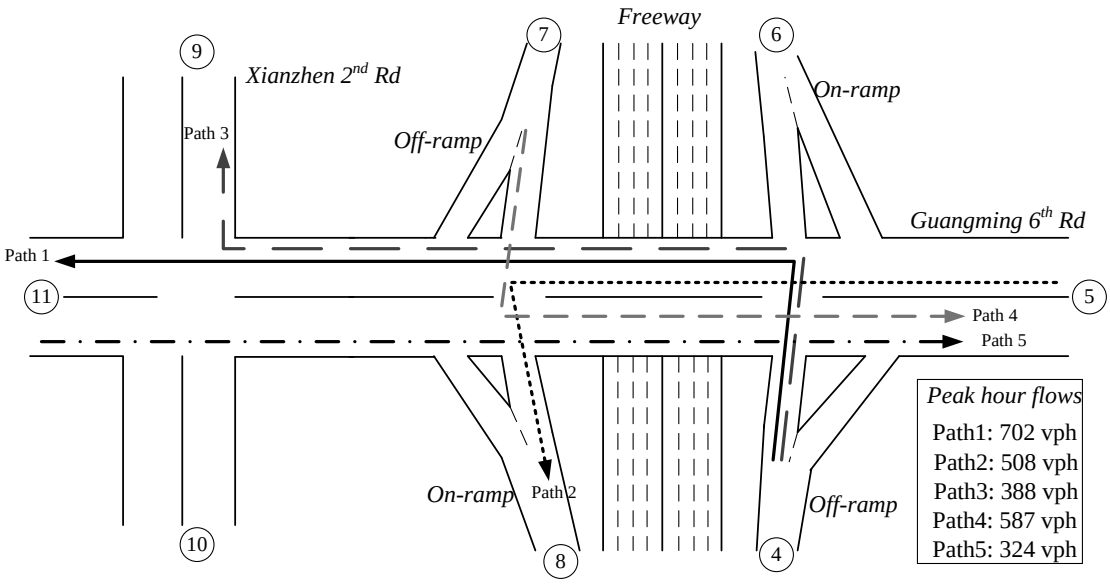


Figure 5.12 Five identified critical paths at the study site

As shown in the previous section, since the M3 model can outperform the other two models, this field study has tested only this model. Based on the volumes along those five paths, the following factors, 0.5, 0.4, 0.3, 0.3, and 0.1, are set for the maximization of progress bands.

Also, to evaluate the effectiveness of the proposed progression model, this study has conducted the comparisons for the following three models:

- Model-1: The TRANSYT 7-F model;
- Model-2: The proposed stage-1 model integrated with a two-way progression model;
- Model-3: The proposed two-stage model.

Table 5.1 The three-hour demand patterns for the three intersections

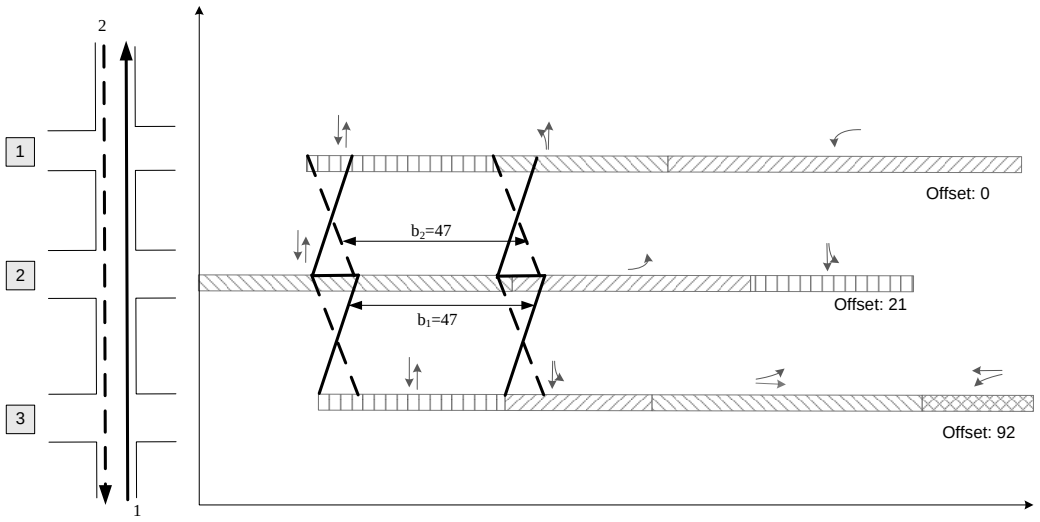
Time	Intersection	Westbound			Northbound			Eastbound			Southbound		
		L	T	R	L	T	R	L	T	R	L	T	R
17:00-	1	/	924	/	1090	/	1073	518	976	/	/	/	/
18:00	2	508	1506	/	/	/	/	/	907	/	587	/	672
	3	332	1056	790	/	274	451	82	834	91	729	272	74
18:00-	1	/	786	/	1050	/	1011	441	968	/	/	/	/
19:00	2	418	1418	/	/	/	/	/	857	/	551	/	729
	3	363	962	822	/	288	371	88	832	86	761	354	64
19:00-	1	/	665	/	995	/	968	443	840	/	/	/	/
20:00	2	269	1391	/	/	/	/	/	921	/	362	/	403
	3	287	872	635	/	238	300	82	653	91	658	286	69

By applying these three models to the target corridor shown in Figure 6.1, the resulting signal plans at the target arterial are presented in Table 5.2.

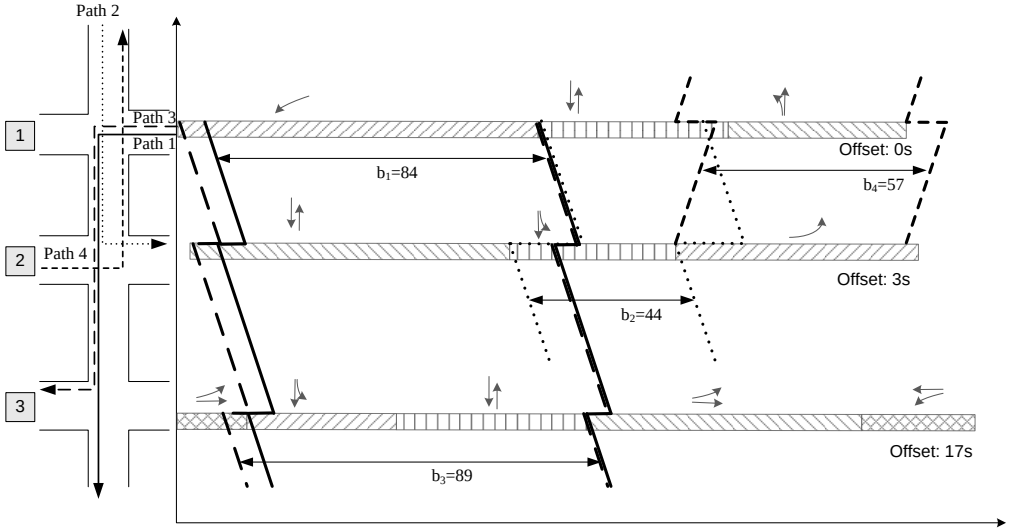
Table 5.2 The optimization results from different models (unit: seconds)

Model	Intersection	CL	Φ1	Φ2	Φ3	Φ4	offset
Model-1	1	160	92	37	31	/	68
	2	160	75	35	50	/	105
	3	160	41	32	60	27	0
Model-2	1	150	89	47	44	/	0
	2	150	41	79	60	/	21
	3	150	47	37	68	28	92
Model-3	1	150	89	47	44	/	0
	2	150	41	79	60	/	3
	3	150	47	37	68	28	17

Based on the signal plans shown in Table 5.2, the resulting green bands produced by Model-2 (MAXBAND) and Model-3 (M3) are shown in Figure 5.13. Notably, the MAXBAND model is a two-way progression model which only offers green bands to the through flows along the arterial (Figure 6.13 (A)). In contrast, the proposed M3 model can concurrently accommodate the progression needs of multiple path-flows. However, to maximize the operational efficiency, path-5 has been eliminated by the model, as shown in Figure 6.13 (B). Also note that the shifts of green bands are caused by the impact of initial queues.



(A) Green bands by MAXBAND for two-way through movements



(B) Green bands by M3 for multiple critical traffic paths

Figure 5.13 The resulting green bands obtained by MAXBAND and M3

VISSIM Calibration

To evaluate the network performance before and after the on-line priority controls, a simulation network is developed with VISSIM 5.20. Recognizing that a

simulation system is useful only if it can faithfully reflect the behavior of its target driving populations, this study has performed the calibration by minimizing the difference between simulated and field-collected queues as well as flow rates. The calibration results for VISSIM simulation are listed is Table 5.3-5.4.

Table 5.3 Percentage difference between simulation and field volume data

Intersection No.	Approach			
	WB	NB	EB	SB
1	1%	0.6%	2%	N/A
2	0.9%	N/A	2%	0.2%
3	2%	3%	0.6%	1%

Table 5.4 Adjusted VISSIM parameters

Parameters	Value
Average stand still distance (Urban)	3.22 ft
Maximum deceleration (Lane Change)	-14.99 ft/s^2
Accepted deceleration (Lane Change)	-6.00 ft/s^2
Maximum deceleration for cooperative braking	-14.99 ft/s^2

Evaluation of the freeway mainline performance

As shown in Figure 5.14, it is clear that both Model-1 and Model-2 fail to prevent the occurrence of queue spillover at the target off-ramp and the travel time along the freeway segment has increased significantly over the congested time period (6:30-7:00). However, with Model-3, the freeway mainline flows can maintain a

higher travel speed during that period and the increase in the freeway travel time is much lower. By comparing Model-2 and Model-3, which share the identical stage-1 model in optimizing signal timings, one can conclude that a simple control method with increased green time of off-ramp flows may not effectively prevent the queue spillover. This is due to that without a specified progression plan for the off-ramp flows, congestion at downstream intersections has propagated the traffic queue to the off-ramp intersection and thus impacts the discharging of flows.

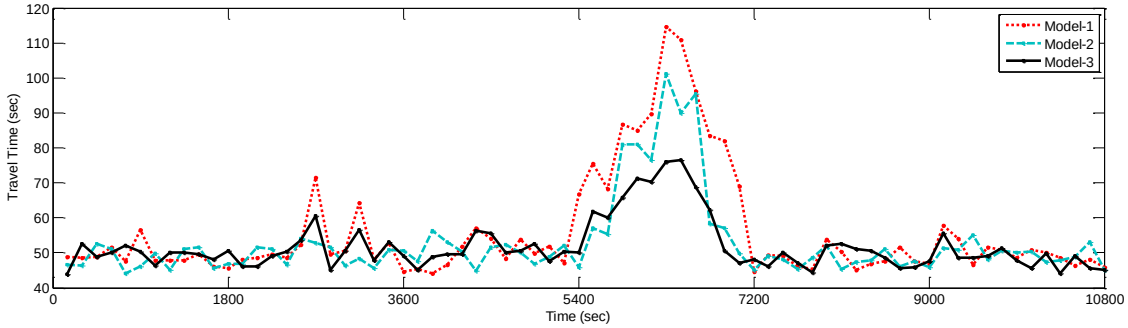


Figure 5.14 Travel time passing the bottleneck

Evaluation of the critical path-flows

Based on the simulation results, Figure 5.15-5.19 presents the time-dependent travel time distribution along each critical path. Compared with the plans from MAXBAND and TRANSYT 7-F, the proposed M3 model can produce much lower travel time along Path 1, as evidenced by the results in Figure 5.15. This is due to the fact that the movement along Path 1 has the largest weighting factor in the M3 model and thus receives the progression priority. Also, the travel times produced by TRANSYT 7-F are slightly higher than those by MAXBAND since Path 1 has shared links with the outbound through path produced with MAXBAND. Similar

observations could be found in Figure 5.17 since Paths 1 and 3 have obtained the green times within the same phases at most intersections.

Figure 6.6 shows the comparison of travel time for Path 2. Obviously, TRANSYT 7-F and M3 can significantly outperform MAXBAND with respect to this path. Based on the trajectory of Path 2, one can observe that vehicles will make left-turns at the second intersection and then leave the arterial. Notably, with the MAXBAND model, the green bands are assigned to the through paths only, so vehicles along Path 2 would consequently experience higher delays. As shown in Figure 5.18, the M3 model can also outperform the other two models since it has also offered a green band to Path 4. An interesting observation revealed in Figure 5.19 is that the travel time differences along Path 5 by different models are not significant, while MAXBAND has produced a much wider green band along this path than with the other two models. By analyzing the simulation results, we have observed that most traffic volumes along Path 5 have only utilized the first half of the green band due to its low flow rate. Hence, the provided green band by MAXBAND has exceeded the need of traffic volume along Path 5, which results in insignificant improvement of travel time as shown in Figure 5.19.

In conclusion, the proposed M3 model, as expected, can efficiently produce progress bands and less travel times for vehicles along all identified critical paths. To further evaluate the progression-based signal plan over the entire network with different MOEs, Table 6.5 summarizes the resulting average delays, average number of stops, and average speeds. Notably, the average delay by M3 for the target arterial has been reduced by 6.7 seconds and 7.8 seconds, respectively, compared with the

other two models. The same trend could be found for the average number of stops. Regarding the average speed, the M3 model can yield about 7 percent and 8.7 percent improvement. In brief, the proposed M3 model has the promising property of producing progression bands to arterials with heavy flow rates on multiple paths, and the resulting signal plan can also achieve the control objective of minimizing vehicle delays and stops.

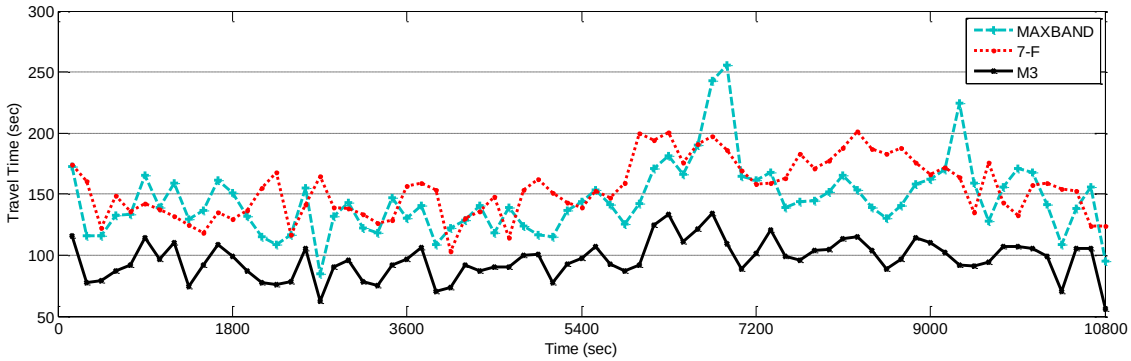


Figure 5.15 The time-dependent travel time along path 1

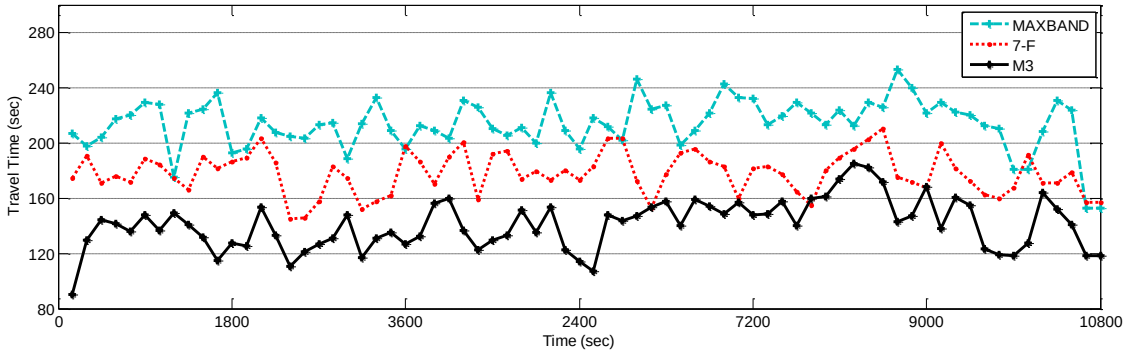


Figure 5.16 The time-dependent travel time along path 2

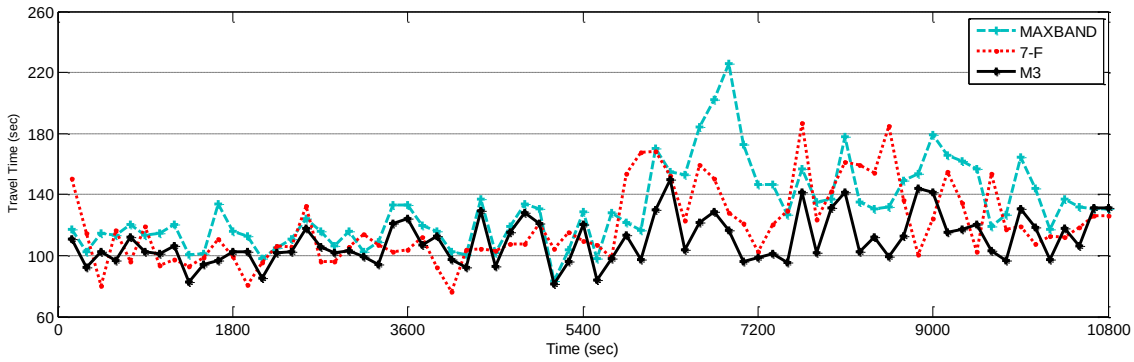


Figure 5.17 The time-dependent travel time along path 3

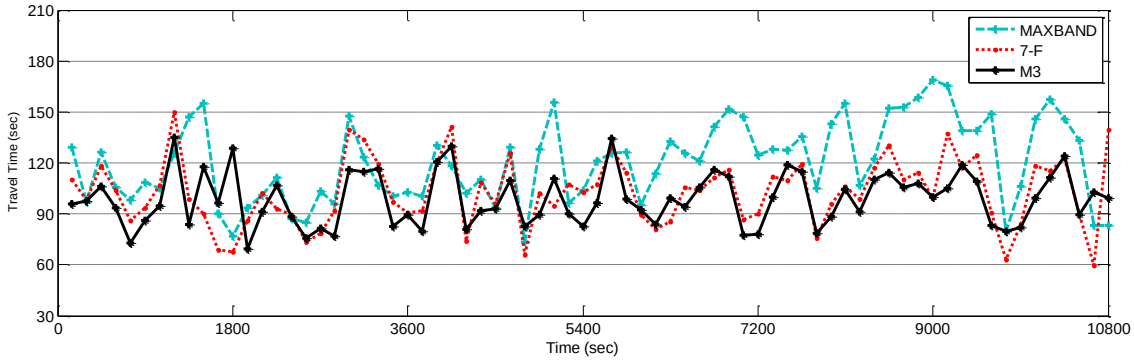


Figure 5.18 The time-dependent travel time along path 4

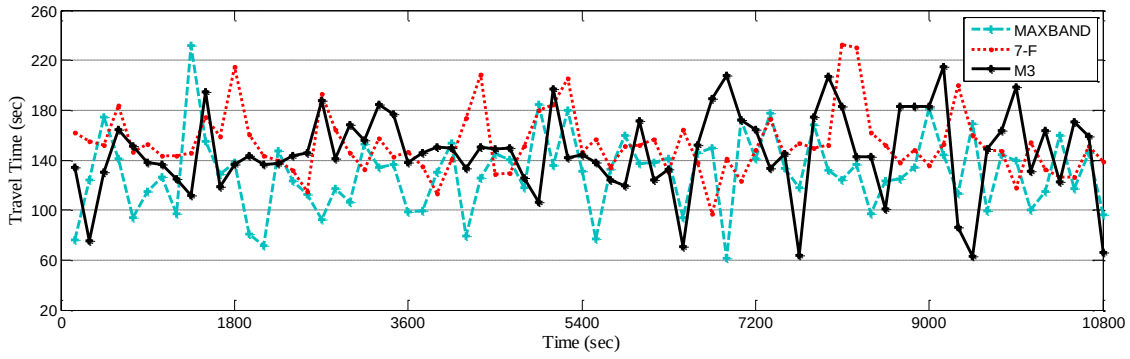


Figure 5.19 The time-dependent travel time along path 5

Table 5.5 Arterial performance under the control of different models

MOEs	Model-1 (MAXBAND)	Model-2 (TRANSYT 7-F)	Model-3 (M3 model)
Average Delay	54.3 secs	55.4 secs	47.6 secs
Average # of Stops	0.972	1.047	0.884
Average Speed	34.7 km/h	31.3 km/h	40.5 km/h

5.6 Closure

Existing models or tools for arterial signal control, focusing either on maximizing the progression for two-way through traffic flows or minimizing their total delay, cannot adequately account for some heavy-path flows that need to take multiple turning movements along the arterial. Hence, providing the progression for not only the through traffic but also other paths of heavy flows is essential in tackling the congestion causing by the overflows from turning bays or link blockage due to spillback in many congested commuting arterials.

This chapter presented a multi-path progression model to concurrently optimize the phase sequence at each intersection and the progression bands for some identified critical path-flows that constitute the complex traffic pattern in congested urban arterials serving mainly as connectors between freeway traffic and surface-street flows. Different from the use of only intersection traffic counts, the proposed model can take full advantage of identified path-flow information in each congested link, and offer the progression for vehicles along each O-D path under the optimized phase sequence. Due to the competing nature among identified path flows, the

proposed model can further identify the set of productive progression paths, based on the distribution of link volumes over each O-D path, and yield the optimal number of progression paths and bandwidths for the target arterial.

The results of extensive numerical investigation with field data have confirmed that the proposed model clearly outperforms the conventional design methods, such as with MAXBAND or TRANSYT, especially for those arterials where heavy path-flows from-and-to the freeway travel over only some of the arterial links but need to execute both left and right turns from the available short turning bays.

Chapter 6 : An Integrated Real-time Control System

6.1 Introduction of the System

Grounded on the pre-timed control models presented in the last chapter, this study has further advanced the proposed control modules for real-time operations, using on-line detector data to adaptively respond to traffic flow fluctuation. The proposed system presented in this chapter has three core modules: off-ramp queue length estimation, arterial adaptive control, and off-ramp priority control. The off-ramp queue length estimation module is used to predict whether or not a potential queue spillover will occur in the following signal cycles. Based on the detected flow data, the arterial adaptive control module functions to dynamically adjust the intersection signal timings and offsets so as to reduce the resulting intersection delay and provide signal progression to those paths of heavy flows. If a potential queue spillover is predicted within the control horizon, the system shall activate its off-ramp priority control module to offer green extension and progression priority to the off-ramp flows.

The interrelations between those key modules in the real-time operational process are shown in Figure 6.1. Notably, the off-ramp queue module will constantly monitor the ramp flows and estimate if any potential queue may spill back to the freeway mainline during the next signal cycle. If causing blockage to the freeway is not of concern within the next control horizon, the real-time integrated system can focus on contending with local arterial congestion by executing the real-time signal progression module with its embedded dynamic programming algorithm for all identified critical paths of traffic flows.

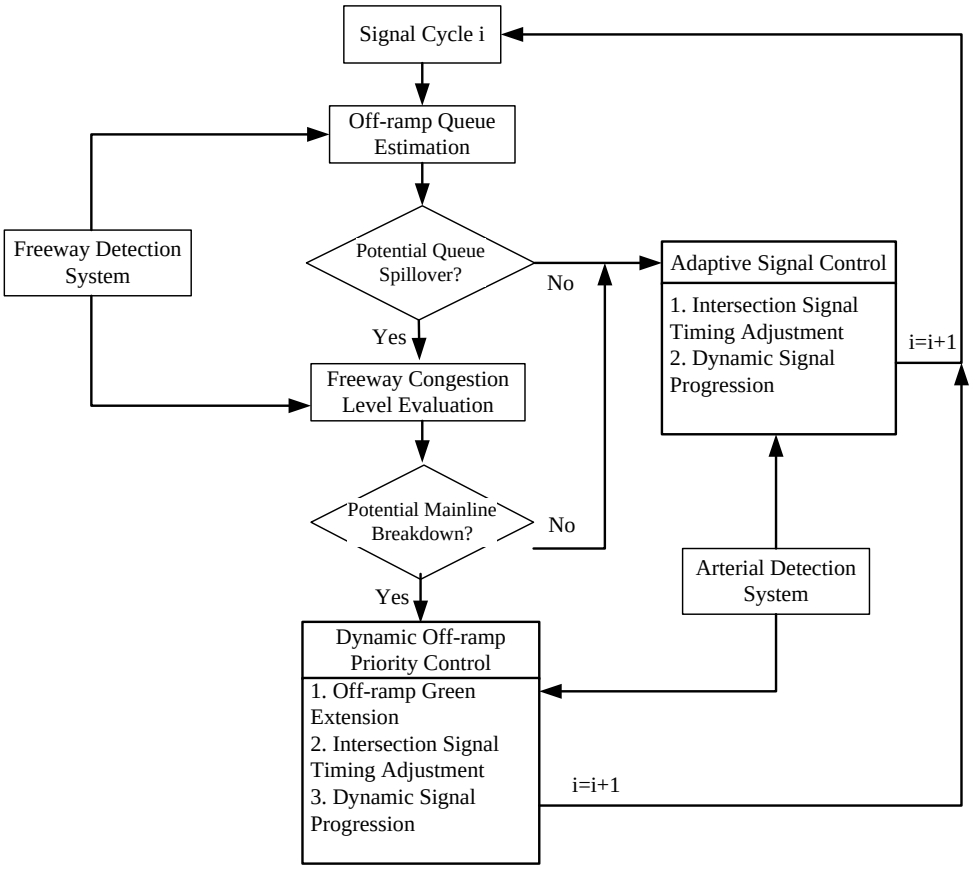


Figure 6.1 The framework of the integrated control system

However, if the off-ramp queue is predicted to spill back to the freeway mainline, the system will first assess whether or not to activate the ramp priority control, based on the estimated impact of ramp queues on the mainline capacity. Note that implementing the real-time priority control needs to accompany by re-optimizing signal timings and offsets at all arterial intersections for all heavy path-flows, but with larger weighting factors for the ramp flows.

Table 6.1 summarizes all key notations used hereafter for description of the proposed real-time integrated control system.

Table 6.1 Key notations in this study

Δt	Length of time interval (seconds)
$q_{up}(t,k)$	Off-ramp upstream flow rate during time interval t in signal cycle k.
$q_{down}(t,k)$	Off-ramp downstream flow rate during time interval t in signal cycle k.
t_{off}	Travel time from an off-ramp upstream detector to the downstream detector.
$\delta(t,k)$	Number of vehicles on the off-ramp during time interval t in signal cycle k.
$\tau_{off}(t,k)$	Off-ramp queue length during time interval t in signal cycle k.
$q_{i,j}(k)$	Number of detected vehciles on link j at intersection i in signal cycle k.
$\mu_{i,j}(t,k)$	Arrival rate on link j at intersection i during time interval t in signal cycle k.
$\gamma_{i,j}(t,k)$	Departure rate on link j at intersection i during time interval t in signal cycle k.
$\tau_{i,j}(t,k)$	Queue length on link j at intersection i during time interval t in signal cycle k.
$s_{i,j}$	Saturation flow rate on link j at intersection i.
$n_{i,j}$	Number of lanes on link j at intersection i.
N_i	Number of intersections on the target arterial.
Np_i	The of signal phases at intersection i.
$g_{i,p}(k)$	Assigned green time for phase p at intersection i in signal cycle k.
$l_{i,p}(k)$	The lost time for phase p at intersection i in signal cycle k.
$\varphi_{p,q}$	Phase sequence parameter: equals "1" if phase p is before q; equals "0"

otherwise.

$\beta_{i,j,p}$	Equals "1" if link *j* receives green in phase *p* at intersection *i*; equals "0" otherwise.
$d_i(k)$	Total intersection delay at intersection *i* in signal cycle *k*.
$c(k)$	Common signal cycle length in signal cycle *k*.
$g_{i,p,max}$	Maximum green time for phase *p* at intersection *i*.
$g_{i,p,min}$	Minimum green time for phase *p* at intersection *i*.
Δg_i	Maximal green difference between two consecutive signal cycles at intersection *i*.
$b_{i,l}(k)$	Green bandwidth for outbound path-flow *l* at between intersections *i* and *i*-1.
$\bar{b}_{i,l}(k)$	Green bandwidth for inbound path-flow *l* at between intersections *i* and *i*-1.
$\theta_i(k)$	Signal offset at intersection *i* in signal cycle *k*.
$\xi_{i,l,p}$	Equals "1" if path *l* receives green in phase *p* at intersection *i*; equals "0" otherwise.
$\Delta\theta_i$	Maximal allowed offset difference between two consecutive signal cycles at intersection *i*.
$g_{off}(k)$	Total green time assigned to the target off-ramp flows in signal cycle *k*.
$e_{off}(k)$	Green extension time assigned to the target off-ramp flows in signal cycle *k*.
L_{off}	Off-ramp link length represented by number of vehicles.
s_{off}	Saturation flow rate for the off-ramp flows.

6.2 Off-ramp Queue Estimation

To develop a reliable off-ramp queue estimation model, this study adopts dual-zone detectors for data detection and collection. As shown in Figure 6.2, such detectors contain a short detection zone for counting traffic flow rates, and a long detection zone for identifying the presence of traffic queues.

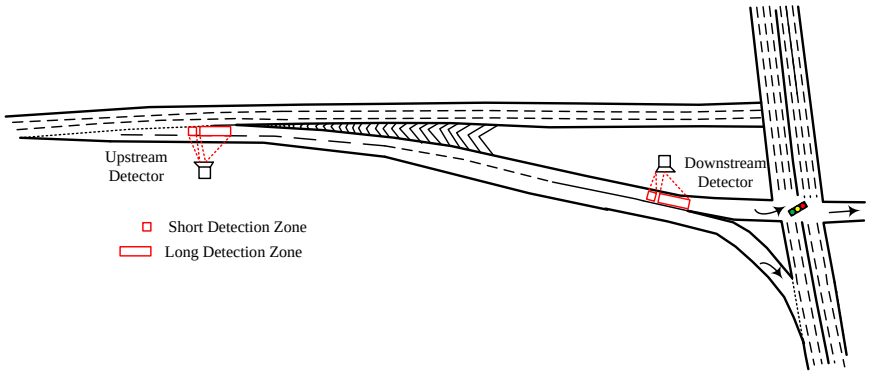


Figure 6.2 Location of dual-zone detectors on the target off-ramp

Since the presence data is more accurate and reliable than the other data provided by the detectors, this study suggests the use of presence data for calculation. In practice, a "0-1" format data with short interval (e.g. 0.1 second) may be obtained from the detectors, as shown in Figure 6.3. For convenience of discussion, the following analysis and computation assume that the data with "0-1" format are available.

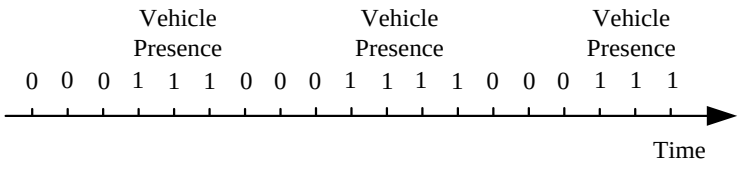


Figure 6.3 An illustrative example for the presence data format

Based on such data, the emerging of multiple continuously “1” or “0” indicates the traffic conditions over the detection area. As shown in Figure 6.4, the presence of “0” from the short detection zone can be used to record the number of passing vehicles within the target time period. Similarly, for long detection zones, the data of multiple “1”, as shown in Figure 6.5, can reflect the formation of queues and, the presence of “0” indicates the clearance of queue.

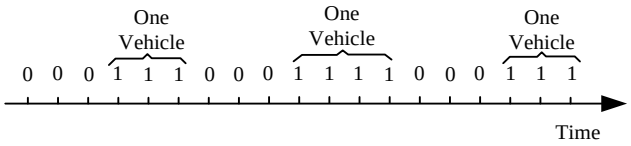


Figure 6.4 The recording of number of passing vehicles by the short detection zone

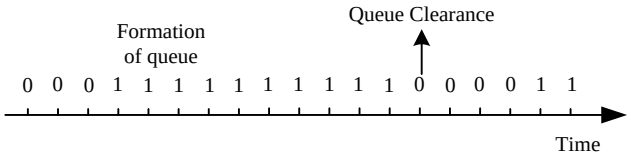


Figure 6.5 The detection of queue formation and clearance by long detection zone

Obviously, some detection errors may exist in practice using the deification methods introduced above. As shown in Figure 6.6, when using the short detection zone to record number of passing vehicles, if the loop distance is longer than the headway between two adjacent vehicles, it may be identified as one large size vehicle. Hence, to ensure the estimation accuracy, the loop distance of short detection zone should be shorter than the minimum vehicle headway. Figure 6.7 shows the

detection errors under two possible conditions. If the loop distance is short than the stopping vehicle headway, the detector may not occasionally identify the formation of queue. Also, when the loop distance is longer than the headway between two consecutively moving vehicles, the detector may mistakenly identify a queue.

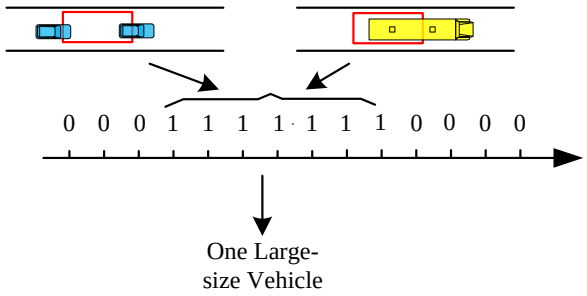


Figure 6.6 Identification errors caused by short detection zone

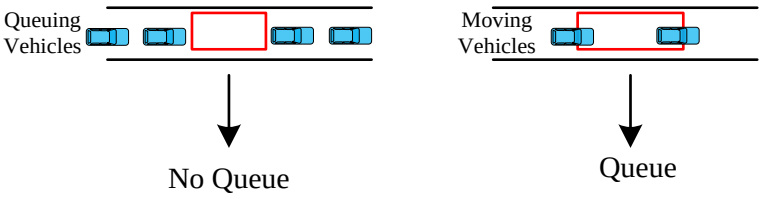


Figure 6.7 Identification errors caused by long detection zone

To overcome the potential errors caused by various traffic conditions, this study develops the following rules for vehicle identification:

- Within the short detection zone: the loop length should be shorter than the minimum vehicle headway;

- Within the long detection zone: 1) the loop length should be longer than the maximum headway between stopping vehicles; 2) only more than *n*

continuously “1” can indicate the formation of queues, where *n* is a pre-calibrated parameter.

With respect to different congestion levels, an effective model for queue estimation needs to tackle the following scenarios: 1) when the inflow exceeds the discharging capacity at the downstream of an off-ramp, the queue may be built up quickly and spillbacks to the freeway mainline; 2) as shown in Figure 6.8, during the green time, the traffic queue could be fully discharged; 3) the traffic queue is not cleared after the green phase and some residual queuing vehicles remains at the off-ramp (see Figure 6.9); 4) as shown in Figure 6.10, queue spillover at the downstream link may affect the queue discharging process and the effective green time would be less than the given green time.

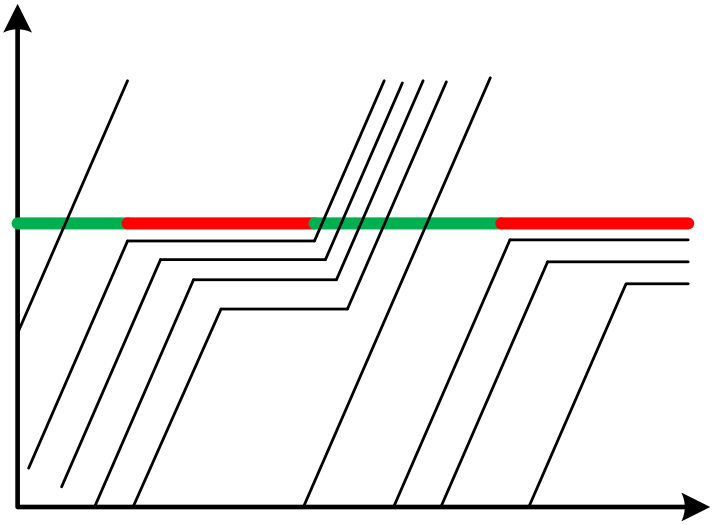


Figure 6.8 Queuing vehicles can be fully discharged during the green phase

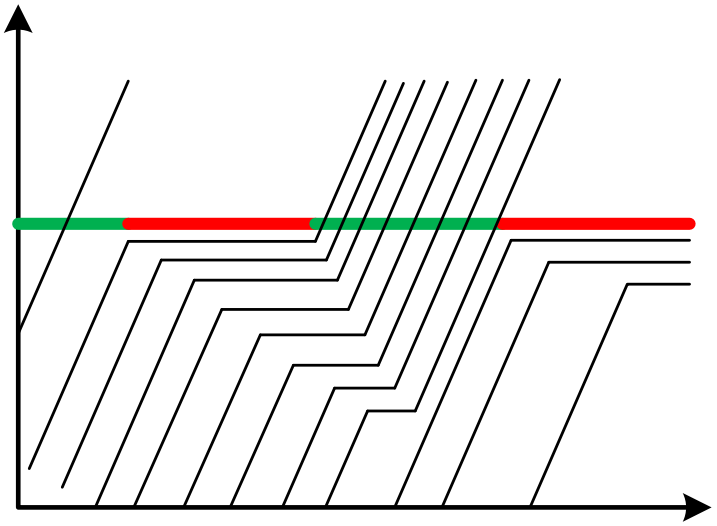


Figure 6.9 Queuing vehicles cannot be fully discharged during the green phase

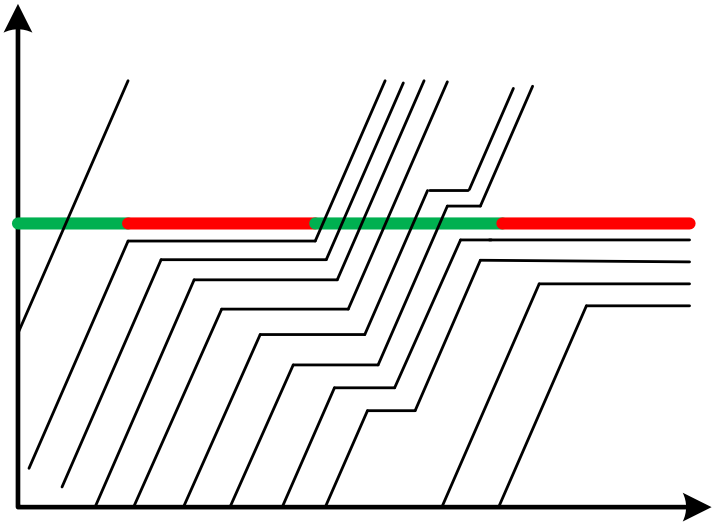


Figure 6.10 Queue spillover at downstream link which affect the discharging process

Assuming that one signal cycle starts from a green phase for off-ramp flow, the proposed model will use the detector data at the end of each signal cycle for analysis.

Figure 6.11 illustrates the estimation process for the off-ramp queue length. In the first scenario where traffic queue already spill back to the freeway mainline, the proposed model will use the data from the long detection zone at the off-ramp upstream for identification. Based on the above analysis method, a queue spillback warning will be provided if the upstream detector data detects the queue.

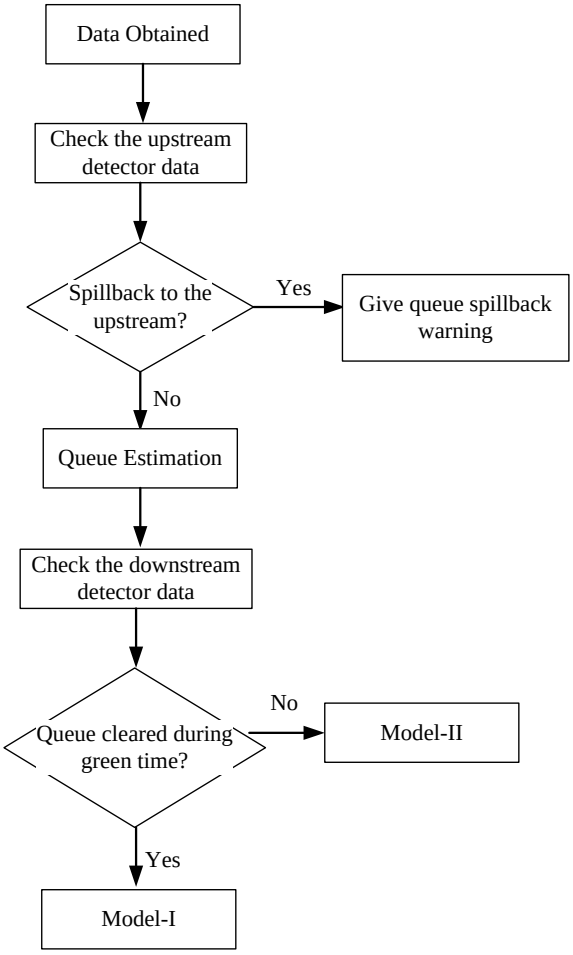


Figure 6.11 Flowchart of the entire queue estimation process

If the off-ramp queue doesn't spill back to the upstream detection zone, the model will further activate queue estimation process. Based on the classification of scenarios in Figure 6.5, the proposed model will then examine the downstream

detector. When the data from the long detection zone indicates the clearance of queue during the green time, Model-I will be used for queue estimation; otherwise Model-II should be implemented.

6.2.1 Formulations for Model-I

As shown in Figure 6.12, assuming that the signal cycle starts from a green phase for the off-ramp flows and the queue is fully discharged at time interval ε, then the number of moving vehicles between two detectors shall equal the number of vehicles passed the upstream detector during time period [$\varepsilon - t_{off}$, ε], and be expressed as follows:

$$\delta(\varepsilon, k) = \sum_{t=\varepsilon - t_{off}}^{\varepsilon} q_{up}(t, k) \tag{6.1}$$

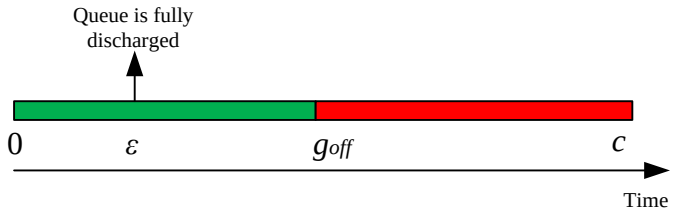


Figure 6.12 Time slots within the target signal cycle

Then, at the end of a green phase, the total number of vehicles between two detectors is given by:

$$\delta(g_{off}, k) = \delta(\varepsilon, k) + \sum_{t=\varepsilon}^{g_{off}} q_{up}(t, k) - \sum_{t=\varepsilon}^{g_{off}} q_{down}(t, k) \tag{6.2}$$

where the second and third terms at the right side of equation (6.2) represent the number of entering and existing vehicles during time period [ε, g_{off}], respectively.

Since no queuing vehicle will be discharged during the red phase, the total number of vehicles between two detectors at the end of a cycle shall be as follows:

$$\delta(c,k) = \delta(g_{off},k) + \sum_{t=g_{off}}^{c(k)} q_{up}(t,k) \tag{6.3}$$

Therefore, the queue length at the end of a cycle could be approximately computed with the following expression:

$$\tau_{off}(c,k) = \delta(c,k) + \delta_0 \tag{6.4}$$

where δ_0 is a constant, reflecting the number of vehicles between downstream detector and the signal stop-line.

6.2.2 Formulations for Model-II

If residual queues exist at the end of a green phase, those queuing vehicles may or may not reach the downstream detector.

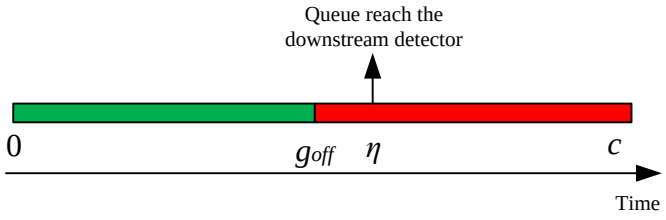


Figure 6.13 Time slots within the target signal cycle

For the former scenario, where traffic queues have reached the downstream detector at time interval η, as shown in Figure 6.13, the total number of vehicles between two detectors at the end of a cycle can be computed as follows:

$$\delta(c,k) = \sum_{t=\eta-t_{off}}^{c(k)} q_{up}(t,k) \tag{6.5}$$

Similarly, the queue at the end of a cycle could be computed with Eq. (6.4).

If the residual queues have exceeded the downstream detector, the queue length at the end of a cycle can be approximated with the following equation:

$$\tau_{off}(c,k) = \tau_{off}(c,k-1) + \sum_{t=1}^{c(k)} q_{up}(t,k) - \sum_{t=1}^{c(k)} q_{down}(t,k) \tag{6.6}$$

With the above logic, the model will compare the estimated queue length at the end of last cycle with the pre-determined threshold, and identify whether or not the potential queue spillover may occur in the projected time horizon.

6.3 Arterial Adaptive Signal Control

To provide the essential information for adaptive signal control, the arterial needs to deploy traffic detectors on all departing links at each intersection, as shown in Figure 6.14, to collect the time-varying flow rate for each approach and movement for optimizing signal timings.

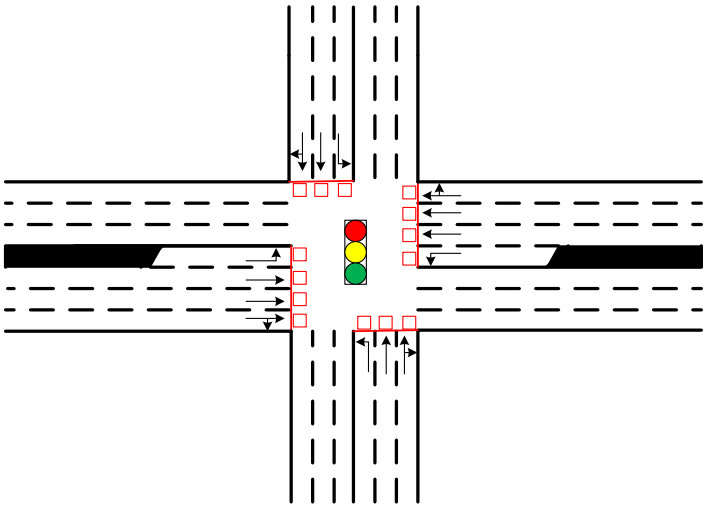


Figure 6.14 Illustration of the detection system at each intersection

To provide a proactive signal control, the system will further adopt its flow prediction module. Based on the detected flow rate at each intersection, one may employs a vast body of algorithms in the statistical literature to perform the prediction. An example of methods which is convenient but reasonably effective is the following smoothing approach:

$$q_{i,j}(k) = \frac{1}{m}\sum_{t=1}^{m} q_{i,j}(k-t) \tag{6.7}$$

where m is the number of previous signal cycles accounted for computation.

The turning flows predicted for each intersection can then serve as the primary input for the adaptive signal control module to compute the cycle length and signal timings. The offsets for all intersections shall also be adjusted accordingly for design of signal progression.

6.3.1 Intersection Signal Timings Adjustment

Given the signal plan and phase sequence at each intersection, this module will increase or decrease the common cycle length by a pre-determined time interval (e.g., 5 seconds) after the computation for each control horizon (e.g, 5 minutes). Similar to the existing adaptive control system, such as SCOOT, the module will select the most congested intersection in the arterial for evaluation. Then, the cycle length will be adjusted to maintain the intersection v/c ratio to be within the target threshold (e.g., 80%).

Within each control horizon, the dynamic process to optimize the signal timings at each intersection can be summarized as follows:

$$M1: \quad Min \quad d_i(k) \tag{6.8}$$

s.t.

$$d_i(k) = \sum_{j=1}^{N_j} \sum_{t=1}^{c} \tau_{i,j}(t,k) \Delta t \tag{6.9}$$

$$\mu_{i,j}(t,k) = \frac{1}{c} q_{i,j}(k) \quad \forall j,t \tag{6.10}$$

$$r_{i,j}(t,k) = \begin{cases} s_{i,j} \Delta t & if \quad green \\ 0 & if \quad red \end{cases} \quad \forall j,t \tag{6.11}$$

$$\tau_{i,j}(0,k) = \tau_{i,j}(c,k-1) \quad \forall j \tag{6.12}$$

$$\tau_{i,j}(t,k) = Max[\tau_{i,j}(t-1,k) + \mu_{i,j}(t,k) - r_{i,j}(t,k), 0] \quad \forall j,t \tag{6.13}$$

$$\sum_{p=1}^{N_p}(g_{i,p}(k)+l_{i,p}(k))=c(k) \tag{6.14}$$

$$g_{i,p,\min} \le g_{i,p}(k) \le g_{i,p,\max} \tag{6.15}$$

$$g_{i,p}(k-1)-\Delta g_i \le g_{i,p}(k) \le g_{i,p}(k-1)+\Delta g_i \tag{6.16}$$

In the above optimization model, the control objective (Eq. (6.8)) is to minimize the total intersection delay. Eq. (6.9) denotes the total delay with the queue length, $\tau_{i,j}(t,k)$, on all intersection links and the length of time interval, Δt. Eq. (6.10) is used to estimate the vehicle arrival rate on all links during each time interval.

Eq. (6.11) estimates the total number of departing vehicles from each link, and Eq. (6.12) approximates the initial link queue at the start of a signal cycle. The queue length at time interval *t*, computed with Eq. (6.13), depends on the queue length during last time interval, the arrival flow, and departure flow. Eq. (6.14) ensures that the summation of effective green times and lost times equals the cycle length, and Eq. (6.15) sets the minimal and maximal green times for each phase. To stabilize the green time transitions between phases, one shall let the adjustment of signal timings be constrained within a pre-calibrated interval (e.g., 6 seconds), as shown in Eq. (6.16).

The algorithm for the proposed M1 model to be efficiently solved for on-line applications is presented below:

Step 1: Initialization. Let p = 1 and get the green time of each phase at the previous

signal cycle;

Step 2: For phase *p*, change the green time by α seconds (could be negative or positive) by solving the following sub-problem:

$$\begin{aligned} &\alpha = \arg\min\{d_i(k);\ m \in N_p,\ m \neq p\} \\ &s.t.\quad g_{i,p}(k) = g_{i,p}(k-1) + \alpha \\ &\qquad g_{i,m}(k) = g_{i,m}(k-1) - \alpha \\ &\qquad -\Delta g_i \le \alpha \le \Delta g_i \\ &\qquad g_{i,p,\min} \le g_{i,p}(k) \le g_{i,p,\max} \\ &\qquad g_{i,m,\min} \le g_{i,m}(k) \le g_{i,m,\max} \end{aligned} \tag{6.17}$$

Step 3: Let $p = p + 1$. If $p > |N_{pi}|$, stop; otherwise go back to Step 2.

It is noticeable that the sub-problem (6.17) in Step 2 is to transfer several seconds of green time between the target phase *p* and one of other phases. To solve this sub-problem and ensure the computation efficiency for on-line applications, the system can use a gradient searching approach to find the optimal value for α. And the solution process could be explained with the following steps:

// initialization:

$$opt\_d_i(k) = \left\{ \sum_{j=1}^{N_j} \sum_{t=1}^{c} \tau_{i,j}(t,k)\Delta t \,\middle|\, g_{i,p}(k) = g_{i,p}(k-1)\ \forall p \right\};\ d_i(k) = 0;$$

// signal timing adjustment for phase p

// Step 1: find the adjustment direction

$\alpha = 1$;

$$d_i(k) = \min_{m \neq p} \left\{ \sum_{j=1}^{N_j} \sum_{t=1}^{c} \tau_{i,j}(t,k) \Delta t \,\middle|\, g_{i,p}(k) = g_{i,p}(k-1) + \alpha,\ g_{i,m}(k) = g_{i,m}(k-1) - \alpha \right\}$$

if $d_i(k) < opt\_d_i(k)$

{ $opt\_d_i(k) = d_i(k)$;

$direction = 1$;

}

else

{ $\alpha = -1$;

$$d_i(k) = \min_{m \neq p} \left\{ \sum_{j=1}^{N_j} \sum_{t=1}^{c} \tau_{i,j}(t,k) \Delta t \,\middle|\, g_{i,p}(k) = g_{i,p}(k-1) + \alpha,\ g_{i,m}(k) = g_{i,m}(k-1) - \alpha \right\};$$

if $d_i(k) < opt\_d_i(k)$

{ $opt\_d_i(k) = d_i(k)$;

$direction = -1$;

}

else

{ $direction = 0$; break}

}

// continue to adjust the green time along the direction until no delay reduction is found

$\alpha = direction$;

$$d_i(k) = \min_{m \neq p} \left\{ \sum_{j=1}^{N_j} \sum_{t=1}^{c} \tau_{i,j}(t,k) \Delta t \,\middle|\, g_{i,p}(k) = g_{i,p}(k-1) + \alpha,\ g_{i,m}(k) = g_{i,m}(k-1) - \alpha \right\};$$

do while ($d_i(k) <= opt\_d_i(k)$)

{ $\alpha = \alpha + direction$;

$opt\_d_i(k) = d_i(k)$;

$$d_i(k) = \min_{m \neq p} \left\{ \sum_{j=1}^{N_j} \sum_{t=1}^{c} \tau_{i,j}(t,k) \Delta t \,\middle|\, g_{i,p}(k) = g_{i,p}(k-1) + \alpha,\ g_{i,m}(k) = g_{i,m}(k-1) - \alpha \right\}$$

}

6.3.2 Adaptive Signal Progression Control

Aside from dynamically adjusting signal timings at each intersection, it is also essential to re-compute the offsets for signal progression. Following the same concept of multi-path progression for off-line applications, this study has further advanced its

functions for use with real-time available detector information. For convenience of discussion, let $t_{i,l,1}$ and $t_{i,l,2}$ denotes the start and end of the green time for path-flow l at intersection i. The notations for key model variables used are shown in Figure 6.15 and Figure 6.16.

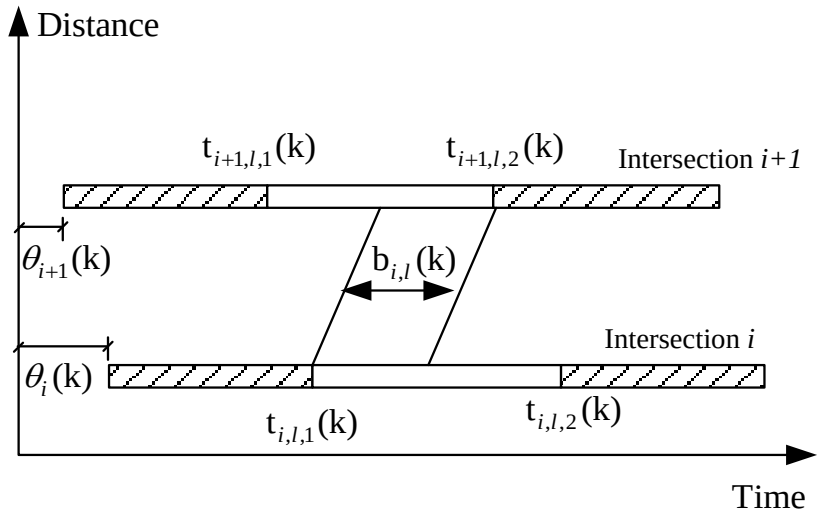


Figure 6.15 Green band of an outbound path between two intersections

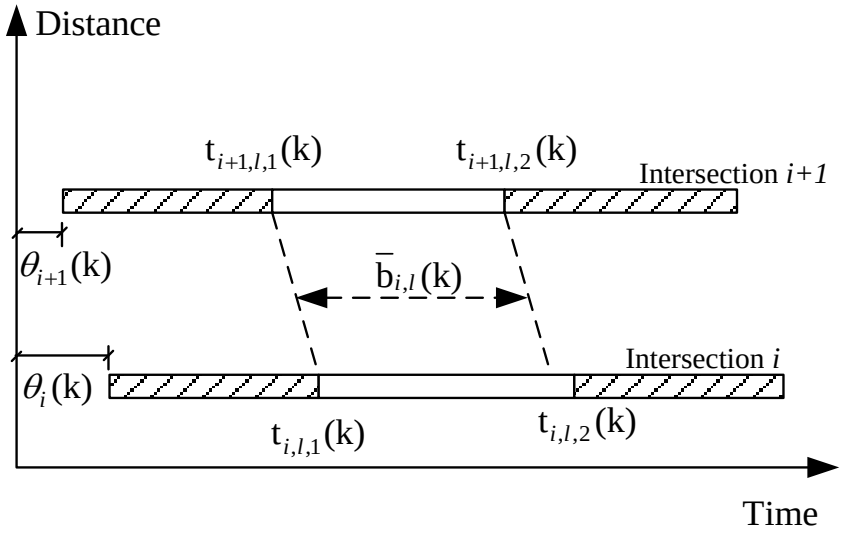


Figure 6.16 Green band of an inbound path between two intersections

Given the cycle length and signal timings generated from the adaptive signal control function, the following equations present the set of functions for the dynamic signal progression model:

$$M2: \ Max \ \sum_i \sum_l \phi_l(\mathrm{k}) \mathrm{b}_{i,l}(\mathrm{k}) + \sum_i \sum_l \bar{\phi}_l(\mathrm{k}) \bar{\mathrm{b}}_{i,l}(\mathrm{k}) \tag{6.18}$$

s.t.

$$b_{i,l}(\mathrm{k}) = Max[Min(\mathrm{t}_{i+1,l,2}(\mathrm{k}), \mathrm{t}_{i,l,2}(\mathrm{k}) + \mathrm{t}_{i,i+1}(\mathrm{k})) - Max(\mathrm{t}_{i,l,1}(\mathrm{k}) + \mathrm{t}_{i,i+1}(\mathrm{k}), \mathrm{t}_{i+1,l,1}(\mathrm{k})), 0] \tag{6.19}$$

$$\bar{b}_{i,j}(\mathrm{k}) = Max[Min(\mathrm{t}_{i,l,2}(\mathrm{k}), \mathrm{t}_{i+1,l,2}(\mathrm{k}) + \mathrm{t}_{i+1,i}(\mathrm{k})) - Max(\mathrm{t}_{i,l,1}(\mathrm{k}), \mathrm{t}_{i+1,l,1}(\mathrm{k}) + \mathrm{t}_{i+1,i}(\mathrm{k})), 0] \tag{6.20}$$

$$\mathrm{t}_{i,l,1}(\mathrm{k}) = \sum_q \sum_p \zeta_{i,l,p} \varphi_{p,q} g_{i,p}(\mathrm{k}) + \theta_i(\mathrm{k}) \tag{6.21}$$

$$\mathrm{t}_{i,l,2}(\mathrm{k}) = \sum_q \sum_p \zeta_{i,l,p} \varphi_{p,q} g_{i,p}(\mathrm{k}) + \sum_p \zeta_{i,j,p} \varphi_{p,q} g_{i,p}(\mathrm{k}) + \theta_i(\mathrm{k}) \tag{6.22}$$

$$\theta_i(\mathrm{k}-1) - \Delta\theta_i \le \theta_i(\mathrm{k}) \le \theta_i(\mathrm{k}-1) + \Delta\theta_i \tag{6.23}$$

where, $\phi_l(\mathrm{k})$ and $\bar{\phi}_l(\mathrm{k})$ are the weighting factors.

The control objective (Eq. (6.18)) is to maximize the sum of weighted green bands between adjacent intersections for all path-flows. Also note that those weighting factors could be determined by traffic volumes over each path. Eq. (6.19) and Eq. (6.20) are used to compute the corresponding green bandwidths for an outbound and an inbound path-flow, respectively, between two adjacent intersections

i and *i* + 1. Eq. (6.21) and Eq. (6.22) are used to compute the start and end of a green phase for path *l* at intersection *i*, based on the intersection offset, signal timings, and phase plan. Eq. (6.23) functions to constrain the change of offsets within a preset range (e.g., 6 seconds).

To solve the proposed optimization model (*M*2), shown with Eqs. (6.18)-(6.23), this study proposes a dynamic programming solution algorithm. Given Eq. (6.23), the feasible solution set for the new offset at each intersection is given below:

$$\Theta_i(\mathrm{k}) = \left\{\theta_i(\mathrm{k}-1) - \Delta\theta_i, \theta_i(\mathrm{k}-1) - \Delta\theta_i + 1, \cdots, \theta_i(\mathrm{k}-1) + \Delta\theta_i\right\} \tag{6.24}$$

Also, the total green bandwidth, $B_i(\theta_i)$, for all path-flows between intersections *i* and *i* – 1 could be calculated with Eq. (6.25):

$$B_i(\theta_i) = \sum_l \beta_{i-1,l,p}\beta_{i,l,p}\phi_l(k)b_{i,l}(k) + \sum_l \beta_{i-1,l,p}\beta_{i,l,p}\bar{\phi}_l(k)\bar{b}_{i,l}(k) \tag{6.25}$$

where the bandwidths, $\mathrm{b}_{i,l}(\mathrm{k})$ and $\bar{b}_{i,l}(k)$, for each outbound and inbound path-flows, respectively, could be obtained by Eqs. (6.19) – (6.22).

Let $f_i(.)$ denote the accumulated performance measure, the algorithm consists of the following steps:

Step 1: set $i = 1$, $\theta_1(\mathrm{k}) = 0$, and $f_i(0) = 0$;

Step 2: $i = i + 1$;

$$f_i(\theta_i^*(k)) = \min_{\theta_i(k)}\left\{ f_{i-1}(\theta_{i-1}^*(k)) + B_i(\theta_i(k)) \middle| \theta_i \in \Theta_i(\mathrm{k})\right\}$$

Record $\theta_i^*(k)$ as the optimal solution in Step 2.

Step 3: if $i < N_i$, go to Step 2.

Else, Stop.

6.4 Dynamic Off-ramp Priority Control

As shown in Figure 6.2, when potential off-ramp queue spillover is detected by the queue estimation module, the real-time system needs to assess the spillover impact on the freeway mainline, and decide the necessity of activating the off-ramp queue priority control. Such functions include two control steps: 1) increasing the green time for the off-ramp flows; and 2) providing signal progression priority to those path-flows from the target off-ramp.

6.4.1 Intersection Signal Timing Adjustment with Off-ramp Priority

To limit the negative impact to the local traffic, the real-time adaptive control system needs to reduce the activation frequency for such priority control. Furthermore, the corresponding extension for the minimal green in signal cycle *k* shall ensure the prevention of queue spillover until the end of the following signal cycle. Eqs. (6.26) and (6.27) show those two constraints for the maximal queue length during cycle *k* and *k*+1, respectively:

$$\tau_{off}(c,k-1) - s_{off}(g_{off}(k) + e_{off}(k)) + q_{off}(k) = \tau_{off}(c,k) < L_{off} \tag{6.26}$$

$$\tau_{off}(c,k) - s_{off} g_{off}(k+1) + q_{off}(k+1) = \tau_{off}(c,k+1) < L_{off} \tag{6.27}$$

Given the two constraints, the minimal green extension shall satisfy the following two constraints:

$$e_{off}^{\min}(k) \geq \frac{L_{off} - \tau_{off}(c, k-1) + s_{off} g_{off}(k) - q_{off}(k)}{s_{off}} \tag{6.28}$$

$$e_{off}^{\min}(k) \geq \frac{L_{off} - \tau_{off}(c, k-1) + \sum_{m=k}^{k+1}(s_{off} g_{off}(m) - q_{off}(m))}{s_{off}} \tag{6.29}$$

Then, for simplicity, the minimal green extension could be determined by the following expression:

$$e_{off}^{\min}(k) = \frac{L_{off} - \tau_{off}(c, k-1) + Max[s_{off} g_{off}(k) - q_{off}(k), \sum_{m=k}^{k+1}(s_{off} g_{off}(m) - q_{off}(m))]}{s_{off}} \tag{6.30}$$

In summary, the adaptive signal control model with an off-ramp priority function is summarized below:

$M3$: $Min \quad d_i(k)$

s.t.

$$d_i(k) = \sum_{j=1}^{N_j} \sum_{t=1}^{c} \tau_{i,j}(t, k) \Delta t \tag{6.31}$$

$$\mu_{i,j}(t, k) = \frac{1}{c} q_{i,j}(k) \quad \forall j, t \tag{6.32}$$

$$r_{i,j}(t,k) = \begin{cases} s_{i,j}\Delta t & if \ \ green \\ 0 & if \ \ red \end{cases} \quad \forall j,t \tag{6.33}$$

$$\tau_{i,j}(0,k) = \tau_{i,j}(c,k-1) \quad \forall j \tag{6.34}$$

$$\tau_{i,j}(t,k) = Max[\tau_{i,j}(t-1,k) + \mu_{i,j}(t,k) - r_{i,j}(t,k), 0] \quad \forall j,t \tag{6.35}$$

$$\sum_{p=1}^{N_p} (g_{i,p}(k) + l_{i,p}(k)) = c(k) \tag{6.36}$$

$$g_{i,p,\min} \le g_{i,p}(k) \le g_{i,p,\max} \tag{6.37}$$

$$g_{i,p}(k-1) - \Delta g_i \le g_{i,p}(k) \le g_{i,p}(k-1) + \Delta g_i \tag{6.38}$$

$$g_{off}(k) - g_{off}(k-1) \ge e_{off}^{\min}(k) \tag{6.39}$$

where Eqs. (6.31)-(6.38) are used in the adaptive signal control module; Eq. (6.39) is an additional constraint that constrains the minimal green extension.

Also, by replacing the minimal green time for off-ramp flows with $e_{off}^{\min}(k) + g_{off}(k-1)$, one can use the solution algorithm summarized below to solve model M3:

Step 1: Initialization. Let p = 1 and get the green time of each phase at the previous signal cycle;

Step 2: For phase p, change the green time by α seconds (could be negative or positive) by solving the following sub-problem:

$$
\begin{aligned}
&\alpha = \arg\min\{d_i(k);\, m \in N_p,\, m \neq p\} \\
&s.t.\quad g_{i,p}(k) = g_{i,p}(k-1) + \alpha \\
&\qquad g_{i,m}(k) = g_{i,m}(k-1) - \alpha \\
&\qquad -\Delta g_i \le \alpha \le \Delta g_i \\
&\qquad g_{i,p,\min} \le g_{i,p}(k) \le g_{i,p,\max} \\
&\qquad g_{i,m,\min} \le g_{i,m}(k) \le g_{i,m,\max} \\
&\qquad g_{off}(k) - g_{off}(k-1) \ge e_{off}^{\min}(k)
\end{aligned}
\tag{6.40}
$$

Step 3: Let $p = p + 1$. If $p > |N_{pi}|$, stop; otherwise go back to Step 2.

6.4.2 Signal Progression Design with Off-ramp Priority

With the green extension, the real-time adaptive system shall also activate the signal progression module to revise the offsets to provide priority control to the off-ramp path-flows. Specifically, a minimal total green bandwidth for those path-flows, $B_{off}^{\min}$, shall be sufficient to discharge the queuing off-ramp vehicles. Hence, the control model could be summarized as follows:

$$
M4:\ \ Max\ \sum_i \sum_l \phi_l(\mathrm{k})\, \mathrm{b}_{i,l}(\mathrm{k}) + \sum_i \sum_l \bar{\phi}_l(\mathrm{k})\, \bar{\mathrm{b}}_{i,l}(\mathrm{k})
$$

s.t.

$$
b_{i,l}(\mathrm{k}) = Max[Min(\mathrm{t}_{i+1,l,2}(\mathrm{k}), \mathrm{t}_{i,l,2}(\mathrm{k}) + \mathrm{t}_{i,i+1}(\mathrm{k})) - Max(\mathrm{t}_{i,l,1}(\mathrm{k}) + \mathrm{t}_{i,i+1}(\mathrm{k}), \mathrm{t}_{i+1,l,1}(\mathrm{k})), 0]
\tag{6.41}
$$

$$
\bar{b}_{i,j}(\mathrm{k}) = Max[Min(\mathrm{t}_{i,l,2}(\mathrm{k}), \mathrm{t}_{i+1,l,2}(\mathrm{k}) + \mathrm{t}_{i+1,i}(\mathrm{k})) - Max(\mathrm{t}_{i,l,1}(\mathrm{k}), \mathrm{t}_{i+1,l,1}(\mathrm{k}) + \mathrm{t}_{i+1,i}(\mathrm{k})), 0]
\tag{6.42}
$$

$$t_{i,l,1}(k) = \sum_{q}\sum_{p}\zeta_{i,l,p}\varphi_{p,q}g_{i,p}(k) + \theta_i(k) \tag{6.43}$$

$$t_{i,l,2}(k) = \sum_{q}\sum_{p}\zeta_{i,l,p}\varphi_{p,q}g_{i,p}(k) + \sum_{p}\zeta_{i,j,p}\varphi_{p,q}g_{i,p}(k) + \theta_i(k) \tag{6.44}$$

$$\sum_{l\in\Gamma_{off}} b_{i,l}(k) > B_{off}^{\min} \tag{6.45}$$

$$\sum_{l\in\Gamma_{off}} \bar{b}_{i,l}(k) > B_{off}^{\min} \tag{6.46}$$

Similar to the control model M2, the solution algorithm with dynamic programming for M4 is given below:

Step 1: set $i = 1$, $\theta_1(k) = 0$, and $f_i(0) = 0$;

Step 2: $i = i + 1$;

$$f_i(\theta_i^*(k)) = \min_{\theta_i(k)}\left\{ f_{i-1}(\theta_{i-1}^*(k)) + B_i(\theta_i(k)) \middle| \theta_i \in \Theta_i(k);\ \sum_{l\in\Gamma_{off}} b_{i,l}(k) > B_{off}^{\min};\ \sum_{l\in\Gamma_{off}} \bar{b}_{i,l}(k) > B_{off}^{\min} \right\}$$

Record $\theta_i^*(k)$ as the optimal solution in Step 2.

Step 3: if $i < N_i$, go to Step 2.

Else, Stop.

6.5 Numerical Tests

Based on the evaluation results given in the last chapter, one can still observe the travel time increase along the freeway mainline with the pre-timed control module

(see Figure 5.13), caused by the traffic fluctuation and the resulting off-ramp queue spillover during the peak hours.

Hence, to evaluate the effectiveness and necessity of the proposed real-time control module, this study further uses the selected field site, shown in Figure 5.11, for numerical evaluation. To simulate the traffic fluctuation in practice, Table 6.2 summarizes the demand patterns used for system evaluation.

Table 6.2 The two-hour demand patterns for the three intersections

Time	Intersection	Westbound			Northbound			Eastbound			Southbound		
		L	T	R	L	T	R	L	T	R	L	T	R
16:30-	1	/	665	/	995	/	968	443	840	/	/	/	/
17:00	2	269	1391	/	/	/	/	/	921	/	362	/	403
	3	287	872	635	/	238	300	82	653	91	658	286	69
17:00-	1	/	924	/	1090	/	1073	518	976	/	/	/	/
17:30	2	478	1506	/	/	/	/	/	907	/	587	/	672
	3	332	1056	790	/	274	451	82	834	91	729	272	74
17:30-	1	/	924	/	1090	/	1073	518	976	/	/	/	/
18:00	2	508	1506	/	/	/	/	/	907	/	587	/	672
	3	332	1056	790	/	274	451	82	834	91	729	272	74
18:00-	1	/	786	/	1050	/	1011	441	968	/	/	/	/
18:30	2	418	1418	/	/	/	/	/	857	/	551	/	729
	3	363	962	822	/	288	371	88	832	86	761	354	64

Based on the collected field data, those key control parameters used in the numerical analysis are set as follows:

- The length of time interval, Δt, is set to be 1 second;
- The average travel time, t_{off}, between two off-ramp detectors is 15 seconds;
- The saturation flow rate at all intersections, $s_{i,j}$, is 1700 vehicles per hour;
- The green lost time, $l_{i,p}$, for each phase is 3 seconds;
- The minimal and maximal green time, $g_{i,p,min}$ and $g_{i,p,max}$, for each phase is 8 seconds and 100 seconds, respectively;
- The maximal allowed green time, Δg_i, difference between two consecutive signal cycles is 6 seconds;
- The maximal allowed offset difference, $\Delta \theta_i$, between two consecutive cycles is 4 seconds;
- The storage capacity of the off-ramp link, L_{off}, is 45 vehicles;
- The number of signal cycles, m, for moving average flow prediction is 10;
- The progression weighting factors of five path-flows are set to be 0.3, 0.2, 0.2, 0.2, and 0.1, respectively.
- The weighting factors for off-ramp path-flows (Path 1 and Path 3) are changed to 0.4 when the priority control is granted.

To demonstrate the effectiveness of the proposed real-time control module, this study concurrently evaluates the following three systems for comparison:

(1) Pre-timed Control System: Using the aggregate data presented in Table 6.2, the target arterial for evaluation adopts the signal optimization model and multi-

path progression model introduced in the last chapter to design the signal plans. The green times and offset at each intersection remain unchanged during the control period.

(2) Adaptive Control System: Only the proposed adaptive signal control model and dynamic signal progression model are implemented for numerical analysis.

(3) Proposed System: Following the control logic shown in Figure 6.2, the target arterial is set to deploy the proposed integrated real-time signal control system, including the off-ramp queue estimation, arterial signal adaptive control, and off-ramp priority control.

Since the proposed integrated real-time system may activate the off-ramp priority control when off-ramp queue spillover is predicted, its effectiveness may vary with the estimation accuracy of the off-ramp queue model. With the well-calibrated simulation platform, the comparison between the estimated off-ramp and actual queues (at the end of red phase), are shown in Figure 6.17 and 6.18, respectively.

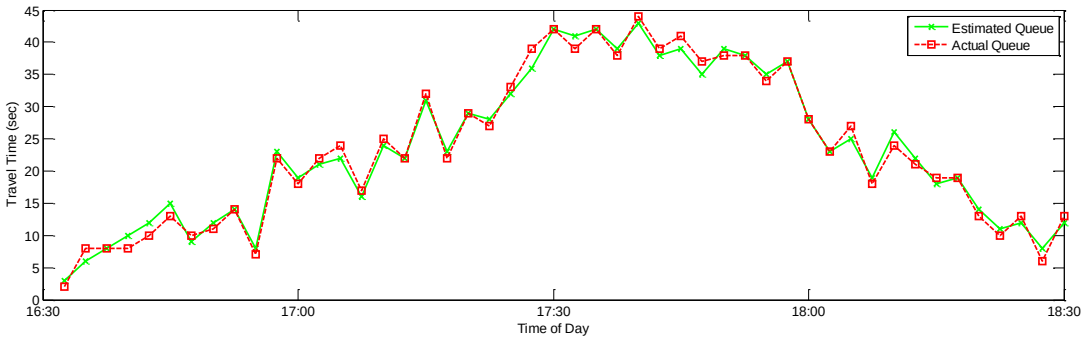


Figure 6.17 Comparison of estimated and actual queue length at the off-ramp

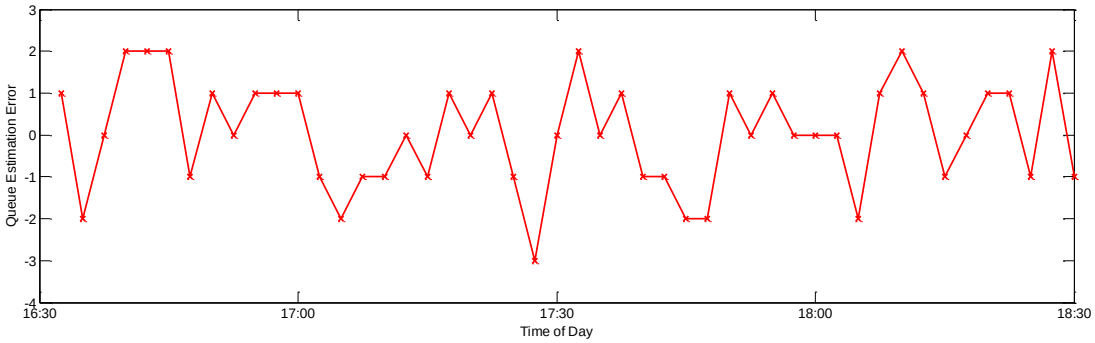


Figure 6.18 The estimation errors of the off-ramp queue estimation model

As shown in Figure 6.17, one can observe that the off-ramp queue length at the end of each signal cycle show a tendency of increasing with off-ramp flows. During the peak period (17:15-18:00), the queue length can reach up to 44 vehicles, which is quite close to the off-ramp storage capacity (i.e., 48 vehicles). However, with the priority control function, no spillover has been detected in this case. Also by comparing the estimated queue with actual queue, no significant difference could be observed here. To further analyze the estimated errors produced by the proposed model, Figure 6.18 shows that the differences between the estimated and actual queue lengths fall within the range [-3, 2] vehicles, which is acceptable for operations.

Figure 6.19 shows the activation frequency of the off-ramp priority control along with the green extension granted to the off-ramp flows. As revealing in the graphical results, the priority function is only activated during the peak period and the maximal extended green time is only 18 seconds. However, with such a function, the off-ramp queue spillover can be successfully prevented, as evidenced by the results shown in Figure 6.17.

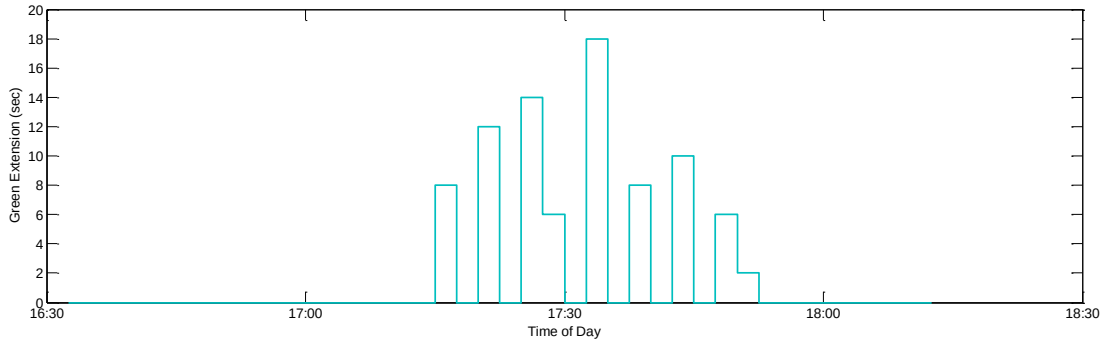


Figure 6.19 Green extension time granted to the off-ramp flows

Evaluation of the freeway mainline performance

Figure 6.20 shows the results that both pre-timed control and adaptive control systems fail to prevent the queue spillover at the target off-ramp, and the travel time along the freeway mainline segment has increased significantly over the congested time period (i.e., 17:15-18:00).

However, with the off-ramp priority control function, the freeway mainline flows under the proposed integrated system can still maintain a relatively higher travel speed during that period, and an insignificant increase in the freeway travel time. Hence, the experimental results with the simulation have shown that the proposed control strategies are likely to yield satisfactory results, and successfully discharge the off-ramp queue to prevent the excessive traffic spillover.

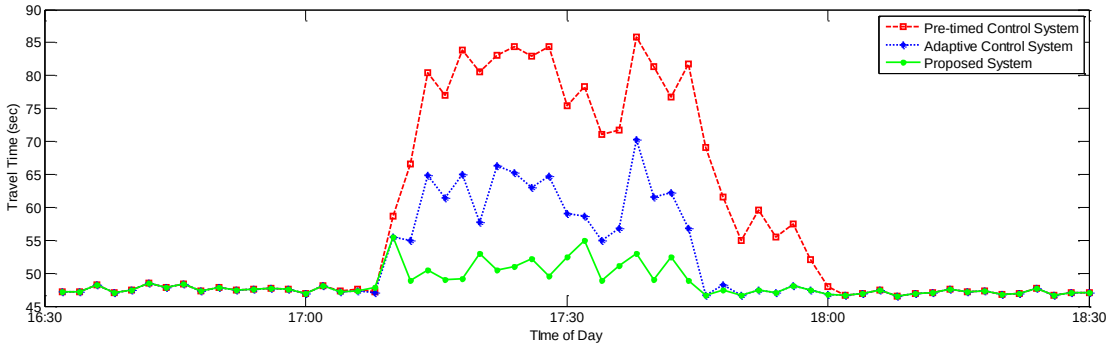


Figure 6.20 The time-dependent travel time along the freeway marline

Evaluation of the critical path-flows

This subsection evaluates the travel times among those five critical path-flows from both off-ramp and local arterial. Compared with the pre-timed control system, the other two systems can produce much lower travel time along Path 1, as evidenced by the results in Figure 6.21. This is due to the fact that the dynamic signal progression function can effectively adjust the intersection offset in response to the change in traffic flow patterns. Also, the proposed real-time system can outperform the adaptive control system since it offers Path-1 with green time extension and progression priority when a potential freeway breakdown caused by the off-ramp queue spillover has been predicted. Similar observations could also be found in Figure 6.23 since Path-3 shares most signal phases with Path-1.

Figure 6.22 shows the comparison of travel times along Path-2. Obviously, the adaptive control system can still outperform the pre-timed system in reducing travel time. However, with the proposed integrated system, the travel time along this path has been increased during the peak-period (i.e., 17:15 to 18:00). This is due to the fact

that with the green extension time granted to the off-ramp flows, the green times for other traffic movements on the arterial will consequently be reduced. Also, the off-ramp flows along both Path-1 and Path-3 could receive progression priority when the control function is activated, which can bring some negative benefits to the traffic along Path-2. Similar observations could also be found in Figure 6.24. Also note that although Path-4 is from the southbound off-ramp, it is not considered for priority control in this system since no queue spillover has been observed during the control period at the study site.

Figure 6.25 further shows that the travel time differences along Path 5 with different models are not significant, even though the proposed model may sacrifice some operation benefits of the local traffic when granting priority control to the off-ramp flows. By analyzing the simulation results, it is noticeable that most traffic volumes along Path 5 have utilized only half of its green band due to its low flow rate. Hence, the provided green bands by the pre-timed control system and adaptive control system have exceeded the need of traffic volume along Path 5, which results in insignificant improvement on travel time, as shown in Figure 6.25.

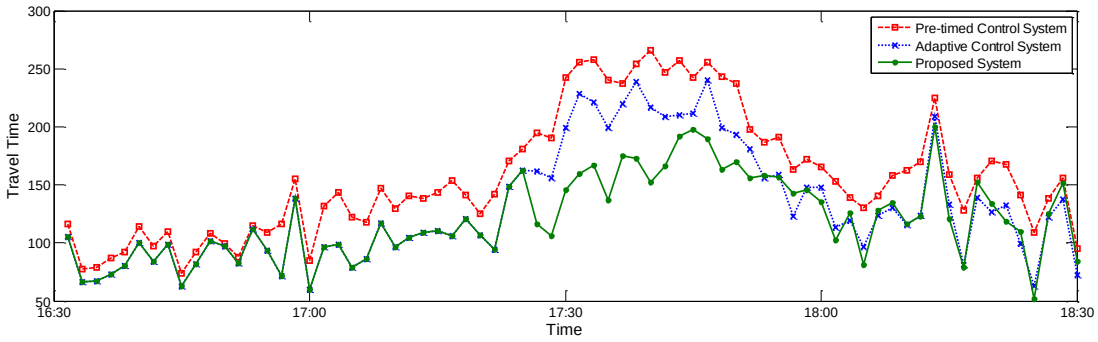


Figure 6.21 The time-dependent travel time along path 1

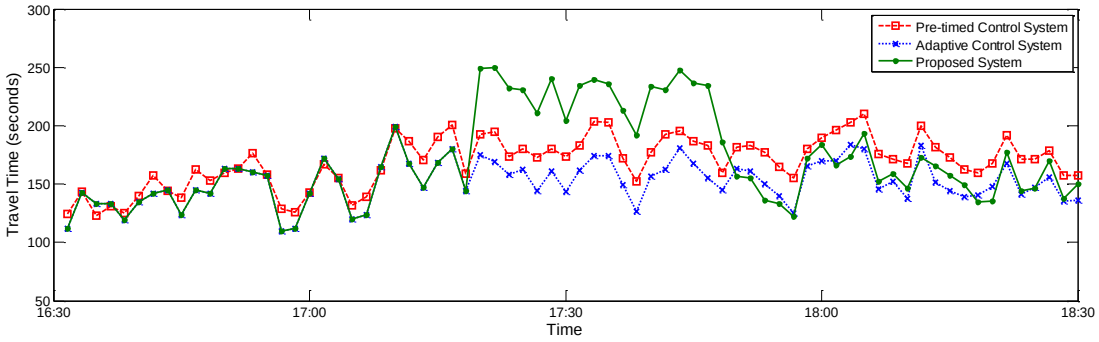


Figure 6.22 The time-dependent travel time along path 2

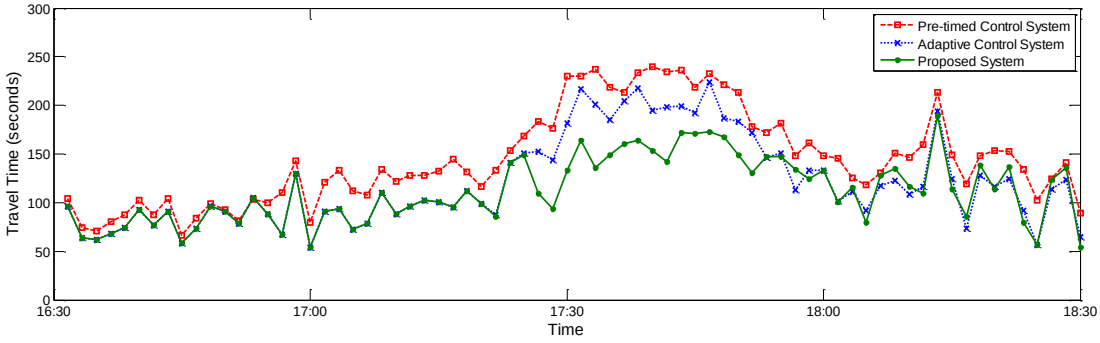


Figure 6.23 The time-dependent travel time along path 3

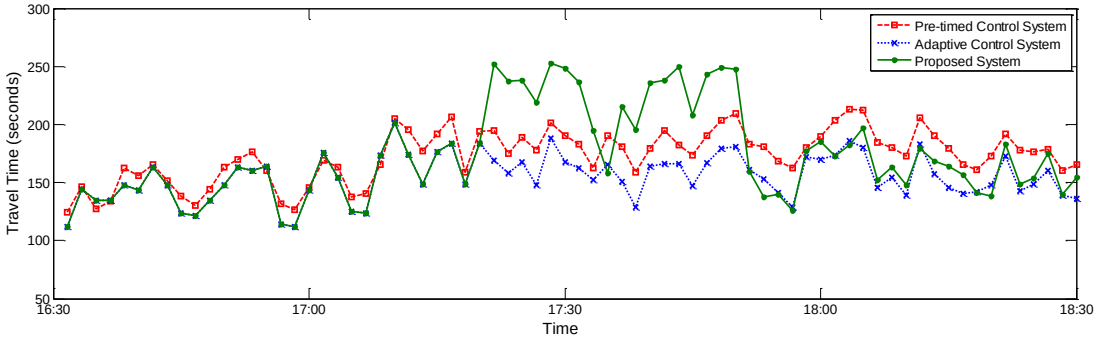


Figure 6.24 The time-dependent travel time along path 4

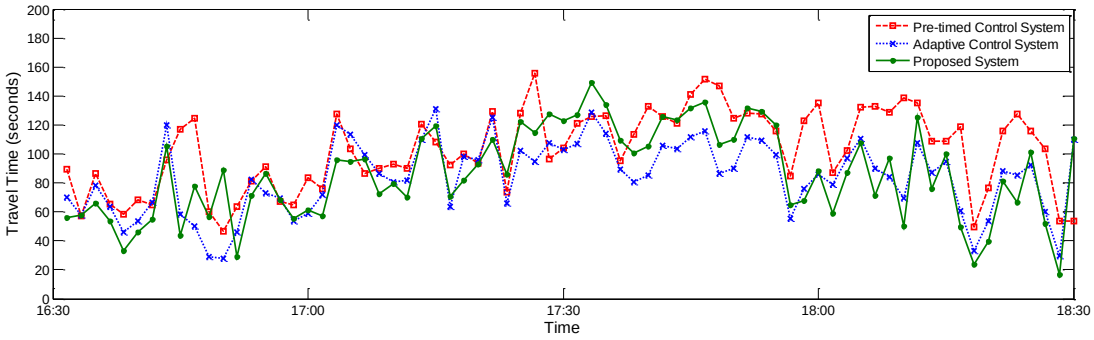


Figure 6.25 The time-dependent travel time along path 5

In conclusion, the adaptive control systems, as expected, can efficiently outperform the pre-timed system in reducing travel times for vehicles along most critical paths. Moreover, the proposed integrated system can successfully prevent the off-ramp queue spillover to the freeway mainline with activation of the priority control function. To further evaluate the systems' operational efficiency over the entire network with different MOEs, Table 6.3 summarizes the resulting average delays, average number of stops, and average speeds.

Notably, the proposed system can produce the lowest network delay and the average number of stops, compared with the other two systems. Regarding the average speed, the proposed system can also yield significant improvement. However, one could also notice that the off-ramp priority control may result in some negative impacts to the local traffic. Hence, further comparisons between the average freeway delay and arterial delay under those three control systems are provided in Table 6.3. Based on the simulation results, it is noticeable that the adaptive system can outperform the pre-timed system on delay reduction. However, the proposed

system can significantly reduce the travel delay on freeway mainline with the slightly increase of delay on local traffic.

Table 6.3 Network performance with different control systems

Performance Index	Pre-timed System	Adaptive System	Proposed System
Average number of stops	2.391	1.711 (-28.4%)	1.621 (-32.2%)
Average speed (km/h)	36.116	38.633 (+7.0%)	39.25 (+8.7%)
Average Network delay (s)	89.065	73.77 (-13.7%)	68.209 (-19.6%)

6.6 Closure

This chapter has illustrated an integrated real-time control system which includes three primary modules: off-ramp queue estimation, arterial adaptive signal control and freeway off-ramp priority control. Different from the conventional adaptive control systems, the proposed system would firstly estimate the queue length on the target off-ramp at the beginning of each signal cycle. Then, the priority control module will be activated to clear the queuing vehicles on the off-ramp when it predicting potential freeway breakdown caused by off-ramp queue spillover. To evaluate the effectiveness of the proposed system, this study has conducted numerical studies on a real-world network comprising both freeway mainline and local arterial, with a well-calibrated simulation platform. The experimental results reveal that the overall network performance has been significantly improved with the proposed

control system. Further evaluation of the freeways travel time distribution has also demonstrated the effectiveness of the proposed system on preventing off-ramp queue spillover.

Chapter 7 : Conclusions

7.1 Research Summary and Contributions

To prevent traffic breakdowns on a freeway mainline due to the off-ramp queue spillover, this dissertation has proposed a three-stage integrated control system that consists of three modules for Origin-Destination (O-D) estimation, pre-timed signal optimization, and real-time signal adaptive control. The objective of the O-D estimation is to first identify the traffic demand pattern at the off-ramp and local arterial within the impact area. Ground on the identified paths of heavy traffic flows, the proposed integrated system employs two sequential models for design of signal plans on the local arterial. The final stage of the system's control actions is to execute its embedded real-time control function to provide signal priority control to the off-ramp flows when potential queue spillback to the freeway mainline has been detected.

In summary, this dissertation has made the following contributions to contend with daily recurrent congestion in urban networks:

- Development of an effective operational structure for integrated traffic control for arterials and freeways plagued by the excessive off-ramp queue spillover during the peak period. With the embedded three-stage control function, the proposed integrated system is able to estimate the demand pattern at the target corridor, and design the base signal plan using its signal optimization and multi-path progression models. The system can further deal with the traffic fluctuation by dynamically adjusting the signal plan at each intersection with its real-time control module.

- Construction of three O-D estimation models based on the data availability to identify the critical paths at the off-ramp and on the connected local arterials. Model-I is to use link accounts as the primary measurements, and Model-II directly takes turning flows at each intersection as its primary input. To take advantage of the newly available information, this dissertation has further proposed Model-III, that is, an enhanced model to incorporate the time-varying queue patterns in the model formulations and key parameter estimation. The results of extensive numerical tests have revealed that Model-III can clearly outperform other models in the literature on identifying the critical traffic paths.

- Formulation of a signal optimization model to optimize the green splits at all target intersections and their common cycle length. With the objective of maximizing intersection capacity, this model, formulated with the linear programming, can ensure the optimality of the resulting signal plans. In addition, to prevent the formation of excessive off-ramp queue, this model has further incorporated a queue constraint to ensure that the off-ram queue length stays below a pre-determined threshold.

- Development of three innovative multi-path progression models based on the traffic flow patterns to facilitate path-flows to reach their destinations. The first proposed model is a direct extension of MAXBAND under a predetermined phasing plan, but using the path-flow data to yield the progression band for each identified path flow. The second model further takes the phase sequence at each intersection as a decision variable, and then

concurrently optimizes the signal plans with offsets for the entire arterial. Due to the competing nature of multi-path progression flows over the same green duration, the second model has been enhanced with a function to automatically select the optimal number of paths and then to maximize their progression bandwidths. The results of extensive numerical investigation with field data have demonstrated that the proposed models can clearly outperform the conventional design methods, such as with MAXBAND or TRANSYT, especially for those arterials experiencing heavy path-flows from-and-to the freeway, and accommodating both left- and right-turns maneuvers from short turning bays.

- Advancement of all key control models for real-time operations. The proposed real-time system includes three primary modules: off-ramp queue length estimation, arterial signal adaptive signal control, and off-ramp signal priority control. Using the flow data obtained from dual-zone detectors, the proposed real-time system can reliably estimate the time-dependent queue length at an off-ramp and subsequently identify whether or not a potential queue spillover will occur. If no spillover is predicted, its arterial adaptive control module will function to dynamically adjust the intersection signal timings and offsets so as to reduce the resulting delay and provide progression to those paths of heavy flows. Otherwise, the off-ramp priority control module will be activated to offer a green extension and progression priority to those off-ramp flows.

7.2 Future Research

Despite the effectiveness of the integrated control system developed in this dissertation, some critical issues remain to be solved. Future research directions along this study are listed below:

Development of an optimal traffic control model to concurrently account for the delay of traffic flows on the freeway and local arterial

The system proposed in this dissertation has focused on the signal optimal control at a local arterial but with a priority control to the off-ramp flows. When a potential queue spillover is predicted, a green time extension and progression priority will be provided to the off-ramp flows. Such control logic has demonstrated its effectiveness in preventing the formation of freeway bottleneck caused by ramp queues. However, depending on the congestion level on the local arterial, some level of off-ramp queue spillback to the freeway mainline may be inevitable. As such, the objective of control shall consider the equity between the freeway and arterial traffic flows.

Hence, it is essential to develop the following models for potential real-world applications: 1) freeway capacity-drop model to estimate the impact by off-ramp queue spillover; 2) total travel delay model to compute those traveling delays on both the freeway mainline and local arterial; and 3) optimization model that incorporates the equity between freeway and arterial flows on minimizing the total travel delay over the entire network considering the impact of a freeway off-ramp queue spillover.

Integration of both on-ramp and off-ramp control strategies for a large-scale corridor traffic management

This research mainly focused on the traffic control at the interchange network connected to the off-ramp. However, congestion on the freeway mainline may be caused by a variety of factors. For instance, aside from the potential queue spillover issue from the off-ramp, the flows via an on-ramp may also result in a bottleneck on the mainline and queue blockage to the surface street. Moreover, when the distance between an on-ramp and its neighboring off-ramp is relatively short, the weavings between on-ramp flows, mainline flows, and off-ramp flows may also cause reduction of the freeway capacity during the peak hours.

Hence, to effectively deal with other issues contributing to traffic congestion in urban networks, one shall further extend this study to a large-scale level, and such a system shall be capable of effectively coordinating Ramp Metering (RM), Variable Speed Limit (VSL), and off-ramp priority control strategies.

Enhancement of the current real-time signal control system with advanced information/communication technologies

The proposed adaptive signal control system relies mostly on the data from traffic sensors, which often suffer from the following limitations: 1) those sensors are point detectors, and provide only instantaneous vehicle locations when a vehicle is passing over the detection zone; 2) the installation and maintenance cost of the detection system is relatively high for on-line applications; and 3) the malfunction of one or more sensors may significant degrade the effectiveness of the control system.

With the recent advancement in wireless technologies, connected vehicles are likely to be popular in the near future. Such vehicles are able to communicate with each other (V2V) and with the infrastructure (V2I). Data from connected vehicles shall include real-time vehicle location, speed, acceleration or deceleration rate, and the information associated with nearby vehicles. Hence, how to best use the available information from connected vehicles so as to minimize the dependence of infrastructure-based sensors will be a challenging task for efficient and effective update of the traffic control plans in real-time operations.

Bibliography

Aboudolas, K., Papageorgiou, M., Kouvelas, A., and Kosmatopoulos, E., (2010), A rolling-horizon quadratic-programming approach to the signal control problem in large-scale congested urban road networks, Transportation Research Part C: Emerging Technologies, 18(5), 680-694.

Allsop, R.E., (1971), Delay-minimizing settings for fixed-time traffic signals at a single road junction, IMA Journal of Applied Mathematics, 8 (2): 164-185.

Allsop, R.E., (1972), Delay at a fixed time traffic signal-I: Theoretical analysis, Transportation Science, 6(3), 260-285.

Allsop, R.E., (1976), Sigcap: A computer program for assessing the traffic capacity of signal controlled road junctions, Traffic Engineering and Control, vol. 17, 338-341.

Antoniou, C., Ben-Akiva, M. E. and Koutsopoulos, H. N., (2006), Dynamic traffic demand prediction using conventional and emerging data sources, in IEEE Proceedings-Intelligent Transport Systems, pp. 97-104.

Ashok, K. and Ben-Akiva, M. E., (2000), Alternative approaches for real-Time estimation and prediction of time-Dependent origin–destination flows, Transportation Science, vol. 34, pp. 21-36.

Ashok, K. and Ben-Akiva, M. E., (2002), Estimation and prediction of time-dependent origin-destination flows with a stochastic mapping to path flows and link flows, Transportation Science, vol. 36, pp. 184-198.

Ashok, K., Ben-akiva, M., (1993). Dynamic origin–destination matrix estimation for real-time traffic management systems. Transportation and Traffic Theory, 465–484.

Bell, M. G., (1983). The estimation of an origin-destination matrix from traffic counts, Transportation Science, vol. 17, pp. 198-217.

Bell, M. G., (1991), The real time estimation of origin-destination flows in the presence of platoon dispersion, Transportation Research Part B: Methodological, vol. 25, pp. 115-125.

Boillot, F., Blosseville, J.M., Lesort, J. B., Motyka, V., Papageorgiou, M., and Sellam, S., (1992), Optimal signal control of urban traffic networks, In Proceedings of the 6th IEEE International Conference on Road Traffic Monitoring and Control, 75–79.

Bretherton, D., Bodger, M., and Cowling, J., (2005), SCOOT-managing congestion, communications and control, 12th World Congress on Intelligent Transport Systems.

Burrow, I. J., (1987), OSCADY: a computer program to model capacities, queues and delays at isolated traffic signal junctions, No. RR 105.

Cai, C., Wong, C. K., and Heydecker, B. G., (2009), Adaptive traffic signal control using approximate dynamic programming, Transportation Research Part C: Emerging Technologies, 17(5), 456-474.

Calabrese F., Di Lorenzo, G. Liu, L., and Ratti, C., (2011), Estimating origin-destination flows using mobile phone location data, IEEE Pervasive Computing, vol. 10, pp. 36-44.

Cao P., Miwa, T., Yamamoto, T., and Morikawa, T., (2013), Bilevel generalized least squares estimation of dynamic origin-destination matrix for urban network with probe vehicle data, Transportation Research Record: Journal of the Transportation Research Board, vol. 2333, pp. 66-73.

Cascetta E. and Nguyen, S., (1988), A unified framework for estimating or updating origin/destination matrices from traffic counts, Transportation Research Part B: Methodological, vol. 22, pp. 437-455.

Cascetta E., (1984), Estimation of trip matrices from traffic counts and survey data: a generalized least squares estimator. Transportation Research Part B: Methodological, vol. 18, pp. 289-299.

Cascetta E., Papola, A., Marzano, V., Simonelli, F., and Vitiello, I., (2013), Quasi-dynamic estimation of o–d flows from traffic counts: Formulation, statistical validation and performance analysis on real data, Transportation Research Part B: Methodological, vol. 55, pp. 171-187.

Cascetta, E., Inaudi, D., and Marquis, G., (1993), Dynamic estimators of origin-destination matrices using traffic counts, Transportation science, vol. 27, pp 363-373.

Cassidy, M. J., Anani, S. B., and Haigwood, J. M., (2002), Study of freeway traffic near an off-ramp, Transportation Research Part A: Policy and Practice, 36(6), 563-572.

Castillo E., Menéndez, J. M., and Jiménez, P., (2008), Trip matrix and path flow reconstruction and estimation based on plate scanning and link

observations, Transportation Research Part B: Methodological, vol. 42, pp. 455-481.

Chang, E. C., Cohen, S. L., Liu, C., Chaudhary, N. A., and Messer, C., (1988), MAXBAND-86: Program for optimizing left-turn phase sequence in multiarterial closed networks, Transportation Research Record, 1181, 61-67.

Chang, G. L. and Tao X., (1998). Estimation of time-dependent turning fractions at signalized intersections. Journal of Transportation Research Board: Transportation Research Record, 1644, 142-149.

Chang, G. L. and Wu, J., (1994), Recursive estimation of time-varying origin-destination flows from traffic counts in freeway corridors, Transportation Research Part B: Methodological, vol. 28, pp. 141-160.

Chang, G. L., Ho, P. K., and Wei, C. H., (1993), A dynamic system-optimum control model for commuting traffic corridors, Transportation Research Part C, 3-22.

Chang, G.-L., and Tao, X., (1996). Estimation of dynamic O–Ds for urban networks. Transportation and Traffic Flow Theory 13, 1–20.

Chang, G.-L., and Tao, X., (1999). An integrated model for estimating time-varying network origin–destination distributions. Transportation Research A 33 (5), 381–399

Chang, T. H., and Lin, J. T., (2000), Optimal signal timing for an oversaturated intersection, Transportation Research Part B: Methodological, 34(6), 471-491.

Chaudhary, N.A., Kovvali, V.G., Chu, C.-L., Kim, J., and Alam, S.M., (2002), Software for timing signalized arterials, Report FHWA/TX-03/4020-1, Texas

Transportation Institute, the Texas A&M University System, College Station, Texas, September.

Chen R., Sun, J., and Feng, Y., (2011), A novel OD estimation method based on automatic vehicle identification data, in Intelligent Computing and Information Science. vol. 135, ed: Springer Berlin Heidelberg, pp. 461-470.

Cornwell, P. R., Luk, J. Y. K., and Negus, B. J., (1986), Tram priority in SCATS, Traffic Engineering and Control, Vol 27, Issue NO. 11.

Courage K. G., Luh, J. Z., and Wallace, C. E., (1989), Development of guidelines for implementing computerized timing designs at traffic actuated signals, Two Volumes on Arterial System Implementation, Transp. Res. Center, Gainesville, FL.

Cremer, M., and Keller, H., (1981), Dynamic identification of O–D flows from traffic counts at complex intersections, In: The Proceeding of the 8th International Symposium on Transportation and Traffic Theory.

Cremer, M., and Keller, H., (1984), A systems dynamics approach to the estimation of entry and exit O–D flows, In: The Proceeding of the 9th International Symposium on Transportation and Traffic Theory, pp. 431–450.

Cremer, M., and Schoof, S., (1989), On control strategies for urban traffic corridors, In Proceedings of IFAC Control, Computers, Communications in Transportation, Paris.

Cremer, M., Keller, H., (1987), A new class of dynamic methods for the identification of origin–destination flows, Transportation Research B 21 (2), 117–132.

D'Ans, G.C., and Gazis, D.C., (1976), Optimal control of oversaturated store- and forward transportation networks, Transportation Science, 10, 1–19.

Daganzo, C. F., (1994), The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory, Transportation Research Part B: Methodological, 28(4), 269-287.

Daganzo, C. F., Cassidy, M. J., and Bertini, R. L., (1999), Possible explanations of phase transitions in highway traffic, Transportation Research Part A: Policy and Practice, 33(5), 365-379.

Daganzo, C. F., Laval, J., and Muñoz, J. C., (2002), Some ideas for freeway congestion mitigation with advanced technologies, Traffic Engineering and Control 43 (10), 397-403.

Day, I., Ag, S., and Whitelock, R., (1998), SCOOT-split, cycle & offset optimization technique, In TRB Mid-Year Meeting and Adaptive Traffic Signal Control Workshop, vol. 7.

Di, X., Zhang, X., Zhang, H. M., and Liu, H. X., (2013), Application of pavement marker to avoid queue-jumping and traffic spillback at off-ramp of expressways, In Transportation Research Board 92nd Annual Meeting, No. 13-3298.

Dixon, M. P. and Rilett, L., (2002), Real-time OD estimation using automatic Vehicle identification and traffic count data, Computer-Aided Civil and Infrastructure Engineering, vol. 17, pp. 7-21.

Dixon, M. P. and Rilett, L., (2005), Population origin–destination estimation using automatic vehicle identification and volume data, Journal of Transportation Engineering, vol. 131, pp. 75-82.

Eisenman, S. M., and List, G. F., (2004), Using probe data to estimate OD matrices, Proceedings of the 7th International IEEE Conference on, pp. 291-296.

Frederix R., Viti, F., Corthout, R., and Tampère, C. M., (2011), New gradient approximation method for dynamic origin-destination matrix estimation on congested networks, Transportation Research Record: Journal of the Transportation Research Board, vol. 2263, pp. 19-25.

Gartner, N. H., (1982), Development and Testing of a demand-responsive strategy for traffic signal control, American Control Conference.

Gartner, N. H., (1983), OPAC: A demand-responsive strategy for traffic signal control, Transportation Research Record 906, pp. 75–81.

Gartner, N. H., and Stamatiadis, C., (2002), Arterial-based control of traffic flow in urban grid networks, Mathematical and Computer Modelling, 35(5), pp. 657-671.

Gartner, N. H., and Stamatiadis, C., (2004), Progression optimization featuring arterial-and route-based priority signal networks, Journal of Intelligent Transportation Systems, 8(2), pp. 77-86.

Gartner, N. H., Assman, S.F., Lasaga, F., and Hou D.L., (1991), A multi-band approach to arterial traffic signal optimization, Transportation Research Part B, 25(1), pp. 55-74.

Gartner, N. H., Pooran, F. J., and Andrews, C. M., (2001), Implementation of the OPAC adaptive control strategy in a traffic signal network, In Proceedings of the 2001 IEEE Intelligent Transportation Systems Conference. Oakland, California, pp. 195-200.

Gartner, N. H., Pooran, F. J., and Andrews, C. M., (2002), Optimized policies for adaptive control strategy in real-time traffic adaptive control systems: Implementation and field testing, Transportation Research Record: Journal of the Transportation Research Board, 1811(1), 148-156.

Gartner, N. H., Stamatiadis, C., and Tarnoff., P. J., (1995), Development of advanced traffic signal control strategies for ITS: a multilevel design, Transportation Research Record, 1494, pp. 98–105.

Gazis, D. C., (1964), Optimum control of a system of oversaturated intersections, Operations Research, 12(6), 815-831.

Günther, G., Coeymans, J. E., Muñoz, J. C., and Herrera, J. C., (2012), Mitigating freeway off-ramp congestion: A surface streets coordinated approach, Transportation research part C: emerging technologies, 20(1), 112-125.

Haddad, J., Ramezani, M., and Geroliminis, N., (2013), Cooperative traffic control of a mixed network with two urban regions and a freeway, Transportation Research Part B: Methodological, 54, 17-36.

Hadi, M. A., and Wallace, C. E., (1993), Hybrid genetic algorithm to optimize signal phasing and timing, Transportation Research Record, 1421, 104-112.

Hagen, L., Lin, P. S., and Fabregas, A. D., (2006), A Toolbox for Reducing Queues at Freeway Off-Ramps, Center for Urban Transportation Research, Florida Department of Transportation, University of South Florida. Report BD544-10.

Hansen, B.G., Martin, P. T. and Perrin, H. J., (2000), SCOOT real-time adaptive control in a CORSIM simulation environment, Transportation Research Record 1727, pp. 27-31.

Hazelton, M., (2000). Estimation of origin–destination matrices from link flows on uncongested networks. Transportation Research B 34(7), 549–566.

Henry, J. J., Farges, J. L., and Tuffal, J., (1983), The PRODYN real time traffic algorithm, Proceedings of the fourth IFAC-IFIP-IFORS conference on Control in Transportation Systems, pp. 307–311.

Hunt, P. B., Robertson, D. I., Bretherton, R. D., and Royle, M. C., (1982), The SCOOT on-line traffic signal optimisation technique, Traffic Engineering and Control, vol. 23, Issue NO. 4.

Iqbal, M. S., Choudhury,C. F., Wang, P. and Gonz á lez, M. C., (2014), Development of origin–destination matrices using mobile phone call data, Transportation Research Part C: Emerging Technologies, vol. 40, pp. 63-74.

Jia, B., Jiang, R., and Wu, Q. S., (2004), Traffic behavior near an off ramp in the cellular automaton traffic model, Physical Review E, 69(5), 056105.

Kashani, H. R., and Saridis, G.N., (1983), Intelligent control for urban traffic systems. Automatica 19, 191–197.

Kell J. H. and Fullerton, I. J., (1998), Manual of traffic signal design, 3rded. Englewood Cliffs, NJ: Institute of Transportation Engineers, Prentice-Hall.

Keller, H., Ploss, G., (1987). Real-time identification of O–D network flow from counts for urban traffic control. Transportation and Traffic Theory, 267–284.

Kim J. T. and Courage, K. G., (2003), Evaluation and design of maximum green time settings for traffic actuated control, Transportation Research Record, vol. 1852, pp. 246–255.

Lan, C.J., (2004), New optimal cycle length formulation for pre-timed signals at isolated intersections. Journal of Transportation Engineering, 130(5), 637–647.

Li, J. Q., (2014), Bandwidth Synchronization under Progression Time Uncertainty, IEEE Transactions on ITS, 15(2), 1-11.

Li, Z., (2012), Modeling arterial signal optimization with enhanced cell transmission formulations, Journal of Transportation Engineering, 137(7), 445–454.

Li, Z., Chang, G.L., and Natarajan, S., (2009), An integrated off-ramp control model for freeway traffic management, In: 15th World Congress on ITS. New York.

Lim, K., Kim, J. H., Shin, E., and Kim, D. G., (2011), A signal control model integrating arterial intersections and freeway off-ramps, KSCE Journal of Civil Engineering, 15(2), 385-394.

Lin, F. B., (1985), Optimal timing settings and detector lengths of presence model full-actuated control, Transportation Research Record, vol. 1010, pp. 37–45.

Lin, P. W., and Chang, G. L., (2007), A generalized model and solution algorithm for estimation of the dynamic freeway O-D distribution. Transportation Research Part B, Vol. 41, No. 5, pp. 554-572.

Lin, P. W., Chang, G-L., (2005). A robust model for estimating freeway dynamic origin–destination matrix. Transportation Research Record 1923, 110–118.

Lin, P. W., and Chang, G. L., (2006), Modeling the measurement errors and the missing initial O-D set in a freeway dynamic O-D estimation system, Transportation Research Part C, Vol. 14, No. 6, pp. 384-402.

Little, J. D. C., (1966), The synchronization of traffic signals by mixed-integer linear programming, Operation Research, 14(4), pp. 568-594.

Little, J. D. C., Kelson, M. D., and Gartner, N. H., (1981), MAXBAND: A program for setting signals on arteries and triangular networks, Transportation Research Record, 795, pp. 40-46.

Liu, Y., and Chang, G. L., (2011), An arterial signal optimization model for intersections experiencing queue spillback and lane blockage, Transportation research part C: emerging technologies, 19(1), 130-144.

Lo, H., (1999), A novel traffic signal control formulation, Transportation Research, Part A, 44, 436–448.

Lo, H., (2001), A cell-based traffic control formulation: strategies and benefits of dynamic timing plan, Transportation Science, 35, 148–164.

Lo, H., Chang, E., Chan. Y. C., (2001), Dynamic network traffic control. Transportation Research, Part A, 35, 721–744.

Lou, Y., and Yin, Y. (2010). A decomposition scheme for estimating dynamic origin–destination flows on actuation-controlled signalized arterials. Transportation Research Part C: Emerging Technologies, 18(5), 643-655.

Lovell, D. J., (1997), Traffic control on metered networks without route choice, Report No. UCB-ITS-DS-97-3.

Lowrie, P R., (1990), SCATS, Sydney Co-Ordinated Adaptive Traffic system: a traffic responsive method of controlling urban traffic, Roads and Traffic Authority NSW, Darlinghurst, NSW Australia.

Lu, C. C., Zhou, X., and Zhang, K., (2013), Dynamic origin–destination demand flow estimation under congested traffic conditions, Transportation Research Part C: Emerging Technologies, 34, 16-37.

Lu, X., Su, D., and Spring, J., (2013), Coordination of Freeway Ramp Meters and Arterial Traffic Signals Field Operational Test, Report CA14-2223.

MacDougall, M.H., (1987). Simulating Computer Systems. MIT Press, Cambridge, MA.

Maher, M., (1983), Inferences on trip matrices from observations on link volumes: a Bayesian statistical approach, Transportation Research Part B: Methodological, vol. 17, pp. 435-447.

Manual, H.C., (2000), Highway Capacity Manual, Transportation Research Board, Washington, DC, fourth edition, ISBN 0-309-06681-6.

Marzano V., Papola, A., and Simonelli, F., (2009), Limits and perspectives of effective O–D matrix correction using traffic counts, Transportation Research Part C: Emerging Technologies, vol. 17, pp. 120-132.

Matson, T. M., (1955), Traffic engineering. McGraw-Hill.

Mauro, V., and Di Taranto, C., (1989), UTOPIA. In: CCCT'89 – AFCET Proceedings, Paris.

Messer, C. J., (1998), Simulation studies of traffic operations at oversaturated, closely spaced signalized intersections. Transportation Research Record: Journal of the Transportation Research Board, 1646(1), 115-123.

Michalopoulos, P. G., and Stephanopoulos, G., (1977a), Oversaturated signal systems with queue length constraints—I: Single intersection, Transportation Research, 11(6), 413-421.

Michalopoulos, P. G., and Stephanopoulos, G., (1977b), Oversaturated signal systems with queue length constraints—II: Systems of intersections, Transportation Research, 11(6), 423-428.

Miller, A.J., (1963), Settings for fixed cycle traffic signals, Operations Research Quarterly, Vol. 14, No. 4, pp. 373-386.

Mirchandani, P. B., and Lucas, D. E., (2004), Integrated transit priority and rail/emergency preemption in real-time traffic adaptive signal control, Journal of Intelligent Transportation Systems, 8(2), 101-115.

Mirchandani, P. B., Head, K. L., Knyazyan, A. and Wu, W., (2000), An approach towards the integration of bus priority, traffic adaptive signal control and bus information/scheduling system, 8th International Conference on Computer-Aided Scheduling of Public Transportation at Berlin, Germany.

Mirchandani, P., and L. Head., (2001), A real-time traffic signal control system: architecture, algorithms, and analysis, Transportation Research Part C, Vol. 9, No. 6, pp. 415-432.

Morgan, J. T. and Little, J.D.C., (1964), Synchronizing traffic signals for maximal bandwidth, Operation Research, 12(6), pp. 896-912.

Muñoz, J. C., and Daganzo, C. F., (2002), The bottleneck mechanism of a freeway diverge, Transportation Research Part A: Policy and Practice, 36(6), 483-505.

Newell, G. F., (1999), Delays caused by a queue at a freeway exit ramp, Transportation Research Part B: Methodological, 33(5), 337-350.

Nihan, N. L., and G. A. Davis, (1987), Recursive estimation of origin-destination matrices from input/output counts, Transportation Research Part B: Methodological, vol. 21, pp. 149-163.

Okutani I., and Stephanedes, Y. J., (1984), Dynamic prediction of traffic volume through Kalman filtering theory, Transportation Research Part B: Methodological, vol. 18, pp. 1-11.

Orcutt F. L.,(1993), The traffic signal book. Englewood Cliffs, NJ: PrenticeHall.

Papageorgiou, M., (1995), An integrated control approach for traffic corridors, Transportation Research Part C 3, 19–30.

Papageorgiou, M., and Kotsialos, A., (2000), Freeway ramp metering: An overview. Proceedings of IEEE in Intelligent Transportation Systems, pp. 228-239.

Papageorgiou, M., Hadj-Salem, H., and Blosseville, J. M., (1991), ALINEA: A local feedback control law for on-ramp metering, Transportation Research Record, (1320).

Papamichail, I., Kotsialos, A., Margonis, I., and Papageorgiou, M., (2010), Coordinated ramp metering for freeway networks–A model-predictive hierarchical control approach. Transportation Research Part C: Emerging Technologies, 18(3), 311-331.

Park, B., Messer, C.J., and Urbanik, T., (1999), Traffic signal optimization program for oversaturated conditions: genetic algorithm approach. Transportation Research Record, 1683, 133–142.

Parry K. and Hazelton, M. L., (2012), Estimation of origin–destination matrices from link counts and sporadic routing data, Transportation Research Part B: Methodological, vol. 46, pp. 175-188.

Payne, H. J., (1971), Models of freeway traffic and control, Mathematical models of public systems.

Pei, Y., and Zhou, K., (2013), Off-ramp control near surface road, Research Journal of Applied Sciences, Engineering and Technology 5(8): 2612-2615.

Pooran, F. J., and Lieu, H. C., (1994), Evaluation of system operating strategies for ramp metering and traffic signal coordination, In Moving Toward Deployment. Proceedings of the IVHS America Annual Meeting. vol. 2 .

Robertson, D. I., (1969), TRANSYT: a traffic network study tool, Report Number: TRRL-LR-253, Transport and Road Research Laboratory.

Robertson, D. I., and Bretherton, R. D., (1991), Optimizing networks of traffic signals in real time: The SCOOT method, IEEE Transactions on Vehicular Technology. vol. 40 no. 1.

Rudjanakanoknad, J., (2012), Capacity change mechanism of a diverge bottleneck, Transportation Research Record: Journal of the Transportation Research Board, (2278), 21-30.

Sherali, H.D., and Park, T., (2001), Estimation of dynamic origin–destination trip tables for a general network. Transportation Research B 35, 217–235.

Silcock, J. P., (1997), Designing signal-controlled junctions for group-based operation, Transportation Research Part A: Policy and Practice, 31(2), 157-173.

Sims, A. G., and Finlay, A. B., (1984), SCATS, splits and offsets simplified (SOS), Australian Road Research, 12(4).

Sohn, K. and Kim, D., (2008), Dynamic origin–destination flow estimation using cellular communication system, Vehicular Technology, IEEE Transactions on, vol. 57, pp. 2703-2713.

Spiess, H., (1987), A maximum likelihood model for estimating origin-destination matrices, Transportation Research Part B: Methodological, vol. 21, pp. 395-412.

Spiliopoulou, A., Kontorinaki, M., Papamichail, I., and Papageorgiou, M., (2013), Real-time route diversion control at congested motorway off-ramp areas-Part I: User-optimum route guidance, In 16th International IEEE Conference on Intelligent Transportation Systems-(ITSC), pp. 2119-2125.

Stamatiadis, C., and Gartner, N. H., (1996), MULTIBAND-96: a program for variable-bandwidth progression optimization of multiarterial traffic networks, Transportation Research Record, 1554(1), 9-17.

Stevanovic, A., Martin, P.T., and Stevanovic, J., (2007), VisSim-based genetic algorithm optimization of signal timings, Transportation Research Record, 2035, 59–68.

Tian, Z. and Urbanik, T., (2007), System partition technique to improve signal coordination and traffic progression, Journal of transportation engineering, 133(2), 119-128.

Tian, Z., Balke, K., Engelbrecht, R., and Rilett, L., (2002), Integrated control strategies for surface street and freeway systems, Transportation Research Record: Journal of the Transportation Research Board, 1811(1), 92-99.

van Aerde M., Hellinga,B., Yu, L. and Rakha, H., (1993), Vehicle probes as real-time ATMS sources of dynamic OD and travel time data, Preprint, Queen's University, Department Of Civil Engineering.

van den Berg, M., De Schutter, B., and Hellendoorn, H., (2006), Effects of on-ramp and off-ramp metering on queue forming in urban traffic networks, In Proceedings of the 11th IFAC Symposium on Control in Transportation Systems (Vol. 1302135).

van den Berg, M., De Schutter, B., Hegyi, A., and Hellendoorn, J., (2004), Model predictive control for mixed urban and freeway networks. 83rd Annual Meeting of the Transportation Research Board, vol. 19.

Wallace, C.E., Courage, K.G., Reaves, D.P., Shoene, G.W., Euler, G.W., and Wilbur, A., (1988), TRANSYT-7F User's Manual: Technical Report, University of Florida, Gainesville, FL.

Webster, F. V., (1956), Traffic signal settings, Road Research Technical Paper, No. 39, HMSO.

Willumsen, L.G., (1984). Estimation time-dependent trip matrices from traffic counts. In: Proceeding of the 9th international symposium on transportation and traffic theory, pp. 397–411.

Wong, C. K., and Wong, S. C., (2003), Lane-based optimization of signal timings for isolated junctions, Transportation Research Part B: Methodological, 37(1), 63-84.

Wu, J. and Chang, G. L., (1999), Heuristic method for optimal diversion control in freeway corridors, Transportation Research Record, 1667, 8-15.

Wu, J., (1997). A real-time origin–destination matrix updating algorithm for on-line application. Transportation Research Part B 31 (5), 381–396.

Wu, J., and Chang, G.L., (1996). Estimation of time-varying origin–destination distributions with dynamic screenline flows. Transportation Research B 30, 277–290.

Yang, H., (1995), Heuristic algorithms for the bilevel origin-destination matrix estimation problem, Transportation Research Part B: Methodological, 29(4), 231-242.

Yang, H., Meng, Q., Bell, M.G.H., (2001). Simultaneous estimation of the origin–destination matrices and travel-cost coefficient for congested networks in a stochastic user equilibrium. Transportation Science 35, 107–123.

Yang, H., Sasaki, T., Iida, Y., and Asakura, Y., (1992), Estimation of origin-destination matrices from link traffic counts on congested networks, Transportation Research Part B: Methodological, vol. 26, pp. 417-434.

Yang, X., Chang, G. L., and Rahwanji, S., (2014), Development of a Signal Optimization Model for Diverging Diamond Interchange, Journal of Transportation Engineering, 140(5), 04014010.

Yang, X., Lu, Y. C., and Chang, G. L., (2014), Dynamic Signal Priority Control Strategy to Mitigate the Off-ramp Queue Spillback to the Freeway Mainline Segment, Transportation Research Record: Journal of the Transportation Research Board, vol. 2438, pp 1-11.

Yang, X., Lu, Y., and Lin, Y., (2013), Interval optimization for signal timings with time-dependent uncertain arrivals, Journal of Computing in Civil Engineering, in press.

Yin, Y., (2008), Robust optimal traffic signal timing, Transportation research Part B, 42, 911-924.

Yun, I., and Park, B., (2006), Application of stochastic optimization method for an urban corridor, Proceedings of the Winter Simulation Conference, pp. 1493–1499.

Zhang, G., and Wang, Y., (2011), Optimizing minimum and maximum green time settings for traffic actuated control at isolated intersections. Intelligent Transportation Systems, IEEE Transactions on, 12(1), 164-173.

Zhang, H., Ma, J., and Nie, Y., (2009), Local synchronization control scheme for congested interchange areas in freeway corridor, Transportation Research Record: Journal of the Transportation Research Board, (2128), 173-183.

Zhang, Y., and Hobeika, A., (1997), Diversion and signal re-timing for a corridor under incident conditions, 77th Annual Meeting of Transportation Research Board, Washington, D.C.

Zhou X. and Mahmassani, H. S., (2006), Dynamic origin-destination demand estimation using automatic vehicle identification data, Intelligent Transportation Systems, IEEE Transactions on, vol. 7, pp. 105-114.