

**OPTIMAL TRAFFIC CONTROL FOR A FREEWAY CORRIDOR
UNDER INCIDENT CONDITIONS**

by

Yunlong Zhang

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in
Civil Engineering

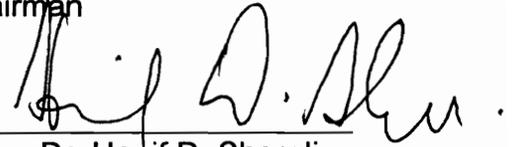
APPROVED:



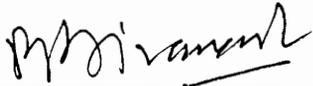
Dr. Antoine G. Hobeika, Chairman



Dr. Donald R. Drew



Dr. Hanif D. Sherali



Dr. R. Sivanandan



Dr. Brian D. Woerner

October, 1996

Blacksburg, Virginia

Keywords: corridor, diversion, signal, timing, optimization

c.2

LD
5655
V856
1996
Z466
c.2

OPTIMAL TRAFFIC CONTROL FOR A FREEWAY CORRIDOR UNDER INCIDENT CONDITIONS

by

Yunlong Zhang

A.G. Hobeika, Ph.D., Chairman

Civil Engineering

(ABSTRACT)

The non-recurring congestion, caused by incidents, is the main cause of traffic delays and causes up to 60 percent of the freeway delay in the United States. When severe incidents occur on freeways, capacity reduction due to lane blockages may cause an extremely high amount of traffic delay. In many cases, parallel surface arterials are available, and provide reasonably high speed and available capacity. In this scenario, to fully utilize the corridor capacity, diversion may be practical and necessary. With the changes of traffic demand levels and patterns on surface streets due to diversion, signal retiming for surface street intersections is necessary.

A nonlinear programming model was formulated to provide an integrated traffic control strategy for a freeway corridor under incident conditions. The objective function of the optimization model considers the interactions among the corridor components, and clearly reflects the primary goals of corridor traffic control under freeway incident conditions: to divert as much traffic away from the freeway as possible, not to over-congest the arterial and surface streets; and properly re-set the signal timing plans at all intersections to accommodate the changed traffic demand levels and patterns.

The gradient projection method is employed to solve diversion and signal retiming control measures simultaneously. By using a specifically developed simple and realistic traffic flow model and employing a sequential optimization approach, the computer program COROPT can obtain optimal traffic control strategies quickly and effectively. The COROPT program also has the flexibility to deal with various corridor configurations, different size of the corridor system, and different timing phasings. The model can address the time-varying factor of traffic flow, and can handle changing traffic and incident conditions over the time.

The model performance was evaluated and validated by running the simulation and optimization programs of TRANSYT-7F and INTEGRATION. It has been found that the proposed model and control strategy reduce the overall system delay, increase the throughput of the corridor, and thus improve the traffic conditions of the entire corridor.

ACKNOWLEDGMENTS

The author is indebted to Dr. Antonie Hobioka for his guidance, suggestions, and support throughout the research. He would also like to give thanks to Dr. Drew and Dr. Sherali for their advice on research, and for their instrumental courses from which the author benefited immensely. He would like to thank Dr. Siva and Dr. Woerner for serving on the committee, and for their helpful suggestions. He is also very thankful to Center for Transportation Research and Department of Civil Engineering for financial support.

The author would like to express his gratitude to his family, especially to his beloved wife Jing Cai, for the moral and financial support throughout his study, without which this work would not have been successful.

TABLE OF CONTENTS

1. INTRODUCTION	1
1.1 BACKGROUND	1
1.2 PROBLEM STATEMENT	3
1.3 RESEARCH OBJECTIVES	6
1.4 ORGANIZATION OF THE DISSERTATION	7
2. LITERATURE REVIEW	8
2.1 INTRODUCTION	8
2.2 VEHICLE AND TRAFFIC ACTUATED CONTROL MODELS AT AN ISOLATED INTERSECTION	9
2.3 OFF-LINE SURFACE STREET NETWORK SIGNAL OPTIMIZATION MODELS	17
2.3.1 TRANSYT Model	18
2.3.2 PASSER and MAXBAND	22
2.3.3 Comparison between TRANSYT-7F and PASSER II/MAXBAND	23
2.3.4 Other Off-Line Software and Studies	25
2.4 ON-LINE NETWORK OPTIMIZATION MODELS	27
2.4.1 UTCS	27
2.4.2 SCOOT and SCAT	29
2.4.3 OPAC, PRODYN and UTOPIA	32
2.4.4 Other On-line Systems	37
2.5 INTEGRATED CORRIDOR CONTROL	39
2.6 SUMMARY OF LITERATURE REVIEW	44
3. MODEL FORMULATION	45
3.1 INTRODUCTION	45
3.2 BACKGROUND	46

3.3 APPROACH	47
3.4 MODEL FORMULATION	52
3.4.1 Assumptions	52
3.4.2 Traffic Flow Model	53
3.4.2.1 Notations	54
3.4.2.2 Decision Variables	57
3.4.2.3 Signal Configuration at Arterial Intersections	57
3.4.2.4 Arrival	58
3.4.2.5 Departure	64
3.4.2.6 Queue	65
3.4.3 Objective Function	66
3.4.3.1 Objective Function Selection	66
3.4.3.2 Discussion on the Objective Function	67
3.4.3.3 Justifications of the Objective Function	68
3.4.4 Constraints	74
3.4.5 Measures of Effectiveness (MOEs)	76
3.4.5.1 Total Delay	76
3.4.5.2 Average Stop Delay	79
3.4.5.3 Throughput	80
4. ALGORITHM AND SOLUTION PROCEDURE	81
4.1 INTRODUCTION	81
4.2 MODEL CHARACTERISTICS	81
4.3 SOLUTION ALGORITHM	83
4.4 MULTI-LAYER SCHEME FOR CORRIDOR OPTIMAL CONTROL	86
4.4.1 Solution Procedures of The Corridor Optimal Control Model	86
4.4.2 Optimizing Splits and Diversion Rates	88
4.4.3 Offset Optimization	96
4.4.3.1 Platoon Dispersion Traffic Model Formulation	97
4.4.3.2 Offset and Phase Length Configuration Scenarios	103
4.4.3.3 Determining Optimal Offsets	106

4.4.4 Optimal Cycle Selection _____	109
5. COROPT: A CORRIDOR OPTIMIZATION COMPUTER PROGRAM _____	111
5.1 FUNCTIONS OF COROPT PROGRAM _____	111
5.2 FEATURES OF COROPT PROGRAM _____	112
5.3 DETERMINING PROGRAM PARAMETERS _____	117
5.3.1 Determining Optimization Time -- the Incident Duration _____	117
5.3.2 Determining Time Periods with Stable Prevailing Traffic and Incident Conditions _____	124
5.3.3 Determining Available Capacity for Incident Section _____	127
5.3.4 Determining Weights of Objective Function Components _____	127
5.4 RESTRICTION OF THE COROPT PROGRAM _____	134
6. CASE STUDY AND MODEL VALIDATION _____	136
6.1 INTRODUCTION _____	136
6.2 CASE STUDY _____	136
6.2.1 Case Study for Arterial with Different Number of Intersections Considered for Diversion Purpose _____	136
6.2.2 Case Study for Different Phasing Scenarios _____	145
6.2.3 Case Study for the Corridor with Different Demand Level and Incident Conditions _____	146
6.2.4 The Effect of Cycle Length on Corridor Traffic Conditions _____	148
6.3 MODEL VALIDATION _____	151
6.3.1 Test Example _____	151
6.3.2 Validation of Optimal Timing Plans _____	154
6.3.3 Evaluating the Effectiveness of the Optimal Diversion Strategy Combined with Arterial Retiming _____	154
7. CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE RESEARCH	158
7.1 CONCLUSIONS _____	158
7.2 RECOMMENDATIONS FOR FUTURE RESEARCH _____	160

LIST OF FIGURES

Figure 2-1. Information Flow at Stage n of OPAC System _____	24
Figure 2-2. Illustration of Rolling Horizon Concept _____	35
Figure 2-3. Section of Freeway with Sub-sections and Model Variables _____	39
Figure 3-1. Typical Lane Groups and Signal Phasing for an Arterial Intersection _____	59
Figure 3-2. Illustration of Diversion Routes and Traffic Movements _____	61
Figure 3-3. Relationship between Square of Queue and Freeway Delay _____	71
Figure 3-4. Relationship between Square of Queue and Delay at a Signal Intersection _____	72
Figure 3-5. Delay Calculation for Arterial Intersection Approach _____	78
Figure 4-1. Illustration of Gradient Projection _____	85
Figure 4-2. Procedures of the Corridor Optimization _____	89
Figure 4-3. Procedures for Optimizing Phase Splits and Diversion Rates _____	95
Figure 4-4. A Sample Scenario of Offset, Platoon Travel Time, and Phase Length for a Pair of Adjacent Intersections _____	99
Figure 4-5. Arrival/Departure Patterns for Different Phase Lengths _____	104
Figure 4-6. Arrival/Departure Patterns for Different Offsets _____	105
Figure 4-7. The Procedure for Optimizing Signal Offsets _____	108
Figure 5-1. Illustration of Intersection/Ramp Connection Scenarios _____	114
Figure 5-2. Different Cases of Phasings at an Intersection _____	116
Figure 5-3. Steps of Incident Management Process _____	118
Figure 5-4. A Sample Survey Form Used for Incident Data Collection In NOVA _____	120
Figure 5-5. A Sample Estimation Tree for Incident Duration Estimation of Property Damage Incidents _____	123
Figure 5-6. A Sample Page of the Survey on Lane Clearance _____	125
Figure 5-7. Dividing Optimization Time into Time Periods with the Same Prevailing Traffic and Incident Condition _____	126
Figure 5-8. The Effect of the Objective Function Weight for the Freeway Component on Delays _____	131
Figure 5-9. The Effect of the Objective Function Weight for the Freeway Component on the Corridor Throughput _____	131
Figure 5-10. The Effect of the Objective Function Weight for the Freeway Component on Diverted Volume _____	132

Figure 5-11. The Effect of the Objective Function Weight for the Freeway Component on the Corridor Throughput at Different Traffic Demand Levels _____	132
Figure 6-1. Illustration of a Corridor with Eight Intersections _____	138
Figure 6-2. Illustration of Phasing Used in Case Studies _____	138
Figure 6-3. The Effect of the Number of Involved Intersections on Delays _____	141
Figure 6-4. The Effect of the Number of Involved Intersections on the Corridor Throughputs _	142
Figure 6-5. The Effect of the Number of Involved Intersections on Total Corridor Diversion Rates _____	142
Figure 6-6. Two Phasing Schemes Used to Determine the Effect of Phasing on Corridor Traffic Flow Conditions _____	145
Figure 6-7. The Effect of the Cycle Length on the Total Delay _____	150
Figure 6-8. The Effect of the Cycle Length on the Corridor Throughput _____	150

LIST OF TABLES

Table 2-1. Comparisons Between TRANSYT-7F and PASSER II/MAXBAND _____	24
Table 2-2. Information Available in SCOOT System _____	31
Table 5-1. Significance of the No. of Police Vehicles Involved on CLT ¹ For Property Damage _____	121
Table 5-2. Fraction of Freeway Section Capacity Available Under Incident Conditions _____	127
Table 5-3. The Effect of the Objective Function Weight for the Freeway Component on Throughput, Delay, and Diverted Volume _____	130
Table 6-1. Optimal Diversion Rates and Timing Plans for a Two-Intersection Arterial _____	139
Table 6-2. Optimal Diversion Rates and Timing Plans for a Four-Intersection Arterial _____	139
Table 6-3. Optimal Diversion Rates and Timing Plans for an Eight-Intersection Arterial _____	139
Table 6-4. Optimal Diversion Rates and Timing Plans for a Sixteen-Intersection Arterial _____	140
Table 6-5. Summary of Optimal Results for Corridors with Different Number of Arterial Intersections _____	140
Table 6-6. The Effect of Phasing on Corridor Traffic Conditions _____	146
Table 6-7. Summary of Optimal Results for the Corridor with Different Demand Level on the Arterial _____	147
Table 6-8. The effect of the Cycle Length on Corridor Traffic Conditions _____	149
Table 6-9. Summary of Optimization Results (Diversion Rates) _____	153
Table 6-10. Summary of Optimization Results (Green Time for Each Phase in Seconds) _____	153
Table 6-11. Comparisons of TRANSYT-7F and COROPT Timing Plans (Green Times in Seconds) _____	155
Table 6-12. Simulation Results for Different Diversion and Signal Retiming Strategies _____	156

1. INTRODUCTION

1.1 BACKGROUND

The increasing volume of traffic on highway networks has put such pressure on existing facilities that one of their main functions, ensuring mobility, has been seriously impeded. Congestion on highways is a daily phenomenon constituting a poor quality of service and causing anxiety and frustration to millions of commuters and business travelers.

The main components of congestion are recurring congestion and non-recurring congestion. Recurring congestion is caused by the high volume of traffic on the highways that exceeds the capacity of the facilities. It often occurs on the urban highway networks during peak periods. The non-recurring congestion, caused by incidents from minor events disrupting normal flow of traffic to major accidents, is the main cause of the congestion and causes up to 60 percent of freeway delay in the United States (Lindley, 1987).

The very aim of traffic control is to provide the best traffic conditions for the users of the networks. Alleviating congestion, thus, becomes one of the main tasks of transportation professionals. Developing and optimizing signal timing plans at intersections of urban streets is essential to reduce vehicle stops, delays, and congestion level, to improve traffic movement efficiency, to enhance safety, to enhance highway operating performance, to improve air quality, and to conserve energy (Transportation Research Circular, 1994). For years, intersection signal control and timing plan optimization technique development have been one of the major concerns of traffic control.

Existing types of signal control are mainly classified as pretimed, semi-actuated, and full-actuated. The later is further divided into full-actuated with volume-density control and without volume-density control (Gartner, 1983, Elahi, 1991).

In pretimed traffic control models, cycle, split, and phasing pattern are fixed over some time periods, thus are unable to respond to traffic fluctuations. Actuated control is more flexible in handling traffic fluctuations. When properly calibrated, traffic-actuated signals in their various forms provide considerable advantages over fixed-time equipment and are widely used (Gartner, 1983). However, the performance of actuated signal control deteriorates under heavy traffic conditions, and the model can not account for unexpected traffic flow conditions such as incidents.

Both pretimed and actuated signal controls are not demand responsive type models. In pretimed or actuated control models, traffic flow process is presented by a single value (flow rate), a statistical distribution (Poisson, Binomial, etc.), and initial conditions (e.g., the initial queue lengths), or in case of oversaturation, by a smooth function of demand versus time. This type of models can not take advantage of the time-variant features of individual vehicle arrival time.

Adaptive control, a new type of operation, is still in the research stage. It represents a real-time, demand responsive traffic signal control and performs the signal operation based on existing traffic conditions. The concept of traffic-responsive control was broadly applied to individual intersection signals as well as to area-wide systems. However, the extent to which traffic responsiveness can be achieved depends on a variety of factors, including the type of control hardware, software capabilities, surveillance and communication equipment, and operator qualifications (Gartner, 1991).

Different traffic control strategies have been developed by traffic engineers. These strategies can be grouped into the following basic categories (FHWA, 1985):

- Isolated intersection control
- Arterial control
- Network control

These categories are further distinguished between off-line strategies, which process manually collected data using batch computer programs to produce signal timing plans, and on-line strategies, which use detector inputs and prediction models to calculate timing plans for immediate implementation.

Another strategy that draws more attention recently is the coordinated operation of freeway ramp metering and adjacent traffic signal systems in order to improve the operational performance of corridors consisting of the freeway, on-ramps, off-ramps, and surface streets. Corridor control is a special case of network control.

1.2 PROBLEM STATEMENT

The Intelligent Transportation Systems (ITS) program has grown rapidly in the United States since the passage of the Intermodal Surface Transportation Efficiency Act of 1992 (ISTEA). One major component of ITS is the Advanced Traffic Management Systems (ATMS) that incorporate state-of-the-art technologies to monitor traffic conditions, adjust traffic operations, and respond to incidents. Clearly, traffic Signal Control Systems (TSCS) and diversion related traffic control will be major building blocks in the development of ATMS.

For the control of traffic signal, an array of strategies and algorithms has been developed, ranging from fixed time control at intersections to demand responsive network coordination based on actual traffic flow data. These control modes,

however, have mostly been designed to handle traffic under normal conditions, i.e., non-saturated conditions, aiming to minimize delays, number of stops, or other performance indexes. With increased traffic demand and congestion level, the United States and many other countries all over the world are confronted with the perennial task of preventing or remedying recurrent conditions of traffic congestion at periods of peak demand, as well as non-recurrent congestion conditions caused by incidents. Development of traffic signal control strategies that will work under non-recurrent condition thus bears more and more significance.

Congestion problems occur both on freeways, major arterials, and in cities. Although congestion normally does not last much longer than several hours, its effect is enormous: Congestion not only increases delay as demand exceeds capacity, it can also create a spin-off effect that is much more dangerous. First, queues of waiting vehicles can block the network considerably. Second, as traffic breaks down the risks increase that an incident will further reduce capacity.

An important issue deserving more research efforts is the integrated corridor optimal control on the freeway, ramps, and surface streets. A freeway with the parallel surface streets and on- and off-ramps form a traffic corridor. The freeway normally carries the majority of the through traffic, while parallel surface streets carry mainly local traffic due to their limited capacity caused by slower speeds and delays at intersections. In case of major freeway incidents, reduced capacity on the freeway section, coupled with long incident duration, will cause tremendous delay on the freeway and diversion might be necessary. The parallel streets often serve as alternate routes in this case.

To achieve optimal traffic flow performance of the corridor system, an integrated control approach should be taken and the following control measures should be treated simultaneously.

- On-ramp metering
- Off-ramp diversion
- Signal timing at surface street intersections.

The time-of-day type of algorithms for ramp metering determine the pre-timed metering rates based on regular demand and roadway traffic pattern, hence they are not effective in dealing with incident situations where diversion is needed and implemented. Furthermore, though it is preferable to divert more traffic from the freeway from the stand point of freeway operation, the capacity of surface streets should be also a factor to determine off-ramp metering rates which determine diversion rate, since overloading surface streets will worsen the overall corridor performance.

With the traffic diverted from the freeway, the traffic volume and patterns on the surface streets are changed. The pretimed traffic signals at intersections can no longer handle the traffic effectively, and retiming of the surface streets becomes necessary.

The overall performance of the corridor traffic flow depends on the following factors that affect each other:

- The volume staying on the freeway
- The diversion and metering rates on on- and off-ramps
- Surface street signal timing plans

To achieve best performance of the corridor traffic flow, integrated optimal traffic control considering these factors simultaneously is necessary.

1.3 RESEARCH OBJECTIVES

The research will address the system operating strategies developed for the coordinated operation of diversion, ramp metering and surface street traffic signal system for a corridor under freeway incident situations. The goal of the research is to devise an effective approach, and an efficient model and algorithm to simultaneously solve for diversion control measures and optimize the overall corridor operational performance.

To fulfill the ultimate goal, the following objectives have been identified:

- To study and develop an appropriate arterial traffic model suitable for near-saturated or saturated traffic condition
- To develop an integrated model to optimize overall corridor operational performance
- To develop evaluation criteria for overall corridor operational performance
- To incorporate congestion relief and congestion avoidance strategies for surface streets in the optimal control model
- To develop an effective and efficient algorithm to find out optimal control measures
- To investigate the feasibility of real-time integrated control
- To use existing software to validate the model developed
- To use simulation models and tools to investigate the benefits of the developed strategies

The optimization model to be developed will have the following control measures:

1. Ramp metering rates that determine traffic diversion strategies including the following two aspects:
 - On-ramp metering rates
 - Off-ramp metering rates
2. Timing plans for arterial intersections including the following components:

- Green times (phase lengths)
- Cycle length
- Offset

1.4 ORGANIZATION OF THE DISSERTATION

In the next chapters, the extensive literature review on signal timing and integrated corridor control is introduced. It is followed by the proposed model formulation, solution algorithm and procedure. The computer program for obtaining optimal control strategies is presented. The model test and validation results are also provided. Finally, the conclusions and suggestions for future research are discussed.

2. LITERATURE REVIEW

2.1 INTRODUCTION

Signal timing optimization at intersections has drawn extensive attentions for decades. Different approaches, models, and algorithms have been developed, tested, and implemented at the following network levels:

- Isolated (single) intersection
- Arterial
- Area wide surface street network
- Corridor consisting freeways, arterials, and ramps

Different optimization Criteria have been used in different models. Among them are the following:

- Total delay
- Total number of stops
- Queue length
- Throughput
- Linear combination of delay and number of stops
- Total cost of losses
- Total fuel consumption
- Total person delay
- Sum of the square of the queue length (This index was used to penalize long queues because long queues have a greater impact on the criterion value than the shorter queues in the optimization procedure.) (Reljic et al., 1990).

The signal timing optimization approaches and models can be divided into the following categories:

- (1). Vehicle and traffic actuated Control models at an isolated intersection;
- (2). Off-line surface street optimization models;

- (3). On-line surface street optimization models;
- (4). Integrated corridor optimization models.

2.2 VEHICLE AND TRAFFIC ACTUATED CONTROL MODELS AT AN ISOLATED INTERSECTION

In general, the term “vehicle-actuated control” infers some form of control strategy that is directly dependent on measure of vehicle activity. Traffic actuated control is used when data of all traffic is available to calculate the signal phase on-line.

One or more of the following traffic flow characteristics usually form the basis of traffic responsive signal control (OCED, 1981)

- Vehicle interval: Traffic volumes on the controlled approaches are measured by small area detectors and green time is assigned to each movement sufficient to serve the measured demand (a pre-selected amount of green time for each vehicle).
- Vehicle waiting time (gap reduction): This method takes into account the length of time that vehicles are waiting to be served. This is accomplished by reducing the amount of green time given to each unit being served as clock passes.
- Volume density: This method, a variation of item 2) above, relates traffic volume and density flow (time between successive vehicles) in determining the amount of green time for each phase.
- Lane occupancy: This method involves the use of a “long loop” placed on the approaches. Green is extended for any phase while a vehicle is in the “long loop”.
- Rate of occupancy: This method takes into account the volume of vehicles waiting to be served.

When utilizing any of the above parameters as the basis for control, several constraints must be considered.

- Maximum green time
- Pedestrian crossing time
- Safe stop distance
- Queue length and storage area
- Intersection geometry, i.e., clearance times

Miller suggested a simple self-optimizing strategy based on the criterion of minimizing the total vehicle delay (Miller, 1963). In Miller's strategy, the decision to extend a phase is made at regular intervals by the examination of a control function. This function represents the difference in vehicle-seconds of delay between the gain made by the extra vehicles that can pass the intersection during an extension and the loss to the queuing vehicles in the cross street resulting from that extension. Miller's method has been further studied by Weinberg (Weinberg, 1966) and Ross (1969).

Bang also use Miller's idea to develop TOL (Traffic Optimization Logic) (Bang, 1976). The method is exemplified the simple case of an intersection between two one-lane approaches A and B. Assuming that lane A has green for the moment. The decision to extend the green is based upon the evaluation of the control function φ_a .

$$\varphi_A = r_A \cdot (a_v \cdot \delta_{AV} + a_b \cdot \delta_{Ab} + a_p \cdot \delta_{Ap}) + b_v \cdot \delta_{AV} + b_b \cdot \delta_{Ab} - h(a_v \cdot n_{BV} + a_b \cdot n_{Bb} + a_p \cdot n_{Bp}) - (b_v \cdot \Delta n_{BV} + b_b \cdot \Delta n_{Bb}) \quad (2-1)$$

Where

- r_A : time interval (red and intergreen) until phase A will get green light again if it is terminated immediately
- a : cost of delay per second
- δ : number of additional cars, buses, pedestrians that can pass the intersection if the green is extended by h seconds

b : vehicle operating cost to bring a vehicle to a complete stop to resume normal speed

h : time interval between the calculation of the control function

n : number of queuing cars, buses etc. in approaches with red light that will suffer an increased delay of h seconds if the prevailing green is extended

Δn : number of extra vehicles, buses etc. that will be forced to a full stop if the prevailing green is extended by h seconds

v, b, p : index for vehicle, bus and pedestrian

Phase A is extend until $\phi_a < 0$ subject to restrictions of maximum green time.

The method does not guarantee that an overall optimal control is based on a large number of short-term optimizations. The results of the study indicated that TOL gave substantial reductions of average delay and number of stops as compared with conventional pretimed or vehicle actuated control.

French Technical Department of Roads and Motorways (SETRA) started conducting an experiment in 1973, in order to evaluate the advantages that could be provided by traffic actuated adaptive control (Jezequel, 1976). Tests were made on critical vehicular gap algorithm, flow occupancy algorithm, the "AERO" occupancy rate algorithm, and Miller's algorithm. The results of optimization using each algorithm were compared with pretimed signal control.

Robertson formulated a model DYPIC (Dynamic Programmed Intersection Control) for an intersection with known sequence of vehicle arrivals (Robertson, 1974). The procedures for calculating the minimum delay arrays at time $(t-1)$ in DYPIC are as follows:

- 1) Select an intersection state at time $(t-1)$.
- 2) Evaluate the total delay after $(t-1)$ assuming:
 - The Signal is not changed, and

- The Signal is changed.
- 3) Choose the “change”, or “no change” decision, whichever gives less total delay.
 - 4) Update the state.

Dynamic programming technique was used to find the optimal policy. The DYPIC can only handle single intersection with a limited number of approaches. Otherwise, the amount of computation increase will make the approach impractical.

Swedish system LHOVRA (Swedish acronym for the system functions truck priority (L), primary road priority (H), accident reduction (O), variable amber (V), control of red-light infringements, (R) and change of mind at all-red (A)) is a signal control optimization model for isolated intersections (Peterson, 1986). It had the feature of reducing intersection accidents in “dilemma zone”, primary road priority, variable amber, and truck priority, etc. The results of the system implementation were reported to be successful and the technique is used as the standard when signaling isolated intersection in Sweden (Peterson, 1982).

English system MOVA (Modernized Optimized Vehicle Actuation) is a strategy for isolated intersection developed by Transport and Road Research Laboratory (Vincent, 1982, 1986). The strategy makes use of two vehicle detectors per approach lane, one at 40 meters and one at 100 meters before the stop line. Initially, each lane receives sufficient green to discharge vehicles queued back to the 40 meter detector. Next, the time intervals between vehicle detectors are examined to find when there is a reduction in saturation flow. Finally, there is an optimization process that balances the merits of extending a green against the disbenefit to the vehicles stopped by the red signal; this aims to avoid the tendency to extend green times unnecessarily when traffic is flowing at considerably less than full saturation flow. Several versions of MOVA have been tested and the results showed that the values of the benefits from MOVA control were likely to far exceed the implementation cost.

Cantarella developed a method for simultaneous computation of cycle time and green time for different groups of a single intersection, and their succession (Improta, 1984). The optimal design of the control system is obtained by solving a Binary-Mixed-Integer Linear Programming Using discrete programming technique.

The Webster's delay formula was used (Webster, 1958). The unitary delay d_h of the uses of stream h belonging to group k is formulated as a function of the effective green g_k and the cycle time c :

$$d_h(g_k, c) = \frac{9}{10} \left(\frac{c(1 - g_k / c)^2}{2(1 - q_h / s_h)} + \frac{x_h^2}{2q_h(1 - x_h)} \right) \text{ (seconds)} \quad (2-2)$$

It is convex in the variable $1/c$ and g_k/c . Piecewise linearization of this delay function was made in optimizing process. A Branch-and Bound algorithm is used for solving the problem.

The model is good for off-line application since Branch-and-Bound algorithm guarantees global optimum with a number of searches. However, for on-line application, search has to be terminated after some appropriate results is obtained, global optimum is not guaranteed or computation time will be too long.

Gallivan discussed two aspects of isolated intersection control problem (Gallivan, 1988). There are: first, the choice of signal groups to run together and the order in which they should run, and the second , the optimal choice of times for which to grant right of way to each signal group. A combination method due to Murchland and Tully was described. The method generates sequences of cliques of signal groups as candidates to determine stage structure and order for an intersection. A convex programming method due to Allosp was described. Given a stage structure,

lost time and extra effective green times, the method allows control performance at the intersection to be optimized by suitable choice of stage duration.

Sen developed a dynamic programming model for single intersection signal control (Sen, 1993). In the model, state variable S_j for stage j was defined as the remaining time after the first $j-1$ "transition", the decision variable x_j denoted the amount of green time allocated to the phase corresponding to stage j . Total number of stops was used as the evaluation criterion and backward recursion was used to find the optimal policy. The model used only one decision variable thus make the strategy very effective. Since total number of stops is not an ideal performance index under saturation traffic condition, delay, stops and queue length were later introduced into the model (Sen, 1994). The model provides the intersection control within a large system called RHODES (Head, 1992).

The phasing and timing of traffic signals requires the use of some heuristic rules of thumb. Because of the need for judgmental knowledge in solving this problem, artificial intelligent technologies have been used in the past.

Linkenheld used the knowledge-base expert system technology to develop a system for the phasing and signal timing (PHAST) of an isolated intersection (Linkenheld, 1992). PHAST takes intersection geometry and traffic volume as input and generates appropriate phase plan, cycle length, and green time for each phase. The knowledge Base is composed of the following rules:

- Design process control rules
- Initial phasing and timing rules
- Initial computation rules to find volume of each approach, sum of approaches, and left-turn difference, etc.
- Signal justification rules to establish the minimum thresholds that would justify the need for a signal phasing and timing design

- Phasing rules
- Degree of saturation rules to compare the critical volume/saturation (v/s) flow ratio for the soon-to-be-compared cycle
- Cycle length rules. The cycle length is computed using HCM (Highway Capacity Manual, 1985) formula.

$$C = \frac{LX_c}{X_c - \sum(\frac{v}{s})} \quad (2-3)$$

Where

L Total lost time for all discrete phases

X_c Critical volume capacity ratio

C Cycle length

- Timing rules. The effective green g is calculated using the HCM formula

$$g = \frac{v}{s} \cdot \frac{C}{X} \quad (2-4)$$

Where X is an estimate of the subphase volume/capacity ratio.

A comparison of delays and levels of service of the phase plans and signal timings generated by PHAST to those of HCM indicated that very comparable optimal solutions were being generated by the PHAST system. Nevertheless, this was an off-line approach, further validation and on-line implementation is needed.

Elahi used Knowledge-Based Expert System to develop an adaptive traffic signal control for isolated intersection (Elahi, 1991). In the system, HCM formula was used for delay equations:

$$d = 0.38C \frac{[1 - (g / C)]^2}{[1 - (g / C)X]} + 173X^2 [(X - 1) + \sqrt{(X - 1)^2 + 16X / c}] \quad (2-5)$$

Where

d average stopped delay per vehicle for the lane group (sec/veh),

C cycle length (sec)

g effective green time (sec)

- X degree of saturation = v/c ratio for lane-group
- c capacity of the lane-group (veh/sec)
- v vehicle flow rate (veh/sec).

To calculate queue length, Miller's formula was used for undersaturated condition:

$$\begin{aligned}
 Q_0 &= \exp[-(4 / 3) * (\lambda * C * s)^{0.5} * (1 - X) / X] / [2(1 - X)] \\
 Q_r &= v * r = v * C(1 - \lambda) \\
 Q &= Q_0 + Q_r
 \end{aligned}
 \tag{2-6}$$

where

- λ G/C ratio
- G effective green (sec)
- s saturation flow (veh/sec)
- v vehicle arrival rate (veh/sec)
- c capacity (veh/sec) = $s \times \lambda = s \times G / C$
- Q_0 average overflow at the beginning of red
- Q_r the number of vehicles that arrived during red
- Q queue length

For near- and over-saturation, a deterministic input-output model is used to estimate queue:

$$Q_i = Q_{i-1} + (v - s) * g + v * r \tag{2-7}$$

where

- Q_{i-1} queue length at the end of cycle $i-1$
- Q_i queue length at the end of cycle i
- r red interval

For an on-line application, a traffic prediction model is needed. The model used in the system uses weighted combination of the most recent volume and data base mean. A smoothing factor was introduced to reduce the adverse impact of abrupt rise and fall in the traffic volume. A separate data base was stored for each 15-min

interval. The concept of delay grade and queue grade and combined grade were used for decision making.

The prototype system can operate both the conventional and adaptive control modes of signal operation. Though the simulation found the prototype to be effective, more rigorous testing for different scenarios is needed.

Various studies have shown that traffic responsive signal optimization models are more efficient than fixed time control for under-saturated conditions, but have not shown any agreements as to the superiority over the other strategies among them. It has been agreed that the performances of this type of control deteriorate as traffic demand increases to near- or over-saturated conditions. The strategies of this type are not obviously efficient at saturation.

This type of signal control requires advance information about traffic. The performances of the strategies depend heavily upon the accuracy of the traffic information. The configuration and locations of detectors and the accuracy of traffic prediction models are therefore vital.

2.3 OFF-LINE SURFACE STREET NETWORK SIGNAL OPTIMIZATION MODELS

Off-line signal optimization models (computer programs) have played important roles in surface street network signal control. They serve as analysis tools for various levels of network traffic control. The most famous model is TRANSYT. Other models that are also extensively used are PASSER II, MAXBAND, etc.

2.3.1 TRANSYT Model

The TRANSYT (Traffic Network Study Tool) model was first developed by Mr. Dennis Robertson of Transport and Road Research Laboratory of the United Kingdom in 1967 (TRANSYT 7F User's Manual, 1984). Over the years, many revisions and improvements have been made. The latest version in the United States is TRANSYT-7F, HFWA's version based on TRANSYT-7.

TRANSYT uses a macroscopic traffic model that is considered to be the most realistic. The traffic model integrates the arrival and departure profiles of traffic at the downstream intersection to calculate delay. The accuracy of the arrival flow profile at downstream intersection in turn depends on the platoon dispersion algorithm utilized by TRANSYT. The platoon dispersion behavior is thus the fundamental principle of traffic representation in TRANSYT.

TRANSYT's basic recursive platoon dispersion model is as follows

$$Q'_{t+T} = FQ_t + (1-F)Q'_{t+T-1}$$
$$\text{or } Q'_t = FQ_{t-T} + (1-F)Q'_{t-1} \quad (2-8)$$

where: $f = 1 / (1 + \alpha \beta t_\alpha)$

Q_t flow rate into the link during time interval t (vehicles per second)

Q'_t the predicted arrival flow rate during time interval t

T a lag time, which is found to be βt_α

β a TRANSYT specific travel time factor

t_α the average travel time from the link's entry point to the tail at the vertical queue at the link's downstream stop link

α a TRANSYT specific platoon dispersion factor (second⁻¹)

In TRANSYT-7F, $\beta = 0.8$ and $\alpha = 0.20 \sim 0.50, 0.35$, $\alpha = 0.35$ is often used

The following measures of effectiveness (MOEs) are considered in the TRANSYT model.

- 1) Stop and queuing: By integrating "IN" and "GO" patterns, "OUT" pattern is defined and number of vehicles in the queue can be determined.

$$m_t = \max [(m_{t-1} + q_t - s_t)] \quad (2-9)$$

where

m_t number of vehicles in the queue in time interval t on a given link,

m_{t-1} number of vehicles in the queue in time interval $t-1$ on a given link

q_t number of vehicles arriving in interval t , given by "IN" pattern

s_t number of vehicles allowed to leave in interval t , given by the "OUT" pattern.

TRANSYT-7F assumes that vehicles that are delayed are also stopped.

- 2) Delay: Delay is composed of a uniform element, d_u , a random element, d_r and the delay due to oversaturation d_s . TRANSYT combines d_r and d_s as d_{rs} . The TRANSYT-7F calculates delay using the following formulas.

$$d_u = \frac{C}{3600N^2} \sum_t^N m_t \quad (2-10)$$

where

d_u uniform delay (veh-hr/hr)

C cycle length (sec)

m_t queue length during step t

N number of steps in the cycle

$$d_{rs} = \left[\left(\frac{B_n}{B_d} \right)^2 + \frac{X^2}{B_d} \right]^{1/2} - \frac{B_n}{B_d}$$

$$B_n = 2(1 - x) + xz$$

$$B_d = 4Z - Z^2 \quad (2-11)$$

$$Z = (2X / V) * 60T$$

where

d_{rs} random and saturation delay

X degree of saturation

V volume on link

T period length, normally 60 minutes, and

$$D = d_u + d_{rs} \quad (2-12)$$

D is the total delay on a link.

- 3) Total travel: This is a simple aggregate of the product of link volume and link length.

$$TT_i = q_i \times L_i \quad (2-13)$$

where

TT_i total travel on link i (veh-km/hr)

q_i traffic volume on link i , (vph)

L_i length of link i (mi)

- 4) Total travel time: This MOE is the product of link volumes and total time spent on links, including delay

$$TTT_i = q_i \left(\frac{L_i}{u_i} + D_i \right) \quad (2-14)$$

where

TTT total travel time on link i (veh-hr/hr)

u_i average free speed on link i (km/hr)

D_i total delay on link i (veh-hr/hr)

- 5) Average speed: This is an indication of overall quality of flow in the network. It is simply the ratio of total travel TT and total travel time TTT .
- 6) Fuel consumption: Fuel consumption is estimated by a regression formula.

$$F = k_1 TT + k_2 D + k_3 S \quad (2-15)$$

where

F fuel consumption (gallons/hr)

S total stops (stops/hr)

K_i coefficients of regression, which are functions of free speed.

When optimizing, TRANSYT-7F minimize an objective function called the performance index (PI). The PI is a linear combination of delay and stops and is expressed as follows:

$$PI = \sum_{i=1}^n (d_i + ks_i) \quad (2-16)$$

where

d_i delay on link i (of n links) (vehicle.hours)

s_i stops on link i in stops/second

k a user input coefficient to express the importance of stops relative to delay

TRANSYT-7F uses a hill-climbing technique (nonlinear programming gradient search method) to explicitly optimize phase lengths and offsets separately for a given cycle length. To determine the best cycle length, an evaluation of a specified range of cycle lengths may also be made.

Some drawbacks of TRANSYT-7F are as follows:

- The “hill-climbing” technique is an iterative, gradient search that requires extensive numerical computation, which makes it hard to use the model in real-time application.
- By using the gradient search technique (hill climbing), global optimum is not guaranteed.
- The software can not analyze alternative phase sequences. To evaluate the alternative phase sequence, multiple runs are required.
- The system does not really optimize cycle length.
- The software requires an initial timing plan. The final optimal plan often depends on the starting solution.
- Progression on the main stream is not considered. (In the TRANSYT-8, this feature is added into the system).

- The traffic model assumes that queue dissipates at the end of green, no queue is built up. Queue is treated as stacking vertically at the stop line.

To optimize signal under saturation, improvements on the traffic model and optimization technique should be made. In the latest version (not available in the US), the effects of queue spillback are considered in the model.

2.3.2 PASSER and MAXBAND

PASSER (Progression Analysis and Signal System Evaluation Routine) represented by the commonly used PASSER II is a serial of programs developed by Texas Transportation Institute (TTI, 1988). PASSER II program can be classified as a macroscopic optimization model. The arterial progression optimization section of the program can analyze the four possible arterial phase sequences per intersection and select from the phase sequences at the intersection the one that provides the best overall progression solution. The optimization procedure is an implicit enumeration of the minimum interference values performed by using a variant of the half-integer synchronization approach for relative offsets. The unique advantage of the PASSER II program over other optimization programs for signalization is that it can be used to consider and select multiple-phase sequences (Rogness, 1983).

The main drawback of PASSER II is that, because it is based on a variant of the maximum bandwidth progression solution, it might not optimally minimize vehicle delay or stops.

Like PASSER II, MAXBAND selects a signal timing plan by maximizing bandwidth efficiency (the ratio of total bandwidth to the cycle length) and has been used primarily for arterial streets. A mathematical programming algorithm, mixed integer linear programming (MILP), is used to select offsets, cycle length, and phase

sequences, which maximize the weighted sum of bandwidths in both directions on an arterial (Cohen, 1983). The program also allows small deviations from the arterialwide progression speed on individual links, a process referred to as speed search. Unlike TRANSYT-7F, MAXBAND program obtains a global optimum, requires no starting solution.

However, the traffic model is oversimplified. No account is taken of secondary flows turning from side streets, platoon dispersion, turning traffic, or platoon shape. As the results, it is not generally true that maximizing bandwidth minimizes such MOEs as stops and delay. Another drawback is that green phase times (green splits) are not optimized. This is because bandwidth provides no criteria for setting green times on side streets.

Both PASSER II and MAXBAND have been used primarily for arterial streets (Hadi, 1993). Though MAXBAND has been reported to be extended for applications to multiple-arterial networks (Chang, 1988), this version is not widely used, primarily because of excessive computer time and its current limitation to mainframe computers. Both programs are not well suited for oversaturated traffic conditions.

2.3.3 Comparison between TRANSYT-7F and PASSER II/MAXBAND

A comparison between TRANSYT-7F and PASSER II / MAXBAD is provided (Table 2-1).

It can be seen that each program has its advantages and disadvantages as well. None of the above available signal timing method is considered adequate to optimize all four design elements (phase sequence, splits, cycle length and offsets), particularly in two dimension networks. Many individual research efforts have contribute to the improvements of the above mentioned programs.

Table 2-1. Comparisons Between TRANSYT-7F and PASSER II/MAXBAND

	TRANSYT-7F	PASSER II / MAXBAND
Traffic Model	Realistic	Oversimplified
Optimization Technique	Hill-Climbing	Mixed Integer Linear Programming
Global Opt. Guaranteed?	NO	YES
Maximize Green Band?	NO	YES
Optimize Split?	YES	NO
Application Level	Arterial/network	Arterial
Needs Initial Timing Plan?	YES	NO
Optimize Phase Sequence?	NO	YES
Good for Saturated Traffic	NO	NO
Optimize Overall Performance?	YES	NO

In the latest release of the TRANSYT program, a progression-based optimization model that uses the progression opportunities (PROS) concept has been implemented, overcoming an early shortcoming of the model. This concept expands upon the maximum bandwidth approach by considering short-term progression opportunities within the system (Hadi,1993)

Since TRANSYT-7F and PASSER II/MAXBAND programs are complementary, several studies have proposed a heuristic signal optimization approach in which PASSER II and MAXBAND have been used as a “preprocessor” before running TRANSYT-7F (Hadi, 1993; Rogness, 1983; Cohen 1983). The approach is described below:

- 1) Using either volume and capacity information or existing green times, execute MAXBAND or PASSER II to provide offsets, cycle length, and phase sequence.
- 2) Using the results of step1) as the input, execute TRANSYT to provide final offsets and green times.

2.3.4 Other Off-Line Software and Studies

SIGOP II (SIGnal Optimization) is a program developed for the Federal Highway Administration for the purpose of computing optimum signal timing plans for either arterial or network systems (OCED, 1981; MacGowan, 1975). The program is composed of two major components:

- 1) A flow model to describe traffic behavior compatible with the applied signal timing;
- 2) An optimization scheme designed to yield signal timings that serve current patterns of traffic with maximum efficiency.

Traffic is modeled as a platoon of vehicles that disperses as it travels along the link. Signal timings (splits and offsets) are obtained from a process whereby a series of optimization problems are solved using gradient search method for every node in the network based on the disutility function. Overall, the program has many similar concepts with TRANSYT-7F. Having special measures to handle oversaturation is one feature of the system.

SIGSET is a maximum bandwidth program (OCED, 1981). Like other green bandwidth maximization programs, it does not have much value in saturation condition. The reason is that the presence of queues, which inevitably form in heavy traffic, is not taken into consideration in designing.

Gartner developed a multi-band approach to arterial traffic signal optimization (Gartner, 1991). The method generates a variable bandwidth progression in which each directional road section can obtain an individually weighted bandwidth. A systematic traffic-depend criterion is incorporated in the model. Mixed-integer linear programming was used for the optimization. Simulation indicated that this approach produced considerable gain when compared with the traditional progression method.

Improta proposed a method to find optimal offsets in the urban network. A binary integer programming model was developed for optimal offset calculation. (Improta, 1982).

MAXBAND 86 was the only program now generally available for progression bandwidth maximization in multiarterial networks, but it does not optimize green split, has a very simplistic traffic model, and is extremely inefficient for practical computations. In addition, it has no capability for reporting traffic measures such as delay, stops, and level of service. Chaudhary developed a practical program PASSER IV (not a member of PASSER family) for optimizing progression bandwidth-based signal timings for arterial as well as multiarterial closed-loop systems (Chaudhary, 1993).

The PASSER IV has the following key Features:

- GUI (Graphical User Interface)
- Computational efficiency
- Minimization of cycle length
- Output reports for multiple solutions

The following additional features were to be implemented:

- green time optimization
- concurrent bandwidth and delay optimization
- delay estimation routine

However, PASSER IV still has the main disadvantage like other progression-based models: it is only suitable for undersaturated situation and no congestion reduction procedures can be incorporated.

The off-line signal optimization models provide the analysis tools and form the basis of adaptive and real-time control. All programs have their strengths as well as weaknesses. One issue that most models failed to address well is the signal optimization under saturated traffic conditions.

2.4 ON-LINE NETWORK OPTIMIZATION MODELS

Ever since the inception of modern traffic signal control, demand responsive traffic control strategies have been attempted by many traffic engineers and researchers hoping to improve traffic performance. Many efforts have been put to develop demand responsive on-line signal optimization system for traffic network. The famous existing on-line systems include, UTCS developed in US, SCOOT system from England, SCAT system of Australia, and OPAC system in US.

2.4.1 UTCS

To develop more advanced (responsive) traffic control strategy, numerous experiments have been conducted. One of the most comprehensive studies was the UTCS (Urban Traffic Control System) that was carried out by the FHWA in Washington D.C. in 1970's.

Three generations of the UTCS have been designed for centralized computer control with a situation-dependent amount of freedom for variations on the basis of Intersection data. (OCED, 1981, Gartner , 1995).

- **First Generation Control (1GC):** It uses prestored signal timing that is calculated off-line based on historical data. It also includes logic to enable a smooth transition between different signal timing plans, and a Critical intersection Control (CIC) feature that enables vehicle-actuated adjustment of green splits at selected signals.

- 1.5GC: It is a strategy in which new timing plans are generated automatically when traffic conditions warrant it.
- Second Generation Control (2GC): It is an on-line strategy that computes in real-time and implements signal timing plans based on surveillance data and predicted values. It contains an optimization algorithm (SIGOP), a traffic prediction model, sub-network configuration models, critical intersection control(CIC), and a transition model to minimize transition time between two plans.
- Third-Generation Control (3GC): It is a fully responsive, on-line control that minimizes networkwide objective using predicted traffic conditions for inputs. The difference compared to 2GC were the period after which timing plans were revised was shortened to 3-5 minutes, and the cycle length was allowed to vary among the signals as well as at the same signal during the control period.

Unfortunately, test results showed that the more “responsive” strategies resulted in poorer performance than the fixed-cycle, non-responsive strategies. 1-GC performed overall best. 2-GC demonstrated some small improvements on the arterial, but degraded traffic flow in the network. 3-GC was unsuccessful in responding to traffic flows and having degraded performance under almost all conditions for which it was evaluated.

Following are some of the reasons that are considered to be the cause of the poor showing of the responsive UTCS strategies:

- Because of the inherent inaccuracies in the measurement-prediction cycle, neither 2-GC nor 3-GC could respond adequately to rapid changes in flows.
- The frequent transition in signal timing may be more harmful than the adaptiveness sought by on-line optimization in 2-GC and in 3-GC. Considerable delays are incurred during transition.

- While 3-GC had a variable-cycle feature, the entire signal switching sequence was predetermined for every control period and therefore not dynamically responsive to the actual traffic condition on the street.
- The control strategy used by 3-GC was centralized and optimization procedure required a relatively long time for convergence. The time allotted for this procedure was insufficient for reaching a good optimum.

To avoid the similar failure of UTCS adaptive control strategies, new concepts and approaches have been developed since 1980's. Two categories of systems were developed:

- The system that makes little changes on a predefined signal plan as SCOOT or those which choose a signal plan among a pre-specified set of plans as SCAT.
- The category of systems that decide to switch the signal or not at each step of time as in OPAC, PRODYN and UTOPIA.

2.4.2 SCOOT and SCAT

Since 1973, the U.K. Transport and Road Research Laboratory has been researching a traffic responsive system--SCOOT (Split, Cycle, and Offset Optimization Technique). It has been widely used since 1981 in more than 30 cities principally in England but also in other cities in the world including Beijing, Hong-Kong, and Red Deer (Boillo, 1992).

SCOOT has the following features which make the system possible to overcome the problems in the earlier responsive systems (e.g., UTCS) (Hunt, 1982).

- **Minimum transients:** SCOOT uses frequent small incremental alteration to split, cycle time and offset. These alterations minimize transients, but can add up to create a new pattern of coordination.

- Short-term prediction: Most SCOOT decisions are based on the current situation and long term predictions are seldom necessary.
- Fast response: In SCOOT, every red/green transition is optimized. The split, offset and cycle optimizers are designed into a hierarchy to ensure compatibility and minimal interaction.
- Faulty detector: Detectors are monitored continuously for faults.
- On-line Traffic Model: SCOOT uses data from the detectors to predict the queues at signal stop lines.
- No background plan. SCOOT needs no initial fixed-time plan and can start from any signal settings.

For each link, the SCOOT traffic model predicts the current value of queue at the stop-line including the length of the queue and the position of the back of the queue that provides the congestion information for the optimizers.

SCOOT is similar to the TRANSYT program in the principle of optimization. It has a set of optimizers. A few seconds before each stage change, the split optimizer estimates whether it is better to make the change earlier, as scheduled, or later and implements whichever alteration will minimize the maximum degree of saturation on the approaches to the intersection. The offset optimizer operates on each intersection for each cycle. The information on the cyclic flow profiles is used to estimate whether or not the alteration of the offset will improve the overall progression on the links immediately upstream or downstream. A performance index using delay, stops and congestion is minimized.

SCOOT operates sub-areas of signals on a common cycle in order to maintain coordination between signals. The cycle time optimizer can vary the cycle of each sub-area in increments of a few seconds at intervals of not less than 2.5 minutes.

The optimizer adjust the cycle so that the heavily congested intersections are operating at about 90% degree of saturation.

The information available in SCOOT is presented in Table 2-2.

Table 2-2. Information Available in SCOOT System

Quantity	Units	Available for every	Frequency
flow	vehicles/hour	link	every 5 min. or cycle
delay	vehicle hour/hour	link	every 5 min.
stops	vehicles/hour	link	every 5 min.
SCCOT congestion	intervals/hour	link	every 5 min.
queue lengths	vehicles	link	once per cycle
queue clear time	second	link	once per cycle
stage timings	second	node	every minute
stage lengths	second	node	every minute
offsets	second	node	once per cycle
cycle time	second	node	every 5 min.
performance index		link or node	every offset opt.

One advantage of SCCOT is that it has congestion reduction feature. The amount of congestion, measured in the model for each link, is used to modify the decisions of the split and offset optimizers. The green time can be increased to reduce congestion and the offset on a link can be improved to reduce the risk of blocking the upstream intersection.

The evaluation of SCOOT system indicated that SCOOT provides better performance than fixed-time signal. Based on the evaluation results, Hunt suggested that SCOOT is most effective where traffic demand approaches the capacity of the intersection, where demand is unpredictable and where the distances between intersections are short.

SCAT has been developed since 1979, and been installed in more than 20 cities. In SCAT, an area for signal coordination is divided into smaller sub-areas of about 1 to 10 signalized intersections with common cycle time. The common cycle time is updated every cycle in steps of up to 6 seconds according to the degree of saturation of that sub-area (given as the ratio of the green duration actually used to the total green duration). Four phase split plans that express green times plus intergreens as percentages of the current cycle time are available at each intersection within a sub-area. The plans include vehicle-actuated tactics like skipping, transfer of spare time, etc. Each sub-area has 5 offset plans as part of predetermined inputs. Internal offsets can vary according to the current cycle time and progression speed factor that governs the percentage change in offset. The determination of each of the three timing elements is independent of others, though all affected by the degree of saturation. Ideally the three parameters should be optimized simultaneously so that delay, stops and other indexes are minimized (or maximized). However, it is not practical because the amount of computation and the data to be collected on-line would be prohibitively high thus make it impractical for on-line application.

Satisfactory test results have been achieved by SCAT. Though global minimal is not guaranteed by the nature of its optimization approach.

2.4.3 OPAC, PRODYN and UTOPIA

Signal timing is a multi-stage decision making process. Dynamic Programming (DP) is very efficient in dealing with this type of control problem and makes itself a good candidate. Miller first proposed a DP strategy in which a decision whether to extend a phase is repeatedly made at very short fixed intervals by the examination of delay-based control function (Miller, 1965). Similar approach was also proposed by Zijverden (1969), and Bang (1976). Grafton developed a continuous-time DP model

(Grafton, 1967) while Robertson developed a discrete-time DP model (Robertson, 1974). In those models, backward DP approaches were used. Those approaches had the following drawbacks thus not suitable for on-line network control.

- The models require advanced knowledge of arrival data for entire horizon period, which is far beyond that can reasonably expected to be obtained from the available surveillance systems.
- The models require extensive computation effort, and, since the optimization is carried backward in time, it precludes the opportunity for updating the input data or correction of future control policy.

OPAC (Optimization Policies for Adaptive Control) was developed initially as an on-line adaptive signal control model at isolated intersection by Gartner (Gartner, 1983), and has been under the test and implementation process to serve as a building block for demand-responsive decentralized control in a network.

To avoid the shortcoming of the other DP based model, OPAC uses a simplified approach described as follows:

- 1) The optimization process is divided into sequential stages of T seconds. (The stage length is in the range of 50-100 s, about the length of a cycle). The information and decision flow at stage n is shown below:

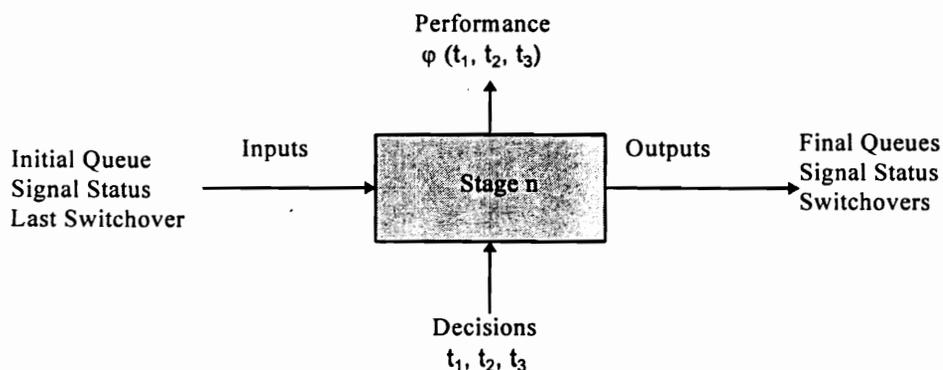


Figure 2-1. Information Flow at Stage n of OPAC System

- 2) During each stage, 1-3 switchovers are required to provide sufficient flexibility for deriving an optimal demand responsive policy.
- 3) For any given switching sequence at stage n , a performance function is defined on each approach that calculates the total delay during the stages (in vehicle intervals).

$$\varphi_n(t_1, t_2, t_3) = \sum_{i=1}^k (Q_0 + A_i - D_i) \quad (2-17)$$

Where

Q_0 initial queue

A_i arrival during interval i

D_i departure during interval i

(t_1, t_2, t_3) possible switching times during this stage

- 4) The optimization procedure used for solving the problem is an optimal sequential constrained search (OSCO) method. The objective function (total delay) is evaluated for all possible switching sequences. At each iteration, the current objective function value is compared with the previously stored value, and, if lower, replaces it. The corresponding switch point times and final queue length are also stored. At the end of the search, the values in storage are the optimal solution.
- 5) The optimal switching policies are calculated independent of other stage, in a forward sequential manner.

The technique requires future arrival information for the entire stage that is difficult to obtain with reliability. The rolling horizon concept is introduced to reduce these requirements in such a way that one can use only available flow data and yet preserve the performance of the computational procedure. The concept is described as follows:

- A stage is divided into k intervals (stage = projection horizon)
- The arrival information for intervals at the head of the stage can be obtained from upstream detectors.
- The tail ($k-r$) interval arrival information is obtained from the model (smoothed average flow for tail arrival data).
- Optimal policy is calculated for the whole stage but implemented for only the head section.
- The projection horizon is shifted by r intervals and the process repeats.

The rolling horizon approaches is illustrated in Figure 2-2.

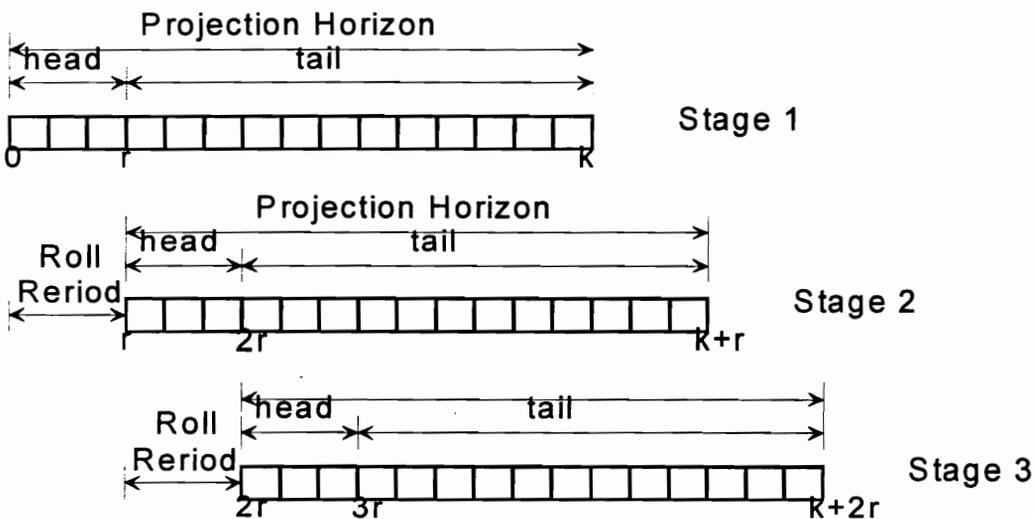


Figure 2-2. Illustration of Rolling Horizon Concept

Tests were made using actual arrival data and NETSIM simulated data.

OPAC has the following features:

- It provide performance results that approach the theoretical optimum.
- It uses on-line data that can be readily obtained from upstream link detectors instead of using predicting models that have been proven to be unsuccessful in arterial real-time applications.
- It is suitable for implementing on existing microprocessors.

Simulation and field tests showed promising results for OPAC model.

PRODYN is a French urban traffic control system developed by CERT and assessed on ZELT experimental field in Toulouse (Henry, 1989). The main characteristics of the system include the following:

- A five second sample time: The control is the decision to switchover from one stage to another stage.
- The use of magnetic loop sensors by lane: One at the entrance of the link, the other at 50 meters upstream of the stopline.
- Explicit minimization of delay.
- Use of automatic control method including Bayesian estimation, dynamic programming and decentralized method.
- Rolling horizon approach.

As a decentralized system, the optimization process is performed at each intersection, but coordination of signals in the network is also considered. The intersection computers that have downstream PRODYN controlled intersections at less than 200 meters simulate the outputs resulting of the application of the optimal control on $[(k+1)\Delta t, (k+K+1) \Delta t]$ using fixed turning movement ratios, and send them to the concerned downstream intersection computers. Between sampling time $(k+1) \Delta t$ and $(k+2) \Delta t$, these computers utilize the received data as a forecast of the arrivals. The arrivals for the remaining part of the horizon $[(k+K+1) \Delta t, (k+K+2) \Delta t]$ are predicted as the average of the 16 last measurements at the entrance of the link.

In PRODYN, the state variables like queues are estimated by Bayesian technique, and an on-line algorithm to estimate turning movements and saturation flow rates based on data from existing magnetic loop sensors is incorporated in the system (Kessaci, 1989)

Like OPAC and PRODYN, Italian system UTOPIA also adopted a dynamic programming approach with a rolling horizon framework (Mauro, 1990).

OPAC, PRODYN and UTOPIA showed good promise as an element of the real-time traffic responsive adaptive control, but there are some limitations.

- They all can only carry out decentralized control. The approach can not be applied in centralized control architecture.
- If more improvements are going to be made to fit the control strategy into the centralized control, the dynamic programming methodology must be abandoned since the "Curse of Dimensionality of DP" as a result of increased control variables will make the DP very ineffective and hardly feasible for any type of real-time application.
- The coordination of traffic within network is not well dealt with.
- The delay models are very simple. Delay is estimated and evaluated by queue length in stead of vehicle-time.
- No congestion restriction measures are considered.
- The constraints in reality might make the strategy less efficient and less effective.
- The turning ratios at intersections are not well accounted for, which might affect the results of optimal policies.
- Very limited field tests have been conducted to prove the advantages of the control strategies.

2.4.4 Other On-line Systems

The other on-line tests and systems include:

- PBIL on-line test in Aachen, Germany. It is a control strategy of one critical intersection plus the related intersections around it. PBIL tries to minimize lost-time that is defined as over-travel-time in the network compared with the normal travel time. (OCED, 1983)

- French system GERTRUDE. The GERTRUDE was specially designed for areas with congestion. The program arranges the splits in such a way that the critical intersection has the maximum throughput at the expense of links. (OCED, 1983)
- RTOP test and implementation in Toronto, Canada. It uses a prediction logic developed for use in the ASCOT (Adaptive Signal Control Optimization Technique, a system evaluated in San Jose). Prediction of traffic is based on historical volume, current volume from detectors, and directional relationship. Splits, cycle lengths and offsets are optimized based on the volume prediction.
- CYRANO of U.S. It was evaluated in Washington D.C., and also known as a 3-GC UTCS system (Luk, 1983).

Most of these on-line tests were not very successful. Evaluation tests even showed some system having consistent worsen performances than fixed timing control. Most of approaches, thus, have been abandoned.

RHODES is a prototype system under development in Tucson, Arizona. It is a hierarchical control also using dynamic programming approach for optimization (Head, 1992).

A TRUSTS (Traffic Responsive and Uniform Surveillance Timing System) has been developed in Taiwan (Tsay, 1989). It generates on-line timing plans using a modified TRANSYT-7F approach, provides timing table selections and time-of-day plans as well. Expert system was used to determine the proper selection of timing plans. Success of the evaluation has been reported.

On-line traffic responsive adaptive control by intuition has advantages over pretimed traffic control. However, inconsistent test results have been reported for many systems. One of the important reasons of the inconsistent performance results from

the unreality and inaccuracy of the traffic prediction models, the difficulties in obtaining and using the detected traffic data, and the problems with reliability and accuracy of the detectors. Besides, the surveillance device requirements increase the system cost.

2.5 INTEGRATED CORRIDOR CONTROL

Integrated corridor traffic responsive strategies (including ramp metering and surface street signal optimization) have been proposed in recent years. Most of them have been grounded on the optimal control theory that usually leads to a large-scale non-linear optimization problem.

Cremer and Schoof (1989) first formulated the integrated control problem. In the model, freeway section flow dynamics is simulated by a set of difference equations. Figure 2-3 shows the freeway section (with subsection 1,2, J) and model variables.

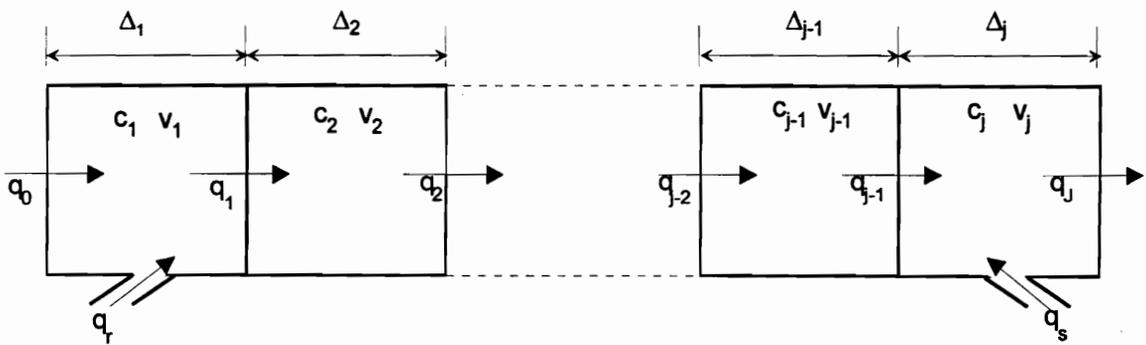


Figure 2-3. Section of Freeway with Sub-sections and Model Variables

From a simple balance of vehicles, the following difference equation holds for the density c_j .

$$c_j(k+1) = c_j(k) + \frac{T}{\tau} [q_{j-1}(k) - q_j(k) + \delta_{ij}q_r(k) - \delta_{js}q_s(k)] \quad (2-18)$$

A corresponding difference equation for the mean speed v_j was derived from empirical considerations:

$$v_j(k+1) = v_j(k) + \frac{T}{\tau} [V(c_j(k), u_2(k)) - v_j(k)] + \frac{T}{\tau} [v_j(k)(v_{j-1}(k) - v_j(k))] - \frac{v}{\tau} \frac{T}{\Delta_j} \left[\frac{c_{j+1}(k) - c_j(k)}{c_j(k) + \chi} \right] \quad (2-19)$$

where

$c_j(k)$ traffic density in subsection j at time kT (veh/km)

$v_j(k)$ mean speed of the vehicles with subsection j at time kT (km/h)

$q_j(k)$ volume for subsection j into subsection $j+1$ during the time interval $kT \leq t \leq (k+1)T$ (veh/h)

$q_r(k), q_s(k)$ entering or leaving ramp volumes during $kT \leq t \leq (k+1)T$ (veh/h)

α, τ, χ, v parameters calibrated from real measurements

T temporal step width

$V(c, u_2)$ average speed of the stationary speed density function, a formula developed by Zackor in 1972.

$$V(c, u_2) = v_f u_2 \left[1 - \left(\frac{c}{c_{max}} \right)^{l(3-2u_2)} \right]^m \quad (2-20)$$

where

u_2 is a normalized control variables with values from 0.5 - 1 with regard to different speed limits

c_{max} maximum density

v_f free flow speed

l, m real positive numbers

For the volumes between the subsections the following approximate expression taken from hydro-dynamical relationship was used.

$$q_j(k) = \alpha c_j(k) v_j(k) + (1 - \alpha) c_{j+1}(k) v_{j+1}(k) \quad (2-21)$$

Ramp metering by control u_1 is modeled by the following expression

$$q_r(k) = \min \left\{ \begin{array}{l} \gamma f(c_{1(k)}) \\ f(\gamma u_1(k)) \\ q_d(k) + q_w(k) \end{array} \right\} \quad (2-22)$$

In the formula, the entering volume is either limited by congestion on freeway ($f(c_1)$), or by green time u_1 of ramp signal or the number of waiting (q_w) and arriving (q_d) vehicles.

For leaving ramp volume, a normal portion and additional diverting portion were considered.

$$q_s(k) = \varepsilon q_j(k) + q_a(k) \quad (2-23)$$

The remaining main stream volume was given by

$$q_j^o(k) = (1 - \varepsilon) q_j(k) - q_a(k) \quad (2-24)$$

where

$$q_a(k) = u_3(k) \varepsilon_a (1 - \varepsilon) q_j(k) \quad (2-25)$$

On the surface street, TRANSYT's Platoon dispersion model was used and control variables were selected as green times at each intersection.

$$u_4^n = (g_1^n, g_2^n, g_3^n, \dots, g_m^n) \quad (2-26)$$

Green sequence was also optimized, no common cycle length was used.

Delay was used as an evaluation criterion. Total delay was calculated as a function of four model variables taking a sum of four contributing terms:

1) momentary delay time in section j of the freeway

$$L_1^j(k) = \Delta_j c_j(k) \frac{T}{v_f} (v_f - v_j(k)) \quad (2-27)$$

2) delay time from waiting queues on i -th ramp

$$L_2^i(k) = T^2 q_w^i(k) \quad (2-28)$$

3) delay time caused by driving on the deviation route

$$L_3(k) = T q_a(k) \left(\frac{l_a}{v_a} - \frac{l_m}{v_f} \right) \quad (2-29)$$

4) delay time caused by waiting time in the j -th road section

$$L_4^j(l) = T_s^2 r_j(l) \quad (2-30)$$

The total delay was determined taking the integral over time and space

$$P = \sum_k^K [\sum_j^J L_1^j(k) + \sum_i^I L_2^i(k) + L_3(k)] + \sum_l^L \sum_j^{J_s} L_4^j(l) \quad (2-31)$$

The formulated problem is a mixed integer nonlinear system, and a two layer optimization approach was proposed. On the upper level, a decision was made for diversion optimization while on the lower level ramp metering with speed limitation for freeway and surface street signaling were optimized independently. A heuristic search method was used for green time optimization

The case study results showed some benefits of congestion reduction of the corridor and reduced performance index (delay). However, the proposed approach has the following disadvantages:

- The control variables are not able to be optimized simultaneously. By using two layer optimization, global optimization is not guaranteed
- For surface street optimization, optimization of green phase lengths is made over the entire time horizon, which is unacceptable for:
 - (1). Too many variables, very computation intensive.
 - (2). Accuracy of prediction of traffic arrival worsens when prediction period becomes longer.
 - (3). The search strategy (Fix a green pattern, define a set of direction, random search, each search requires several simulation runs (TRANSYT-7F Run)) is very time consuming and makes the real-time application impossible
- Coordination of signals on surface is not considered.
- The system involves many parameters. Calibration of them requires much effort.
- Overall, the formulated system is too complicated and too computation intensive. Real time application is almost impossible.

Chang formulated a simplified control problem for the corridor integrated control (Chang, 1994). In the model, similar difference equations to Cremer's representing freeway and ramp traffic dynamics were used. But several simplifications were made:

- The distance from ramp to surface street intersection was negligible.
- For surface street signals, control variables used were g/c (green cycle ratio) instead of green splits and cycle length.
- The flow-density relation was approximated with a two-segment linear function.

With the approximations, the successive linear programming approach was proposed to determine the optimal control measures (on-ramp, off-ramp metering rates, and g/c ratios for street signals) with the time varying demand pattern and incident information. The system parameters were predicted based on some prediction models. In case study, INTRAS, a microscopic simulation Software developed by FHWA was used to simulate traffic data obtained by surveillance systems in reality.

Some disadvantages of this approach are list below:

- Simplified and identical freeway, ramp structure was used for each segment, which might not be good for real world scenario applications.
- Distance from intersection to on-ramp, from off-ramp to intersection is neglected.
- Two-phase intersection timing at off-ramp and intersection is over simplified.
- Only g/c ratios instead of green splits and cycle length were used as control variables.
- Prediction models can not predicted unexpected future flow accurately especially for arterial traffic, not to mention incident conditions.
- The need for surveillance and detecting equipment on the freeway, ramps, and surface streets will result in significant cost for real-time implementation.

Papageorgious (1980,1989), Goldstain and Kumar (1982), Gartner and Reiss (1992) studied freeway corridor control using similar simulation models based on a series of difference equations of flow dynamics. Most of these studies concentrated on freeway mainstream control, and some also addresses the ramp control measures. However, surface street signal timing was not the main concern. Modeling and simulation conducted by Papageorgious (1989), and Pooran (1994) showed the benefits of ramp metering and signal coordination. More research is need in this area.

2.6 SUMMARY OF LITERATURE REVIEW

Off-line pretimed signal control based on historical volume enjoys success under normal traffic demand level, but failed to handle unexpected demand level and traffic pattern changes, therefore, they are not suitable to provide effective control for incident situation.

On-line models for isolated intersection perform better than pretimed strategies for undersaturated conditions but worsen in performances as demand level increases.

On-line network control models showed inconsistent test results. Many of those models suffered from the inaccuracy and unreality of prediction models. Besides, high implementation costs associated with extra surveillance and traffic detecting equipment might make the gain from the control strategies not worthwhile.

Not much research has been done in the field of integrated corridor control. The gain and loss for model complexity, reality, and emphasis on different aspects should be weighed carefully in the future research to seek optimal system performances.

3. MODEL FORMULATION

3.1 INTRODUCTION

Many busiest traffic networks consist of a corridor with freeway sections and signalized arterials. During the peak periods, the congestion on one component of the network often spills over to the other components. Effective traffic operation control strategies should take into account all components of the network.

More than half of the congestion and delay are caused by incidents. Severe incidents occurred on freeway sections significantly reduce freeway capacity, thus incurring significant congestion and delay. In situations where an adjacent arterial is available, diversion is often a logical consideration, moreover, a viable solution.

To determine diversion routes and rates, to find out the appropriate ramp metering rates at on-ramps and off-ramps, and to retime arterial traffic signals are among the most important issues of developing proper diversion strategies. They need to be studied to achieve the best performance of the corridor traffic operation. All those control decisions are inter-related and should be made simultaneously considering each other from an integrated system point of view. This necessitates the development of an integrated corridor control strategy.

This chapter deals with the development of a nonlinear optimization model for integrated corridor control. It is organized to first present the background of the program. This is followed by a description of the proposed approach. Then the model formulation is presented.

3.2 BACKGROUND

Control models for traffic signal networks and freeways have shared success to some degree in the past. For example, traffic signal optimization programs and packages for isolated intersections (Highway Capacity Software), coordinated arterials (PASSER), or area-wide networks (TRANSYT-7F) have been developed and successfully applied to generate timing plans (fixed-time). SCOOT and SCAT have contributed to the implementation of the real-time adaptive signal control of signalized networks. Similarly, freeway simulation models and packages such as FRESIM, FREQ have been employed to develop control strategies for freeway traffic operation (Van Aerde, 1989). The studies, model development, and simulation/optimization packages for the surface network intersection signal control and freeway operation have been proved to be useful and successful tools to analyze respective components of the corridors.

However, most of the control models for traffic signal, ramp metering, and freeway operation were developed independently. Consequently, the integrated nature and interactions of traffic flow between different components of the corridor (freeway sections, arterial intersections, and ramps) are not well addressed when the isolated models are used to analyze and develop control strategies for their respective components. Conflicts often exist when an "optimal" control strategy for one component lead to congestion or adverse operation condition of the other components of the system. The situation worsens when severe incidents occur in some sub-networks or corridor components that results in dramatic temporal and spatial changes of traffic patterns of the entire corridor. Integrated control models that can coordinate and optimize the overall traffic operation of the entire network systems are thus inevitable.

Integrated traffic control has been defined as the joint control of more than one different traffic subsystem, each of which could act to the detriment of the others if they were controlled independently (Yagar, 1983). Such integration could possibly include any combinations of signalized intersections of arterials or surface networks, freeways, service roads, ramps, busways, light rail transit systems. For integrated corridor control, the main components of interest are: freeway sections, signalized arterial, ramps (including on-ramps and off-ramps).

Integrated corridor traffic control models should address the control measures for all the components including traffic operation on freeway sections, arterial and intersection signals, on- and off ramps simultaneously. For the situation of having incidents occurring on freeway sections, diversion strategies including diversion routes and rates, and arterial intersection signal retiming should be considered jointly. In the past, only limited studies have attempted to deal with this control philosophy. Most of the studies, as discussed in the previous chapter, are either simulation based approaches that normally require multiple lengthy simulation runs, or optimization models that often lead to large scale nonlinear models that demand an extreme amount of computation where global optimum is yet not guaranteed. The computation intense nature of those approaches limited the promise of successful development of quick response control strategies and real-time control implementation. More practical and less computation intense models and algorithms need to be developed for optimal integrated corridor control.

3.3 APPROACH

The goal of the research is to develop a model and algorithm to obtain optimal control measures (diversion routes, diversion rates, and intersection signal timing plans) in order to achieve system optimal performance for the corridor composed of

a freeway, ramps, and an arterial, under the circumstance that an incident occurs on the freeway and diversion is needed.

An off-line control approach is adopted in this research. For a corridor with severe incident occurring on a freeway section and diversion being implemented, near-saturated, or saturated traffic condition prevails in the entire system, on-line control has been reported and proved to be ineffective due to the difficulties resulting from traffic detection and prediction for saturated condition. The off-line approach does not require detectors installed that are essential for on-line control. On most of the surface streets and freeways in the U.S., the traffic detection devices suitable for on-line adaptive traffic control are not available; moreover, the installation of loops or other detectors are very expensive. On the other hand, historical traffic volume counts on freeway sections and surface streets are often available and far less expensive to obtain. An off-line control approach will make the control strategy more economical, practical, and promising for implementation. Nevertheless, the model will have the time-varying nature to account for change of traffic patterns over time due to changing demand levels for different time periods of the day, and more important, due to the dynamic clearance process of freeway lanes blocked by incident.

In the literature of diversion and alternate route development, (single) shortest path approaches prevail. For the corridor with some freeway sections blocked partially or fully, the shortest alternate route to bypass the incident is simply the route that starts from the immediate upstream off-ramp, through the parallel arterial, and ends at the immediate downstream on-ramp (of the blocked freeway section). It might not be a good philosophy to implement this shortest path diversion strategy, however, for the corridor of our interest. The reasons are that the capacity of the on-ramp and off-ramp is limited by its physical structure, and the traffic signal of the first intersection from the off-ramp and the last intersection before the on-ramp may also further limit

the extra traffic (diverted traffic) because the approaching traffic demand of the diversion traffic approach is much higher than that of any other approaches, and there should be some limits on the length of the particular phase designated for the approach of the diversion traffic.

The philosophy of diversion adopted in the model development is that more than one upstream off-ramps and downstream on-ramps should be used to divert traffic. That is to say that multiple diversion routes will be employed, the routes use different on-ramps, off-ramps and different portions of the arterial. With the traffic being diverted from the freeway to the arterial, traffic patterns and volumes at the arterial intersections are changed. The arterial signal retiming is necessary to better accommodate the changed traffic status.

The model will try to determine the following issues:

- Which upstream exit ramps (off-ramps) will be used to divert traffic from the freeway?
- Which downstream entrance ramps (on-ramps) will be used to divert traffic back to the freeway?
- What are the appropriate diversion rates at all the ramps on the diversion routes?
- Consequently, how should the signal timing plan at each intersection be re-set?
- How can the arterial intersection signals be synchronized if needed?

Many network or corridor traffic simulation models are based on Origin-Destination (O-D) trip tables to generate vehicles and traffic. The O-D tables are normally very expensive to obtain, and in many sub-networks they are not available. Moreover, the reliability of the tables when available is often questionable. On the other hand, link traffic volume counts are less expensive to acquire, and are often available on a historical time-of-day basis. In the model the traffic generation and movement in the

network will be derived from link volume counts instead of generating from O-D Tables.

Most well known traffic engineering models, such as TRANSYT-7F and HCS, use macroscopic traffic models, while NETSIM and FRESIM provide microscopic performance models. However, these models were developed without the concepts of routing, signal retiming, and other Advanced Traffic Management Systems (ATMS), and Advanced Traveler Information Systems (ATIS) control mechanisms. INTEGRATION falls between microscopic and macroscopic: the individual vehicles are modeled, but their movement is governed by macroscopic flow equations (Van Aerde 1988, 1989, 1994). THOREAU adopts a microscopic and object-oriented approach, the models are suitable for ATMS and ATIS evaluation (Codelli 1992, 1993). However, some of the ATMS and ATIS routines such as diversion and signal retiming are not implemented in THOREAU. In this research, the traffic model takes a macroscopic approach instead of a microscopic one that keeps track of each individual vehicle. By taking a macroscopic approach, the developed model will be simple and practical, and it will be easier to develop efficient solution algorithms.

A uniform and consistent performance index/measure of effectiveness will be used for evaluate the operation of arterial intersections, and freeway sections as well. In terms of model formulation, this means the same form of objective function/cost function/performance index will be used for the different components of the corridor.

The proposed approach and model will have the following features:

1. The control variables of the model are signal timing plans at each intersection represented by green times of each approach, the diversion rates at each on-ramp and off-ramp. If the diversion rate at any particular ramp is determined to be non-zero, that ramp is on one of the diversion route. The diversion routes are reflected by the diversion rates at ramps rather than explicitly presented. The

ramp metering is modeled as a signalized intersection, the green times are determined based on normal rates and diversion rates at ramps.

2. The time-varying nature of the network traffic and dynamic clearance process of incident lanes are considered. The control variables are optimized over several time periods for which a particular traffic pattern prevails and remains unchanged.
3. The effects of queue on performance index and signal coordination are considered.
4. Congestion reduction measures are taken by limiting the queue length to avoid spillback and by penalizing queues that can not be cleared after green time is over for that approach.
5. The prediction of future mainstream arrivals is based on the upstream departures and signal timing
6. Arrivals of approaches other than mainstream are decided by historical data. Time-dependent variation can be considered.
7. Performance indices are evaluated for the entire corridor including the freeway, ramps, and the surface arterial. Same form of the objective function/performance index is used for all components.
8. A mathematical programming algorithm will be developed according to assumptions, constraint formulations, and objective function selection.
9. The optimal control model and strategy is developed with simulated data or historical data based on reasonable assumptions so that effective control strategies can be obtained without significant implementing cost increase. The real-time traffic data can be also used in the system optimization process if the data is available.

3.4 MODEL FORMULATION

3.4.1 Assumptions

The following assumptions will be adopted for the proposed approach and model development:

- 1) A common cycle length is used for all intersections of the arterial within the corridor of interest. This assumption is made because of similar traffic volumes and patterns at intersections. In addition, utilizing a common cycle length makes coordination of mainstream flow possible.
- 2) A fixed green sequence is used for each intersection signal. Optimization of phase sequence provides some advantages for under-saturated traffic conditions. By selecting the proper green sequence or skipping some phases according to the traffic flow, better system performances can be achieved. However, under near-saturation or saturated conditions, changing the green sequence or skipping phase will result in a prohibitively long delay and queue lengths for some approach(es), and cause driver confusion as well.
- 3) Control variables are optimized over some successive time periods. For an optimization period, the arrival patterns on the freeway and the arterial remain unchanged, and neither does the lane blockage situation. If there is any change in any of these factors, a new optimization period should be assigned.
- 4) Traffic is diverted to the arterial at the off-ramp just before the incident section and other off-ramps further upstream as well; Traffic is diverted back to the freeway from on-ramp right after the incident section and from on-ramps further downstream if necessary. The model will determine which on-ramps and off-ramps will be involved in diversion.
- 5) The delay at on-ramps and off-ramps is neglected, and only stop delay on the freeway and arterial intersections is considered.

- 6) For an arterial intersection, the arriving volume on mainstream (the direction of the diverted flow) is determined by normal flow superimposed with the diverted volume from upstream off-ramps (if exists), as well as from the off-ramp that directly leads to that intersection.
- 7) On the opposite direction of mainstream, normal traffic volume is used.
- 8) Normal arriving volumes from side streets and their turning movements are assumed to be either unchanged, or the change patterns are known.
- 9) For freeway sections, only unidirectional traffic flow is considered. The traffic flow in the other direction without incident and lane blockage is regarded as unaffected. However, for arterial traffic, all approaches are considered. Traffic in both directions of the arterial, as well as the traffic movements to/from side streets, all play roles in determining proper signal timing plans.
- 10) Due to the incident on the freeway, the traffic normally gets onto the freeway via on-ramps (if there is no incident on the freeway) will stay on the arterial, moves in the mainstream direction, and gets on the freeway only after the blocked freeway intersection.
- 11) All "normal" variables (volume, turning movements, etc.) denote the values under non-incident conditions. They are system inputs (known) and can vary over time. Real time values will be used if available. Or, fixed rates based on historical data or predicted rates will be used for different time periods.

3.4.2 Traffic Flow Model

A macroscopic traffic model is developed to simulate the characteristics of the traffic flow in the system. A macroscopic traffic model does not keep track of the movement of any individual vehicle in the system, the temporal and spatial characteristics of the traffic are described at an aggregate level. Traffic arrivals and departures for the approaches at arterial intersections and freeway sections, queue

lengths and other measures of effectiveness are all determined at an aggregate level by considering the average performance of the flow.

3.4.2.1 Notations

As previously stated, control variables are optimized over some successive time periods. For each optimization period, the arrival patterns on the freeway and the arterial, and lane blockage situation remain unchanged. A particular time period may last the length of several cycles (plus the possible incomplete cycle at the beginning or end of the period). For each period, the arrival and departure patterns are assumed to be stable for each cycle length.

The following notations for one cycle will be used in the formulation of the models.

- NI total number of intersections on the arterial
- i index for the arterial intersection, and for the interchange on the freeway as well
- δ^i constant indicates the intersection and interchange matching status:
 $\delta^i = 0$, if there is an off-ramp i directly leading to the intersection i , or an on-ramp directly from intersection i , or
 $\delta^i = 1$, otherwise
- P_i total number of phases for a signal cycle at intersection i ,
 $i = 1, 2, \dots, NI$
- M_p^i total number of approaches (movements) for phase p , at intersection i ,
 $p = 1, 2, \dots, P_i; i = 1, 2, \dots, NI$
- m_p^i approach index for phase p , at intersection i ,
 $m_p^i = 1, 2, \dots, M_p^i; p = 1, 2, \dots, P_i; i = 1, 2, \dots, NI$
- L lost time for each green period (sec)
- G_p^i green time for phase p^i , (sec)
- G_{min} minimum green time (sec)

G_{max} maximum green time (sec)

g_p^i effective green time for phase p^i (sec)

$$g_p^i = G_p^i - L$$

$NL_{m_p}^i$ Number of lanes for approach m_p^i

C cycle length (sec)

$N_{m_p}^i$ arrival rate for approach m_p^i at normal condition (veh/sec)

$VRIn^i$ arrival rate into intersection i from ramp going toward mainstream direction at normal condition (veh/sec)

$VROn^i$ rate leaving intersection i from mainstream direction to on-ramp at normal condition (veh/sec)

DVR^i rate into intersection i from off-ramp going toward mainstream direction under incident condition, or diversion rate (veh/sec)

$DVRO^i$ rate leaving intersection i from mainstream direction to on-ramp under incident condition, or, diversion rate (veh/sec)

F^i volume on the freeway section between ramp to intersection $i-1$ and i (veh)

$v_{m_p}^i$ volume arriving at intersection i , from approach m_p^i (veh)

$vn_{m_p}^i$ normal volume arriving at intersection i , from approach m_p^i , if there is no diversion (veh)

$xsg_{m_p}^i$ queue length for approach m_p^i at the start of green (veh)

$xeg_{m_p}^i$ queue length for approach m_p^i at the end of green (veh)

$xsm_{m_p}^i$ allowed maximum queue length for approach m_p^i at the start of green (veh)

$xem_{m_p}^i$ allowed maximum queue length for approach m_p^i at the end of green (veh)

xf_t queue length on freeway incident section at time t (veh)

$vd_{m_p}^i$ departure volume from approach m_p^i (veh)

VF rate arriving to the corridor on the freeway (veh/sec)

- VFD** departure volume of the freeway incident section (veh)
- IOFF** index of the intersection corresponding to the last ramp before freeway incident section
- ION** index of the intersection corresponding to the first ramp after freeway incident section
- lr** constant indicating the relative position of the parallel arterial to the freeway direction of incident occurrence:
lr = 1: arterial is to the left of freeway traffic movement of interest
lr = 2: arterial is to the right of freeway traffic movement of interest
- $D_{m_p^i}$ delay on arterial approach m_p^i (veh.sec)
- D_f total delay on the freeway (veh.sec)
- D_a total average delay on arterial (veh.sec)
- VT_f throughput on the freeway (veh)
- VT_a through put on the arterial (veh)
- ψ weight of freeway objective function component (constant)
- $\xi_{m_p^i}$ weight of arterial objective function component for movement m_p (constant)
- s** saturation flow at intersections (veh/sec)
- u_f free flow speed on freeway (mi/hr)
- CFA** available freeway capacity for corridor under incident (veh/sec)
- $c_{m_p^i}$ capacity of lane-group (arterial) for approach m_p^i (veh/sec)
- $x_{m_p^i}$ degree of saturation for approach m_p^i

3.4.2.2 Decision Variables

The goal of the developed model is to find the optimal control parameters of the corridor, mainly, signal timing plans at arterial intersections, and diversion routes and rates.

The following are the decision variables to be optimized by the model:

- G_p^i $p = 1, 2, \dots, P_i; i = 1, 2, \dots, NI$
- $DVRI^i$ $i = 1, 2, \dots, IOFF$
- $DVRO^i$ $i = ION, ION+1, \dots, NI$

Whether an off-ramp upstream of the incident location or an on-ramp downstream of the incident location is used for diversion can be determined by the values of $DVRI^i$ or $DVRO^i$.

If $DVRI^i > VRIn^i \quad \forall \quad 1 \leq i \leq IOFF$, off-ramp i is used in diversion. Otherwise, if $DVRI^i = VRIn^i \quad \forall \quad 1 \leq i \leq IOFF$, the off-ramp is not used for the diversion purpose.

Similarly, If $DVRO^i > VROn^i \quad \forall \quad ION \leq i \leq NI$, on-ramp i is used in diversion, otherwise, if $DVRO^i = VROn^i \quad \forall \quad ION \leq i \leq NI$, the on-ramp is not used for the diversion purpose.

3.4.2.3 Signal Configuration at Arterial Intersections

The model has the flexibility to handle different intersection geometric layout, different relative positions between freeway and arterial, as well as different signal phasing and lane group combinations. Figure 3-1 shows an arterial intersection i ,

with the typical lane groups and signal phasing. As illustrated, the intersection has 4 phases, each phase has 2 approaches of traffic with left turning movement denoted as approach 1, through and right turning as approach 2. The arterial is to the right of the freeway traffic movement direction in which incident is occurring. The subsequent formulations of arrivals, departures, and queues are developed for this typical freeway arterial configuration, arterial intersection layout, and phasing and lane group combinations.

3.4.2.4 Arrival

It is assumed that links that are not part of diversion routes carry normal traffic volumes during the incident period. The changes to traffic volumes on these links are known regardless of diversion volumes. While for all the movement on the diversion routes, the diverted volumes need to be considered when formulating arrivals.

Let NDR denote the set of the movements (approaches) that are not on the diversion routes, and let DR denote the set of the movements that are on the diversion routes. For the case that the parallel arterial is to the right of the freeway direction with incident occurring, movements on the diversion routes are the following:

- for the intersections upstream of the incident section, the left turn movements to the arterial from off-ramps, and the through movements on the arterial in the direction of diversion
- for the intersections downstream of the incident section, the left turn movements from the arterial to on-ramps, and the through movements on the arterial in the direction of diversion
- for the intersections within the incident section (if there is any), the through movements on the arterial in the direction of diversion

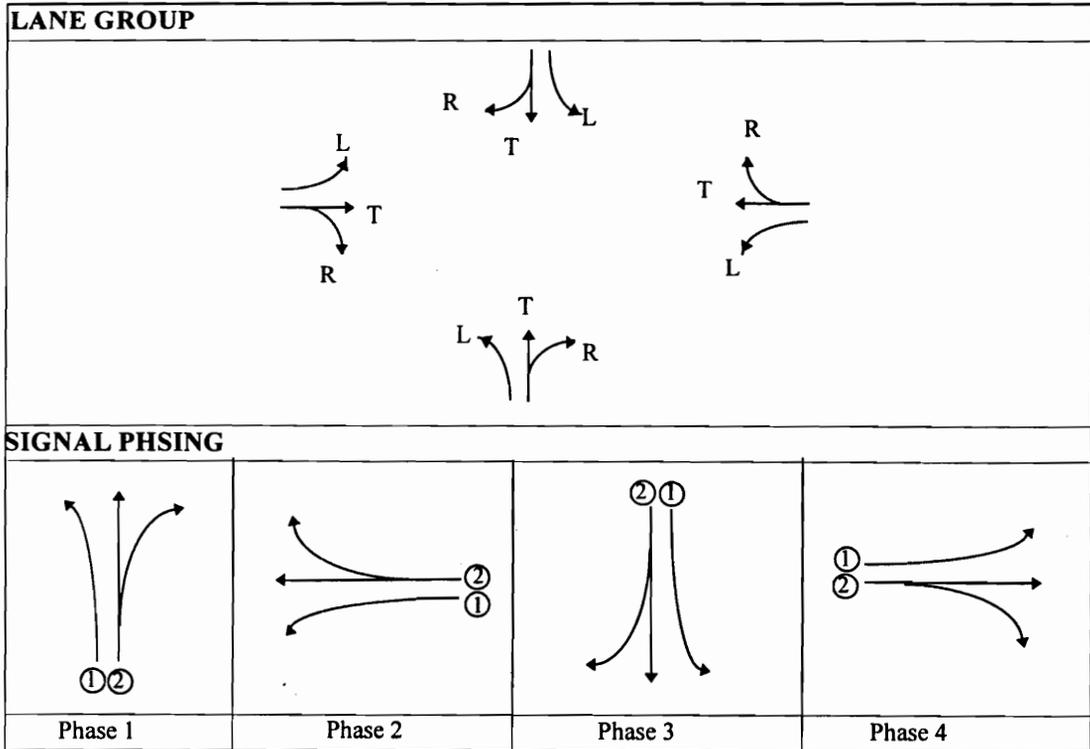
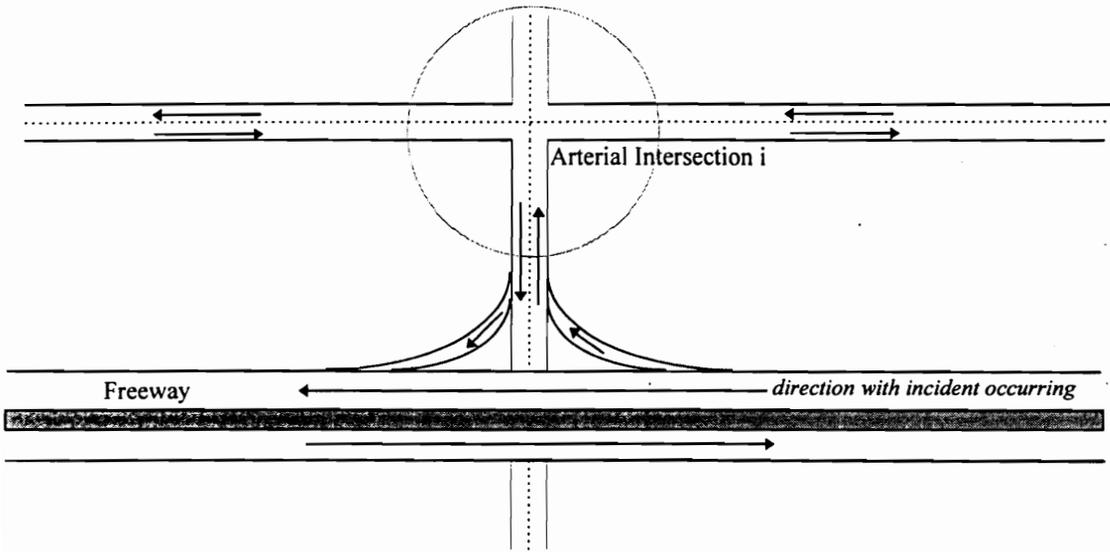


Figure 3-1. Typical Lane Groups and Signal Phasing for an Arterial Intersection

Expressed using the notation introduced early, we have the following:

$$DR = \begin{cases} \forall m_p^i, & \text{if } m_1^i = 1, m_2^i = 2, \text{ and } 1 \leq i \leq IOFF \\ \forall m_p^i, & \text{if } m_2^i = 2, \text{ and } IOFF < i < ION \\ \forall m_p^j, & \text{if } m_2^j = 1, m_2^j = 2, \text{ and } ION \leq j \leq NI \end{cases} \quad (3-1)$$

Accordingly, we have:

$$NDR = \begin{cases} \forall m_p^i, & \text{if } m_1^i = 2, m_2^i = 1, m_p^i = 1,2, \text{ and } 1 \leq i \leq IOFF, 3 \leq p \leq P_i \\ \forall m_p^i, & \text{if } m_2^i = 1, m_p^i = 1,2, \text{ and } IOFF < i < ION, p = 1,3 \leq p \leq P_i \\ \forall m_p^j, & \text{if } m_1^j = 2, m_2^j = 1, m_p^j = 1,2, \text{ and } ION \leq j \leq NI, 3 \leq p \leq P_i \end{cases} \quad (3-2)$$

Here, $m_1^i = 2$ is equivalent to the notation of 2_1^i , standing for the 2nd approach of phase 1 at intersection i . The notations $m_p^i = x$ and x_p^i are used interchangeably in the subsequent sections.

In Figure 3-2, the movements on the diversion routes are illustrated in dashed lines.

Arrivals Independent of Diversion

For all the movements $m_p^i \in NDR$, the arrival rates are independent of diversion volumes and rates, and are all treated as known inputs to the model. The normal arrival rate values can be used.

$$v_{m_p^i} = vn_{m_p^i} = VN_{m_p^i} * C \quad \forall m_p^i \in NDR \quad (3-3)$$

Or adjusted arrival rates ($N_{m_p^i}'$, also assumed to be known) can be used to replace normal arrivals ($N_{m_p^i}$) in formula (3-3). The adjustments are necessary if the changes of arrivals on these approaches become significant due to the fully or partially blockage of freeway sections.

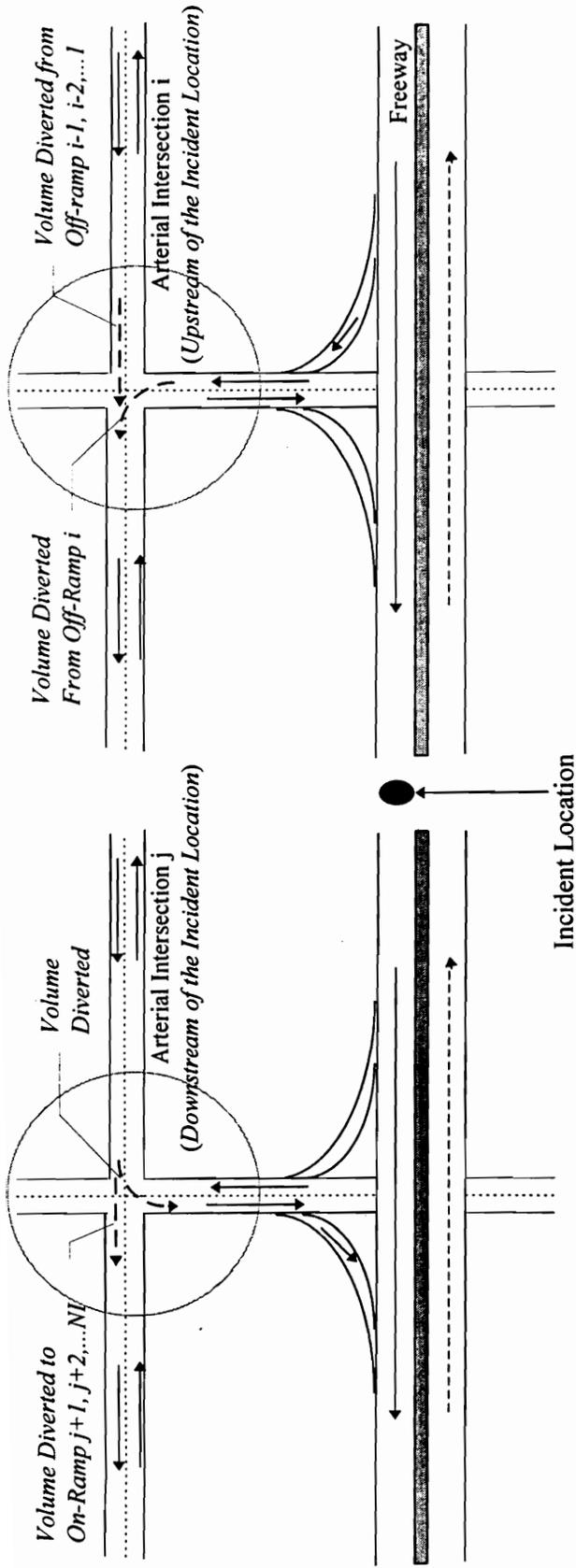


Figure 3-2. Illustration of Diversion Routes and Traffic Movements

Arrivals Affected by Diversion

Arrival volumes on the diversion routes will be affected by the diversion rates on those routes. The arrival rates are determined by superimposing extra diversion traffic onto the normal traffic. Based on this assumption, the arterial arrival volume in a cycle for movements on the diversion routes is formulated below.

For arterial intersections upstream of the freeway incident section, arrivals of left-turning movements from off-ramps to the arterial and through approaches in the direction of diverted traffic are on the diversion routes thus considered to be affected by diversion rates.

$$v_{1_i}^i = DVRI^i * C \quad \forall \quad 1 \leq i \leq IOFF \quad (3-4)$$

where

$$DVRI^i = VRIn^i + \Delta VR^i \quad (3-5)$$

ΔVR^i is the extra diversion rate (from the freeway through off-ramp i)

$$v_{2_2}^i = (VN_{2_2}^i + \sum_{j=1}^{i-1} \Delta VR^j * \delta^j) * C \quad \forall \quad 1 \leq i \leq IOFF \quad (3-6)$$

For arterial intersections within the range of the freeway incident section, arrivals of the through movements in the direction of diverted traffic are on the diversion routes and are considered to be affected by diversion rates.

$$v_{2_2}^i = (VN_{2_2}^i + \sum_{j=1}^{IOFF} \Delta VR^j * \delta^j) * C \quad \forall \quad IOFF < i < ION \quad (3-7)$$

For arterial intersections downstream of the freeway incident section, arrivals of left-turning movements from the arterial to on-ramps and through movements in the

direction of diverted traffic are on the diversion routes thus considered to be affected by diversion rates.

$$v_{2_1}^i = DVRO^i * C \quad \forall \quad ION \leq i \leq NI \quad (3-8)$$

where

$$DVRO^i = VROn^i + \Delta VR^i \quad (3-9)$$

ΔVR^i is the extra diversion rate (back to the freeway through on-ramp i)

$$v_{2_2}^i = (VN_{2_2}^i + \sum_{j=1}^{IOFF} \Delta VR^j * \delta^j - \sum_{j=ION}^i \Delta VR^j) * C \quad \forall \quad ION \leq i \leq NI \quad (3-10)$$

For the freeway, the arrival demand of the incident section (bottleneck) is the normal demand minus diverted traffic (through off-ramps to the arterial). The arrival during the period of a cycle length is given by Equation (3-11).

$$F^{IOFF} = (VF - \sum_{j=1}^{IOFF} \Delta VR^j * \delta^j) * C \quad (3-11)$$

where $DVR^j = DVRI^j - VRIn^j \quad j = 1, 2, \dots, IOFF$

Arrival formulations for the scenario in which the arterial is to the left of the freeway direction with incident occurring, or there are 2 arterial, one on each side of the freeway are different. The major differences are changes of the movements in sets DR and NDR. For example, if arterial is to the left of the freeway direction with incident occurring, the diversion from off-ramps to arterial are right turns and $m_1^i = 2 \in DR$, and, diversion back to on-ramps are left turns, i.e., $m_2^j = 2 \in DR$. Nevertheless, the same principles are followed to formulate all arrivals.

3.4.2.5 Departure

The departure volume of an approach m_p^i in a given cycle is determined by the discharging capacity s , the effective green time $g_{m_p^i}$, the arrival volume, and the number of vehicles in the queue at the stop line at the beginning of the cycle.

For an arterial intersection approach, if there is enough upstream traffic feeding, the departure volume is determined by the effective green time and discharging capacity of the approach of the lane group. Otherwise, the departure volume will be the vehicles initially in the queue at the stop line at the beginning of the green, plus the vehicles arriving when the approach is given green signal.

$$vd_{m_p^i} = \min \left\{ \begin{array}{l} s * NL_{m_p^i} * g_{m_p^i} \\ xsg_{m_p^i} + (v_{m_p^i})_g \end{array} \right. \quad (3-12)$$

$$g_{m_p^i} = G_{m_p^i} - L \quad (3-13)$$

where

$g_{m_p^i}$ is the effective green time for approach m_p^i ,

$G_{m_p^i}$ is the green time for approach m_p^i ,

L is the lost time due to startup delay and green time lost at the end of the green time, and

$(v_{m_p^i})_g$ is the arrival volume for approach m_p^i during the current green period.

For the freeway incident section, the departure volume is limited by the available capacity of the bottleneck section.

$$VFD = CFA * C \quad (3-14)$$

3.4.2.6 Queue

The queue length is represented in number of vehicles in the queue. For arterial intersections, the queue length at the beginning of the green for approach m_p^i is the queue length at the end of last green for this approach plus the arrival during the period the approach is not given the green. The queue length at the end of the green for approach m_p^i is the queue length at the beginning of the green, plus the arrival volume during the period when the approach is given green, minus the departure volume. i.e.,

$$xsg_{m_p^i} = (xeg_p^i)' + (v_{m_p^i})_r \quad (3-15)$$

$$xeg_{m_p^i} = xsg_{m_p^i} + (v_{m_p^i})_g - vd_{m_p^i} \quad (3-16)$$

where

$(xeg_p^i)'$ is the queue length at the end of last the green for approach m_p^i ,

$(v_{m_p^i})_r$ is the arrival volume of approach m_p^i during the period when green is not given, and

$(v_{m_p^i})_g$ is the arrival volume of approach m_p^i during its green period.

For the freeway bottleneck section, the queue length is formulated as follows:

$$\begin{aligned} xf_{t_1} &= \int_{t_0}^{t_1} (VF - \sum_{i=1}^{i=IOFF} \Delta VR^i * \delta^i - VFD) dt \\ &= xf_{t_0} + (VF - \sum_{i=1}^{i=IOFF} \Delta VR^i * \delta^i - VFD) * (t_1 - t_0) \end{aligned} \quad (3-17)$$

where

xf_{t_0} queue length at time t_0

xf_{t_1} queue length at time t_1 , and

$VDF=CFA$

3.4.3 Objective Function

3.4.3.1 Objective Function Selection

In the literature of traffic signal timing optimization, a number of different objective functions/performance indexes/cost functions have been used, with the following being the most popular ones among the others.

- number of stops
- queue length (represented by the number of the vehicles in the queue)
- delay (in vehicle-hours)
- travel speed
- throughput
- fuel consumption

Some optimization models use operating cost or weighted combinations of several parameters listed above.

All those parameters are related to certain degree, but are not necessarily equivalent or interchangeable. The selection of the objective function should be based on the traffic condition. For light traffic conditions, minimizing stops is highly desirable; in medium traffic conditions, minimizing overall delay is most appropriate; while in congested condition, maximizing throughput seems to deserve more attention.

The operation performance of traffic flow on both the arterial and the freeway should be considered for the corridor of interest. The objective of corridor control for severe freeway incident case is to divert as much traffic via the arterial so as to increase the throughput of the corridor, and meanwhile, not over-congesting the arterial.

To achieve this goal, an objective function representing the integrated performance of the corridor is developed. It is in the form of minimization of the summation of the square of the difference between arrival volume and discharging capacity at each intersection approach and on the freeway bottleneck section, i.e.,

$$\begin{aligned}
 \text{Minimize } & \sum_{\forall \tau} \sum_{\forall \text{cycles in period } \tau} \left\{ \sum_{i=1}^{NI} \sum_{j=1}^{PI} \sum_{m=1}^{M_j^i} \left(\xi_{m_j^i} * \left(v_{m_j^i} - NL_{m_j^i} * s * (G_j^i - L) \right)^2 \right) \right\} \\
 & + \sum_{\forall \tau} \sum_{\forall \text{cycles in period } \tau} \left\{ \psi * \left((VF - \sum_{j=1}^{IOFF} (DVRI^j - VRIn^j) * \delta^j) * C - CFA * C \right)^2 \right\}
 \end{aligned} \tag{3-18}$$

where

$\xi_{m_j^i}$ weighting factor for approach m_j^i of the arterial intersection

ψ weighting factor for the freeway portion of the objective function

τ the time period during which arrival patterns and rates remain unchanged for all arterial approaches and the freeway (not considering the diversion factor). The total optimization time horizon is divided into several time periods and prevailing arrival patterns and rates are assumed for each period.

3.4.3.2 Discussion on the Objective Function

The proposed objective function (3-18) takes into account the performance of both the arterial intersections and the freeway section, it is in a form of the summation of

each individual element of the corridor of interest (the freeway section, or the arterial intersection). Basically, for each element, the objective function takes the form of the square of the difference between demand and capacity. For any intersection, the demand (arrival) of an approach is the linear function of the diversion rates at upstream ramps,

$$v_{k_j}^i = \begin{cases} f(DVRI^i, DVRO^i, DVRI^{i-1}, DVRO^{i-1}, \dots, DVRI^1, DVRO^1) & \text{if } m_j^i = k \in DR \\ \text{constant,} & \text{if } m_j^i = k \in NDR \end{cases} \quad (3-19)$$

where $v_{k_j}^i$ is the arrival demand evaluated at the end of the green time for approach $m_j^i = k, (or k_j^i)$, for a particular cycle. The capacity is the linear function of the effective green time assigned to the approach, g_j^i .

Similarly, for the freeway section, the demand function is the linear function of the diversion rates before the blocked freeway section, while the capacity is a constant for any optimization period τ .

It can be seen clearly that the objective function (3-18) is of the quadratic form of all the decision variables listed below:

- $G_p^i \quad p = 1, 2, \dots, P_i; i = 1, 2, \dots, NI,$
- $DVRI^l \quad l = 1, 2, \dots, IOFF, \text{ and}$
- $DVRO^l \quad l = ION, ION+1, \dots, NI.$

3.4.3.3 Justifications of the Objective Function

As discussed in 3.4.3.2, each element of the objective function is the square of the difference between arrival demand ARR and capacity CAP . For the scenario that demand exceeds capacity, $ARR-CAP$ actually stands for queue length (the number of

vehicles in the queue). For an arterial intersection approach, *ARR-CAP* stands for the queue length at the end of the green that is assigned for that approach. For the freeway section, *ARR-CAP* stands for the queue length at the time when an arterial cycle ends.

Queue length *ARR-CAP* has been used as an objective function as a form of delay for signal optimization. However, it only accounts for the number of vehicles in the queue, with no special consideration being given to the actual lengths of the queues. For example, by using this as the objective function, the following 2 cases are considered to be equivalent since they yield the same objective values.

- Case 1: For n intersection approaches, each has 1 vehicle in the queue.
- Case 2: For n intersection approaches, 1 approach has n vehicles in the queue, while the other $n-1$ approaches have no vehicle in the queue.

Apparently, Case 2 is a much worse scenario in term of traffic operation if n is considerably large. However, if we use $(ARR-CAP)^2$ as the objective function, the effect of long queue will be evidently considered since $n^2 \gg n$, when n is considerably large.

$(ARR-CAP)^2$ is a good objective function for the following reasons:

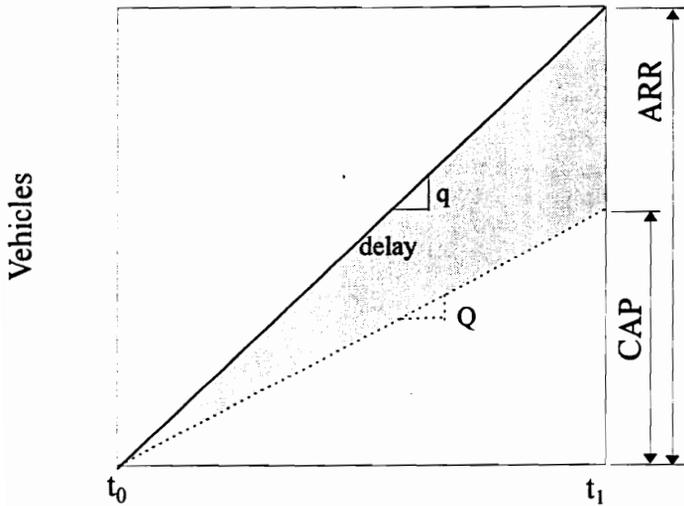
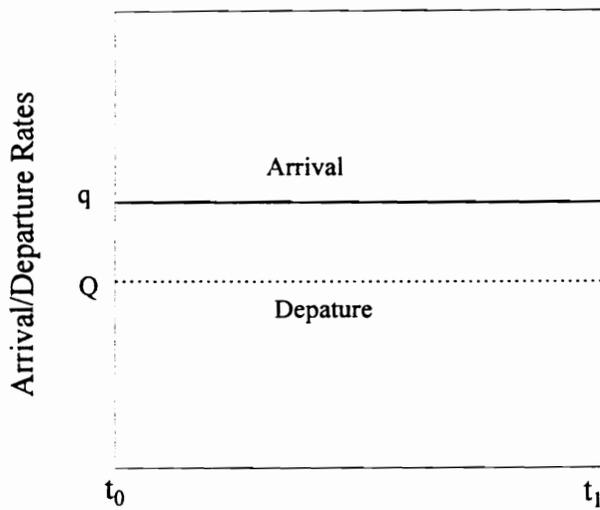
- 1) By using it as the objective function, extremely long queues will be penalized.
The effect of penalizing long queues is already shown in the previous simple example. Limiting long queues is a highly desirable feature, for extra long time of waiting in the queue should be always avoided. By limiting the queue length at the end of green, the chance of cycle failures due to over-saturation will be reduced.
- 2) It is a good representation of delay in a more conventional form (in vehicle.hours).

$(ARR-CAP)^2$ is a close approximation of delay (in vehicle.hours). For a freeway section where arrival rate and departure rate are constant. It can be easily seen that $(ARR-CAP)^2$ is related to delay in a linear fashion (Figure 3-3). For arterial intersections, the relationship between queue length and delay is not explicit. Nevertheless, the connection between queue length and delay is evident (Figure 3-4). Other previous researches showed that the square of the queue length reflects the delay (in veh.hours) very well (Heydecker, 1990).

The other advantage of using this form of objective function is that it is differentiable while the queue length formula is non-differentiable because of the departure formulation (equation 3-12). With the objective function being continuous and differentiable, more efficient solution algorithms are easier to obtain.

For an arterial intersection movement, if $ARR \gg CAP$, the approach in question is over-congested, and either less volume should be diverted from the freeway or more green time is needed for this phase. Increasing the green time by reallocating green times or reducing upstream diversion rates from the freeway will reduce the objective function value for the arterial approach. Meanwhile, the objective value for the freeway section could be increased (due to smaller diversion rates). Tradeoffs have to be made. This accounts for one of the most important reasons why the operation performance of the whole corridor needs to be considered in an integrated manner.

On the other hand, if $ARR \ll CAP$ for an arterial intersection approach, the green time assigned to the approach is not fully utilized. Either more volume should be diverted from the freeway to the arterial via upstream ramps, or the green time for this approach should be reduced. The objective function will automatically drive to increase diversion rates, or reduce green time. This will lead to better utilization of green time for the approach, assignment of more green time to other phases and



$$CAP = Q * (t_1 - t_0)$$

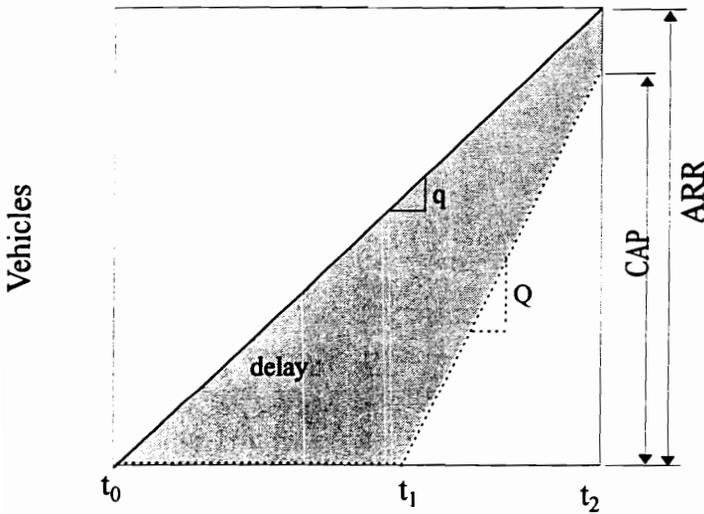
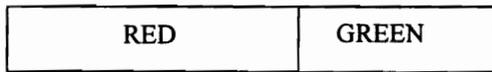
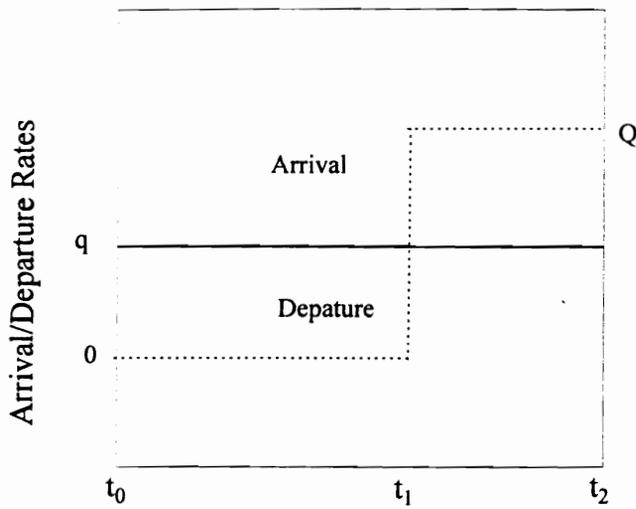
$$ARR = q * (t_1 - t_0)$$

$$(ARR - CAP)^2 = (q - Q)^2 (t_1 - t_0)^2$$

$$delay = 0.5 * (q - Q) * (t_1 - t_0)^2$$

$$\frac{(ARR - CAP)^2}{delay} = 2(q - Q)$$

Figure 3-3. Relationship between Square of Queue and Freeway Delay



$$CAP = Q * (t_2 - t_1)$$

$$ARR = q * (t_2 - t_0)$$

$$(ARR - CAP)^2 = [q(t_2 - t_0) - Q(t_2 - t_1)]^2$$

$$\begin{aligned} \text{delay} = & 0.5q * (t_1 - t_0)^2 \\ & + 0.5(t_2 - t_1) * [q * (t_1 - t_0) + \\ & q * (t_2 - t_0) - Q * (t_2 - t_1)] \end{aligned}$$

Figure 3-4. Relationship between Square of Queue and Delay at a Signal Intersection

approaches, or less queuing vehicles on the freeway section (due to increased diverted volume).

The ultimate goal of the model is to divert as much traffic as possible to bypass the freeway incident, and meanwhile not to over-congest the arterial intersections, and assign the green times to all phases in an optimal way. The objective function proposed can serve exactly that goal.

It should be also noticed that the objective function gives the incentive to fully use the green time for each approach. This is reasonable and desirable. The scenario we are dealing with is the traffic operation of a corridor with a severe incident on the freeway, normally the overall demand exceeds the capacity (even optimal control strategies are employed). Under this circumstance, fully utilization of available or potential capacity of all corridor components is highly desirable. Since all the corridor components are operating at their near capacity respectively, the levels of service are not high. This is the case for real-world scenario. The intersections normally operate at level of service D, or E.

The total delay is normally the most desirable objective function, but with the integrated interactions between the freeway and the arterial being considered in the objective function, the non-convexity nature of the function will make the optimization process inefficient, time consuming, and intractable. The total throughput along the main direction of the corridor is another important performance index. The model and algorithm consider the factors of total delay and the throughput when selecting and calibrating weighting factors for arterial intersections and the freeway section, ξ_{m_p} and ψ . This implementation issue will be discussed in the next chapter.

The proposed objective function is simple in form, reasonable from the traffic operation point of view, and mathematically tractable in optimization process.

3.4.4 Constraints

The following are the constraints of the model:

(1). Common cycle length

A common cycle length is used for all intersection signals. The utilization of a common cycle length makes it possible to synchronize signals on the arterial.

$$\sum_{p=1}^{P_i} G_p^i = C \quad i = 1, 2, \dots, NI \quad (3-20)$$

(2). Minimum and maximum phase lengths

Minimum phase length is set to provide adequate time for a certain number of vehicles to clear the intersection and for pedestrian approach. An upper limit of the phase length is adopted to eliminate extra long delay for other approaches that are not given green, and to avoid driver anxiety and confusion.

$$G_p^i \geq G_{min} \quad p = 1, 2, \dots, P_i, i = 1, 2, \dots, NI \quad (3-21)$$

$$G_p^i \leq G_{max} \quad p = 1, 2, \dots, P_i, i = 1, 2, \dots, NI \quad (3-22)$$

(3). Maximum on- and off-ramp metering rates

The capacity of the ramps is determined by the geometry, speed limit, and other physical features such as the number of lanes for ramps, and metering signals where applicable.

$$DVRI^i \leq VRI_{max}^i \quad \forall \delta^i = 1, i = 1, 2, \dots, IOFF \quad (3-23)$$

$$DVRO^i \leq VRO_{max}^i \quad \forall \delta^i = 1, i = ION, ION + 1, \dots, NI \quad (3-24)$$

(4). Minimum on and off-ramp metering rates

During the course of diversion, the metering rates at on- and off-ramps should maintain a minimum level to serve the normal traffic needs. i.e.,

$$DVRI^i \geq VRIn^i \quad \forall \delta^i = 1, i = 1, 2, \dots, IOFF \quad (3-25)$$

$$DVRO^i \geq VROn^i \quad \forall \delta^i = 1, i = ION, ION + 1, \dots, NI \quad (3-26)$$

(5). Maximum allowable queue length

The maximum allowable queue length is limited by the length of the link, or length of the left turn bay. Certain limits can also be set in the model to eliminate extra-long queues.

$$xsg_{m_p}^i \leq xsm_{m_p}^i \quad \forall m_p^i \quad (3-27)$$

Unlike the other constraints (equation 3-20 through 3-27) that are directly on the control variables and retain the same values throughout the optimization period, this constraint is on the queue length whose value changes from cycle to cycle within an optimization period. For under-saturated condition, the queue dissipates on or before green ends, and limiting the queue length is not much a concern. However, for saturated condition, some vehicles can not clear the intersection after the green is given to that approach. The accumulating vehicles and prolonging queue cycle by cycle might cause spill back problem (blocking the upstream intersections). It is essential to limit the queue length to avoid this adverse case.

Since the rates of arrival of all approaches remain the same during the optimization period, the longest queue (if there is any) exists at the moment when last time the approach is given green for the current optimization period. The constraint on queue length is thus formulated as follows:

$$xsg_{m_p}^0 + \sum_1^{ng-1} (v_{m_p}^i - s * NL_{m_p}^i * g_p^i) \leq xsg_{max}^i \quad \forall m_p^i \quad (3-28)$$

where

$xsg_{m_p}^0$ is the initial queue length at the moment the approach m_p^i is given green the very first time for the current optimization period.

n_{g_i} : number of times the approach m_p^i receives green in the current optimization period.

In equation (3-28), $v_{m_p^i}$ is a function of decision variables $DRVI$ and $DRVO$.

3.4.5 Measures of Effectiveness (MOEs)

For each optimal solution, a set of measures of effectiveness (MOEs) are calculated. The MOEs are used for evaluation of the traffic flow operation performance of the system components, and for comparisons with other simulation and optimization packages. The MOEs formulated include total delay for the freeway and the arterial, average stop delay per vehicle at intersections, the degree of saturation for each lane group, and the throughput of the corridor.

3.4.5.1 Total Delay

The total delay expressed in vehicle.hours (or vehicle.seconds) is one of the most common measure of effectiveness of the traffic condition. The total delay for arterial approaches and the freeway section is formulated below for the time period of one cycle.

Delay for the arterial intersection approach with under-saturated traffic condition

For an arterial intersection approach with under-saturated traffic condition, the queue is cleared during the green time, and there is no queue accumulated at the end of green time for the approach that has been receiving green (Figure 3-5, Case 1).

First, the time at which the queue is cleared, t_c , is calculated as:

$$t_c = t_1 + \frac{xsg + q(t_1 - t_0)}{Q - q} \quad (3-29)$$

The total delay for approach m_p^i is formulated as follows:

$$D_{m_p^i} = 0.5 * [xsg_{m_p^i} + (xsg_{m_p^i} + q*(t_1 - t_0))] * (t_1 - t_0) + 0.5 * (xsg_{m_p^i} + q*(t_1 - t_0)) * (t_c - t_1) \quad (3-30)$$

Substitute (3-20) $t_1 - t_0 = C - g_p^i$, $q = v_{m_p^i}$, and $Q = vd_{m_p^i}$ to (3-30), we have:

$$D_{m_p^i} = 0.5 * [2 * xsg_{m_p^i} + v_{m_p^i} * (C - g_p^i)] * (C - g_p^i) + 0.5 * (xsg_{m_p^i} + v_{m_p^i} * (C - g_p^i)) * \frac{xsg + v_{m_p^i} (C - g_p^i)}{vd_{m_p^i} - v_{m_p^i}} \quad (3-31)$$

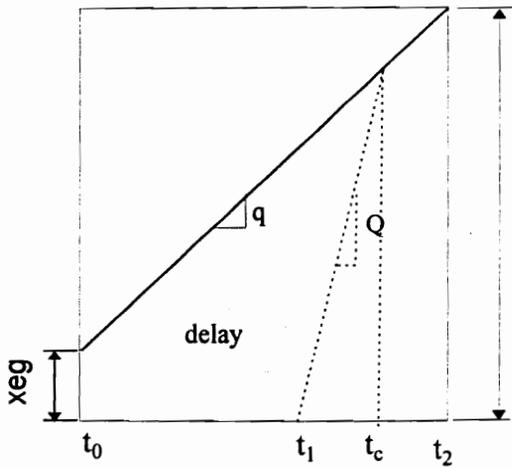
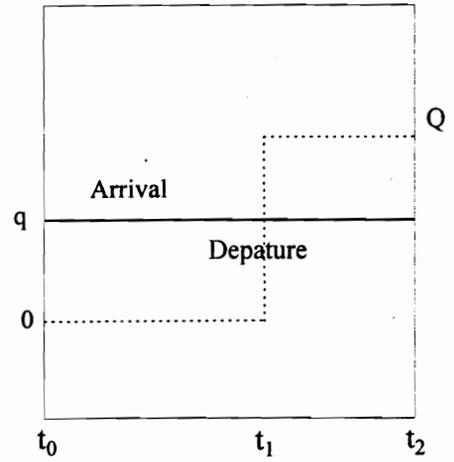
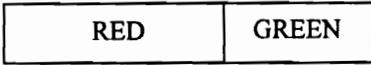
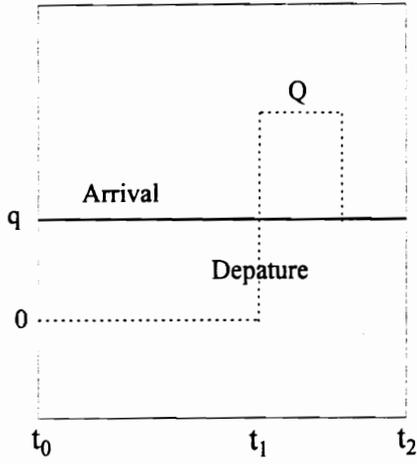
Delay for the arterial intersection approach with over-saturated traffic condition

For an arterial intersection approach with over-saturated traffic condition, the queue is not cleared during the green time, and there is queue accumulated at the end of green time for the approach that has been receiving green (Figure 3-5, Case 2).

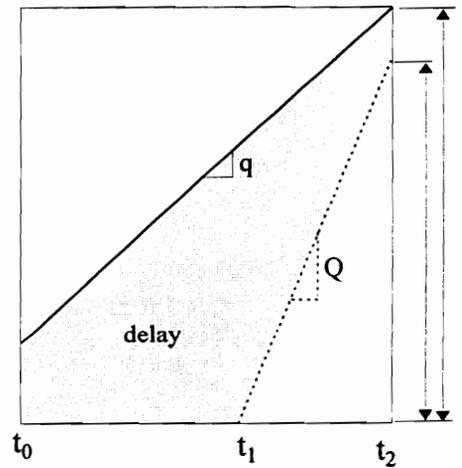
The total delay for approach m_p^i for this case is formulated as follows:

$$D_{m_p^i} = 0.5 * [xsg_{m_p^i} + (xsg_{m_p^i} + q*(t_1 - t_0))] * (t_1 - t_0) + 0.5 * [(xsg_{m_p^i} + q*(t_1 - t_0)) + (Q - q)*(t_2 - t_1)] * (t_2 - t_1) \quad (3-32)$$

Substitute $t_1 - t_0 = C - g_p^i$, $t_2 - t_1 = g_p^i$, $q = v_{m_p^i}$, and $Q = vd_{m_p^i}$ into (3-32), we have:



Case I Delay Calculation
for Under-saturated Traffic Condition



Case II Delay Calculation
Over-Saturated Traffic Condition

Figure 3-5. Delay Calculation for Arterial Intersection Approach

$$D_{m_p^i} = 0.5 * \left[2 * xsg_{m_p^i} + v_{m_p^i} * (C - g_p^i) \right] * (C - g_p^i) + 0.5 * \left[xsg_{m_p^i} + v_{m_p^i} * (C - g_p^i) + (vd_{m_p^i} - v_{m_p^i}) * g_p^i \right] * g_p^i \quad (3-33)$$

Delay on the Freeway

The delay on the freeway incident section (for the time period of one cycle length) is calculated by equation (3-34).

$$D_f = 0.5 * C * \left[xf_0 + (VF - \sum_{i=1}^{i=IOFF} \Delta VR^i * \delta^i - CFA) * C \right] \quad (3-34)$$

where xf_0 is the initial queue length at the time the cycle starts.

3.4.5.2 Average Stop Delay

The level of service of the intersection is directly related to the average stop delay value. The average stop delay per vehicle is estimated for each lane group and averaged for approaches and intersections as a whole.

The average stop delay per vehicle for an intersection is given by Equation (3-35) (HCM, 1994).

$$d_a^i = 0.38C \sum_{p=1}^{P^i} v_p^i \left\{ \frac{[1 - (g_p^i / C)]^2}{[1 - (g_p^i / C)X_p^i]} + 173(X_p^i)^2 [(X_p^i - 1) + \sqrt{(X_p^i - 1)^2 + 16X_p^i / c_p^i}] \right\} \quad (3-35)$$

Where

X_p^i degree of saturation = v_p^i / c_p^i ratio for lane-group

c_p^i capacity of the lane-group (veh/sec), $c_p^i = s \times g_p^i / C$

3.4.5.3 Throughput

The throughput of the corridor is defined as the total vehicles that exit the corridor on the direction of the freeway incident link during an optimization period. Vehicles exiting both from the freeway and the arterial contribute to the corridor throughput. The corridor throughput in the time period of a signal cycle (on the direction of the freeway incident link) is defined in equation (3-36).

$$\begin{aligned} TPC &= VT_f + VT_a \\ &= C * CFA + \frac{\sum_{i=IOFF}^{ION} \sum_{p=1}^{P^i} \sum_{\forall m \in DR} vd_{m_p^i}}{ION - IOFF} \end{aligned} \quad (3-36)$$

where TPC is the throughput of the corridor.

4. ALGORITHM AND SOLUTION PROCEDURE

4.1 INTRODUCTION

The optimization model for corridor optimal traffic control has been formulated and presented in the previous chapter. In this chapter, the characteristics of the model will be discussed. This is followed by the proposed algorithm and the solution procedure. A practical approach for optimizing all control variables will be addressed.

4.2 MODEL CHARACTERISTICS

To summarize, the complete corridor optimal traffic control model formulated to optimize phase splits and diversion rates is as follows:

$$\begin{aligned}
 \text{Minimize } & \sum_{\forall \tau} \sum_{\forall \text{cycles in period } \tau} \left\{ \sum_{i=1}^{NI} \sum_{j=1}^{P_i} \sum_{m=1}^{M_j^i} \left(\xi_{m_j^i} * \left(v_{m_j^i} - NL_{m_j^i} * s * (G_j^i - L) \right)^2 \right) \right\} \\
 & + \sum_{\forall \tau} \sum_{\forall \text{cycles in period } \tau} \left\{ \psi * \left(VF - \sum_{j=1}^{IOFF} (DVRI^j - VRIn^j) * \delta^j \right) * C - CFA * C \right\}^2
 \end{aligned} \tag{4-1a}$$

Subject to:

$$\sum_{p=1}^{P_i} G_p^i = C \quad \forall i \tag{4-1b}$$

$$G_p^i \geq G_{min} \quad \forall p, i \tag{4-1c}$$

$$G_p^i \leq G_{max} \quad \forall p, i \tag{4-1d}$$

$$DVRI^i \leq VRI_{max}^i \quad \forall \delta^i = 1, i = 1, 2, \dots, IOFF \tag{4-1e}$$

$$DVRO^i \leq VRO_{max}^i \quad \forall \delta^i = 1, i = ION, ION + 1, \dots, NI \tag{4-1f}$$

$$DVRI^i \geq VRIn^i \quad \forall \delta^i = 1, i = 1, 2, \dots, IOFF \tag{4-1g}$$

$$DVRO^i \geq VROn^i \quad \forall \delta^i = 1, i = ION, ION + 1, \dots, NI \quad (4-1h)$$

$$xsg_{m_p}^i \leq xsm_{m_p}^i \quad \forall i, p, m \quad (4-1i)$$

The decision variables of the optimization model (4-1) are the following:

- Green time for each phase at all intersections:

$$G_p^i \quad p = 1, 2, \dots, P_i, \quad i = 1, 2, \dots, NI$$

- Off-ramp diversion rates at upstream off-ramps:

$$DVRi^i \quad i = 1, 2, \dots, IOFF$$

- On-ramp diversion rates at downstream on-ramps:

$$DVRO^i \quad i = ION, ION + 1, \dots, NI$$

The characteristics of the model (4-1) are summarized as follows:

- (1). The objective function is nonlinear (quadratic), continuous, differentiable, and convex

For the arterial portion of the objective function, $NL_{m_j}^i * s * (G_j^i - L)$ is a linear function of G_j^i , while $v_{m_j}^i$ is either a constant if $m_j^i \in NDR$, or a linear function of decision variables $DVRi^i$ and $DVRO^i$ if $m_j^i \in DR$. For the freeway portion, the only variable is $DVRi^i$. The objective function is thus nonlinear, or quadratic to be exact.

The only variables in the objective function are G_j^i , $v_{m_j}^i$ and $DVRi^i$, there are linear functions of decision variables as formulated earlier. They are uniquely formulated over the entire optimization time horizon. From equations (3-4) through (3-10) it can be clearly seen that G_j^i , $v_{m_j}^i$ and $DVRi^i$ are continuous and differentiable over the entire optimization time. Therefore, the objective function (4-1a) is continuous and differentiable.

The objective function (4-1a) is also a convex function. It is in the form of the summation of terms, each of which is the square of a linear function of decision variables. It can be easily proven that a quadratic function in the form of the square of a linear function has a positive semidefinite Hessian, and is a convex function. With each component (term) being a convex function, the objective function (4-1a) is convex indeed.

(2). The constraints are linear

The constraints of the model (Equations 4-1b through 4-1i) are linear functions of control variables. In addition, they are also continuous and differentiable.

To summarize, the corridor optimal traffic control model is a nonlinear minimization problem with nonlinear, continuous, differential, and convex objective function, and linear, continuous, and differentiable constraints. More specifically, it is a quadratic programming problem.

4.3 SOLUTION ALGORITHM

The method of feasible directions can be applied to obtain optimal solutions effectively for optimization problems with convex, and differentiable objective functions. This class of methods solves a nonlinear programming problem by moving from a feasible point to an improved feasible point. The following strategy is typical of feasible directions algorithms:

- Given a feasible point x_k , a direction d_k is determined such that for $\lambda > 0$ and sufficiently small, the following properties are true:
 - (1). $x_k + \lambda d_k$ is feasible, and
 - (2). The objective value at $x_k + \lambda d_k$ is better than the objective value at x_k .

- After such a direction is determined, a one-dimensional optimization problem is solved to determine how far to proceed along d_k . This leads to a new point x_{k+1} , and the process is repeated.

Among the family of algorithms of this class, the gradient projection method was selected to solve the formulated optimization problem.

Gradient Projection Method

The direction of the steepest descent is that of the negative gradient. In the presence of constraints, however, moving along the steepest descent direction may lead to an infeasible point. The gradient projection of Rosen(1960) projects the negative gradient in such a way that improves the objective function and meanwhile maintains feasibility (Bazaraa et al. 1993).

Problems with linear Constraints

Consider the following problem:

$$\begin{aligned} & \text{Minimize } f(x) \\ & \text{subject to } \begin{aligned} Ax &\leq b \\ Ex &= e \end{aligned} \end{aligned} \tag{4-2}$$

where A is an $m \times n$ matrix, E is an $l \times n$ matrix, b is an m vector, e is an l vector, and $f: E_n \rightarrow E_1$ is a differentiable function. Given a feasible point x , the direction of steepest descent is $-\nabla f(x)$. However, moving along $-\nabla f(x)$ may destroy feasibility. To maintain feasibility, $-\nabla f(x)$ is projected so that we move along $d = -P\nabla f(x)$, where P is a suitable projection matrix. Figure 4-1 illustrates the projection of the gradient with $-P\nabla f(x)$ being the projected direction. P is the projection matrix.

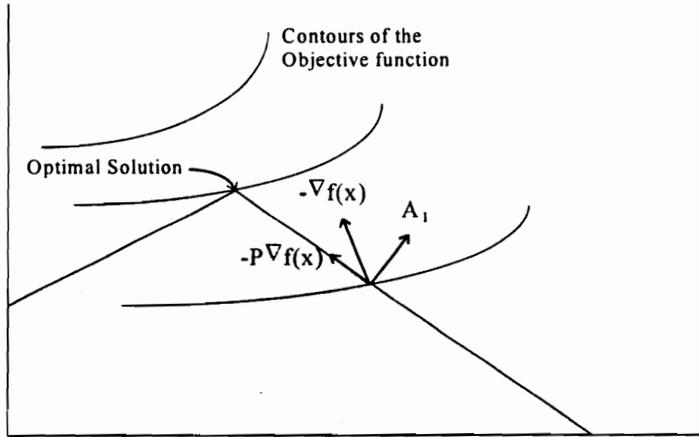


Figure 4-1. Illustration of Gradient Projection

The following lemma gives the form of a suitable projection matrix P and shows that $-P\nabla f(x)$ is indeed an improving feasible direction, provided that $-P\nabla f(x) \neq 0$.

Lemma

Consider the minimization problem (4-2). Let x be a feasible point such that $A_1x = b_1$ and $A_2x < b_2$, where $A^t = (A^t_1, A^t_2)$, and $b^t = (b^t_1, b^t_2)$. Furthermore, suppose that f is differentiable at x . If P is a projection matrix such that $-P\nabla f(x) \neq 0$, then $d = -P\nabla f(x)$ is an improving direction of f at x . Furthermore, if $M^t = (A^t_1, E^t)$ has full rank, and if P is of the form $P = I - M^t(MM^t)^{-1}M$, then d is an improving feasible direction.

Summary of the Gradient Projection Method of Rosen (Linear Constraints)

The procedure of Rosen's gradient projection method for solving a problem of the form to minimize $f(x)$ subject to $Ax \leq b$, and $Ex = e$ is summarized below:

Initialization step:

Choose a point x_1 with $Ax_1 \leq b$, and $Ex_1=e$. Suppose that A^t and b^t are decomposed into (A^t_1, A^t_2) and (b^t_1, b^t_2) such that $A_1x_1=b_1$ and $A_2x_1 < b_2$. Let $k=1$ and go to the main step.

Main step:

1. Let $M^t = (A^t_1, E^t)$. If M is vacuous, stop if $\nabla f(x_k)=0$, let $d_k=-\nabla f(x_k)$, and proceed to step 2. Otherwise, let $P=I-M^t(MM^t)^{-1}M\nabla f(x_k)$ and set $d_k=-P\nabla f(x_k)$. If $d_k \neq 0$, go to step 2. If $d_k=0$, compute $w=-(MM^t)^{-1}M\nabla f(x_k)$ and let $w^t=(u^t, v^t)$. If $u \geq 0$, stop; x_k is a KKT point, with w yielding the associated Lagrange multipliers. If $u \geq 0$ does not hold, choose a negative component of u , say, u_j , update A_1 by deleting the row corresponding to u_j and repeat step 1.
2. Let λ_k be an optimal solution to the following line search problem (4-7):

$$\begin{aligned} \text{Minimize} \quad & f(x_k + \lambda d_k) \\ \text{subject to} \quad & 0 \leq \lambda \leq \lambda_{max} \end{aligned}$$

$$\text{where } \lambda_{max} = \begin{cases} \text{minimum} \left\{ \frac{\hat{b}_i}{\hat{d}_i} : \hat{d}_i > 0 \right\} & \text{if } \hat{d} > 0 \\ \infty & \text{otherwise} \end{cases} \quad (4-7)$$

$$\begin{aligned} \hat{b} &= b_2 - A_2x_1 \\ \hat{d} &= A_2d_k \end{aligned}$$

4.4 MULTI-LAYER SCHEME FOR CORRIDOR OPTIMAL CONTROL

4.4.1 Solution Procedures of The Corridor Optimal Control Model

The basic corridor optimal traffic control model (4-1) has been developed to solve for optimal phase splits and diversion rates. It is formulated based on a given cycle length, and, the offsets of the signals are not taken into account.

It is perceived that the cycle length affects the traffic operation at intersections. Very short cycle lengths are not economical since the lost time takes higher percentage of the cycle time. On the other hand, more delay will be incurred if the cycle lengths are set to be too long. Also play a very important role on the traffic flow on signalized network are the offsets at intersections.

The offset (or yield point) is normally a time from a system reference point to the beginning point of the cycle at each of the signal controllers in the system. The ideal scenario is that the offsets are set in a way that through progression can be achieved and the traffic can flow through a number of intersections without stopping. This, however, is hard to achieve for signalized network with a high demand level. Nevertheless, the proper selection of offsets plays a very important role in improving traffic flow.

Incorporating offset factor into the model will destroy the differentiability of the objective function. Having cycle length as a variable in the model makes the queue constraints nonlinear since the number of cycles over the optimization period will be a variable accordingly. To make things worse, it may hamper the unique formulations of the constraints for queues for different forms of queue expressions have to be used. The same is true for the objective function. All these would make the model much more complicated and an efficient solution algorithm hardly obtainable.

To account for these two important factors while not complicating the model, a multi-layer scheme for optimal corridor control is developed.

- At layer 1, the basic optimization model described in previous sections is solved using the gradient projection method to optimize splits of signal timing at each arterial intersection and diversion rates for a given common cycle length with 0 offset for each arterial intersection.

- At layer 2 the offset of arterial intersection 2 through N is optimized (intersection 1 is used as the mast controller location, or, the reference point). TRANSYT-7F's platoon dispersion model is used to simulate the arterial traffic and the "best" offset is obtained based on comparing the total delay at each intersection.
- Finally at layer 3, the optimal cycle length selection is accomplished by evaluating system delay or throughput for cycle lengths in the range between minimum and maximum cycle length at a pre-specified increment.

Figure 4-2 illustrates the optimization procedures of the corridor control. The technical details of procedures for optimizing different control measures are discussed in succeeding sections.

4.4.2 Optimizing Splits and Diversion Rates

For a given cycle length, the phase splits (green times) of intersections and diversion rates are optimized over an optimization time period by finding the solution to the optimization problem (4-1) in the following steps. The optimization problem (4-1) is of the form of:

$$\begin{aligned} & \text{Minimize } f(x) \\ & \text{subject to } Ax \leq b \end{aligned} \tag{4-8}$$

Notice that there are no equality constraints for the model other than those for equal cycle length, and the objective function $f(x)$ is quadratic.

The following are the steps for building and obtaining optimal solutions for the model.

Step 1.

The first step is model parameters input. The following are the required inputs to the model:

- (1). Corridor configuration and geometry including:

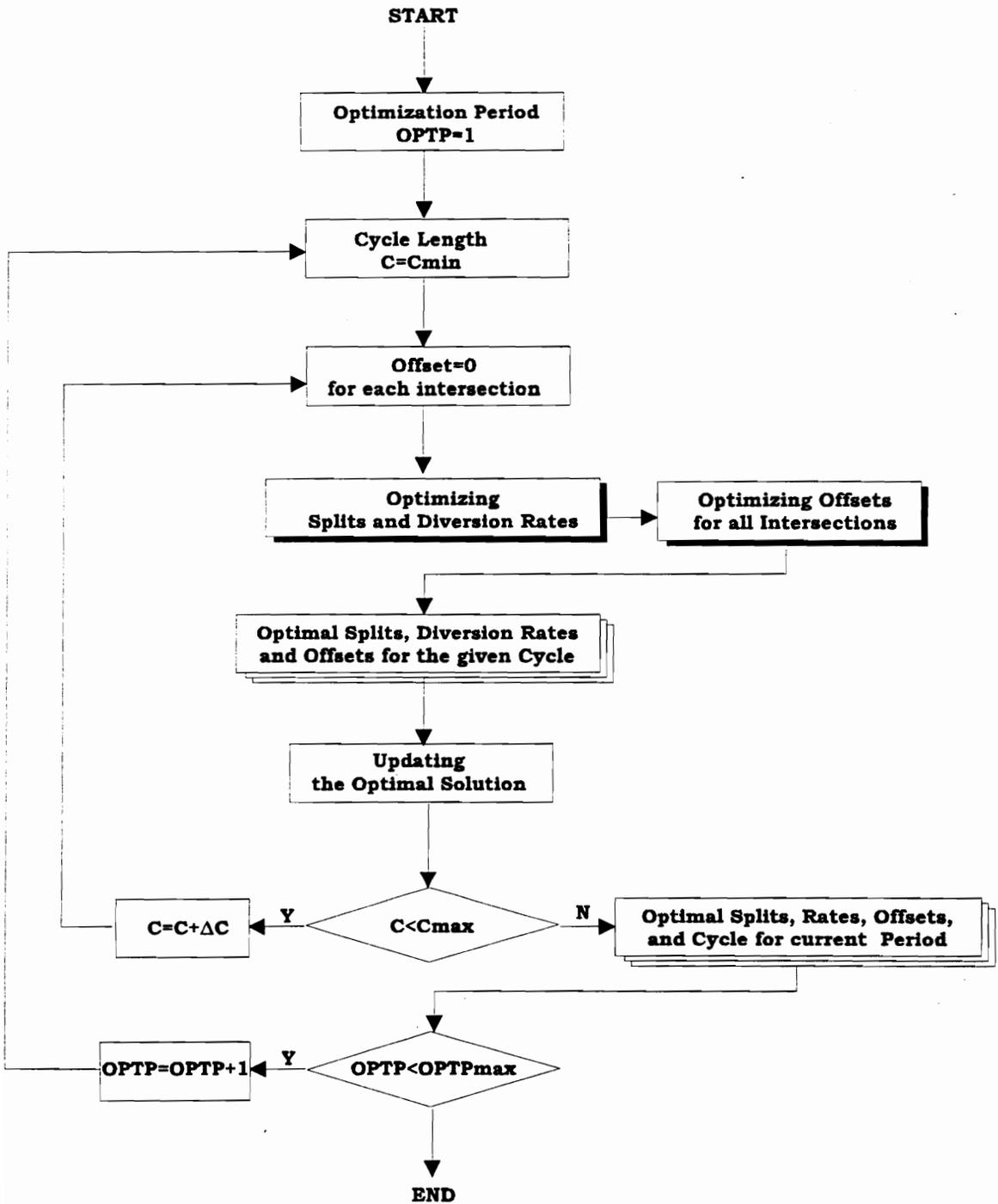


Figure 4-2. Procedures of the Corridor Optimization

- Number of arterial intersections
- Number of the entrance (on-) and exit (off-) ramps, and their relative location to arterial intersection Number of lanes for each approach at each intersection
- Distance between adjacent arterial intersections

(2). Arterial signal configuration including:

- Number of phases at each intersection
- Number of movements for each phase

This information, along with the ramp information, generates the control variables of the model.

(3). Signal timing and capacity related data including:

- Lost Time
- Saturation flow rate for each movement

(4). Prevailing traffic flow data for the current optimization period including:

- Arrival rate to each surface street approach under normal traffic condition (no incidents)
- Demand level on each on, or off-ramp under normal traffic condition
- Arrival demand to the freeway (on the direction with incident occurring)

(5). Incident information including:

- Incident location (on the freeway)
- Available freeway capacity under the incident condition

(6). Constraints information including:

- Minimum phase length
- Maximum phase length

- Maximum metering rate for each on-ramp
- Maximum metering rate for each off-ramp
- Maximum allowable queue length for each approach

The minimum on- or off-ramp metering rates are determined by the normal demand level without the incident occurrence.

Step 2:

Once the input data required by the model and algorithm is acquired, it starts to build the model coefficients for the objective function including the quadratic terms, linear terms, and the constant, the coefficients for each constraints, A , and the right hand side value for each constraint, b . The coefficients and right hand side values for constraints (4-1b) through (4-1h) are straightforwardly obtainable. Calculation based on equation (3-28) is needed to get the coefficients and right hand side values for constraints (4-1i).

Step 3:

An initial solution is also needed for the first optimization period. For succeeding optimization periods, the optimal solution of the previous period is used as the initial solution. The initial solution x_0 for the first period is generated by either one of the following two ways:

- The green times (in term of percentage of the cycle) of timing plans for normal traffic condition at all intersections are used as starting green times (percentages). The diversion rates at on, or off-ramps can be set initially to be the normal demand (lower bounds), or between the lower and upper bounds.
- A “quick optimization” technique is used to allocate green times for each phase based on the normal traffic demand of the critical approaches. The objective is to have the same degree of saturation for each approach. The diversion rates at on, or off-ramps are set, for this case, to be the normal demand.

Step 4

With all constraints being formulated, a feasibility check on the initial solution X_0 is performed. If any of the constraints (4-1b) through (4-1l) is violated, the initial solution X_0 is not feasible. A new set of initial value for X needs to be re-chosen. If the initial solution was chosen by the second method described in step 3, the new starting solution would need to be picked using the other approach. Once a feasible initial solution is at hand, set $k=1$, and a series of iteration of step 5, 6 that is described below will be performed to obtain the improving and, finally, the optimal solution.

Step 5

First, each constraint will be examined, and all binding constraints, if there is any, will be identified. The gradient of the objective function will then be evaluated at X_k . With those pieces of information, the projection matrix for current solution X_k will be formed, and the improving feasible direction d_k will be determined as described in section 4.3.

Then, after the improving feasible direction is found, a step length λ_k from the current solution x_k to the new improving solution x_{k+1} along the improving direction d_k is determined by solving the line search problem defined by equation (4-7). The optimization problem (4-7) is a one-dimensional line search problem with the objective function being convex and nonlinear (quadratic). The only variable to the problem is the step length λ_k , which has a search range from 0 to λ_{max} . λ_{max} is so determined as in (4-7) to not violate the previous unbinding constraints.

Various line search techniques can be employed to solve the optimization problem (4-7). The Golden Section Method summarized below is used for its simplicity and effectiveness for this one-dimensional line search problem with a convex and non-linear objective function.

Summary of the Golden Section Method

Initial Step:

Choose an allowable final length of uncertainty $l > 0$. Let $[a_1, b_1] = [0, \lambda_{max}]$ be the initial interval of uncertainty, and let $\lambda_1 = a_1 + (1 - \alpha)(b_1 - a_1)$ and $\mu_1 = a_1 + \alpha(b_1 - a_1)$, where $\alpha = 0.618$. Evaluate $f(\lambda_1)$ and $f(\mu_1)$, let $k=1$, go to the main step.

Main Step:

- (1). If $b_k - a_k < l$, stop; the optimal solution lies in the interval $[a_k, b_k]$. Otherwise, if $f(\lambda_k) > f(\mu_k)$, go to step 2, and if $f(\lambda_k) \leq f(\mu_k)$, go to step 3.
- (2). Let $a_{k+1} = \lambda_k$ and $b_{k+1} = b_k$. Furthermore, let $\lambda_{k+1} = \mu_k$, and let $\mu_{k+1} = a_{k+1} + \alpha(b_{k+1} - a_{k+1})$. Evaluate $f(\mu_{k+1})$, and go to step 4.
- (3). Let $a_{k+1} = a_k$ and $b_{k+1} = \mu_k$. Furthermore, let $\mu_{k+1} = \lambda_k$, and let $\lambda_{k+1} = a_{k+1} + (1 - \alpha)(b_{k+1} - a_{k+1})$. Evaluate $f(\lambda_{k+1})$, and go to step 4.
- (4). Replace k by $k+1$, and go to step 1.

Step 6

After the feasible improving direction is found and the step length is determined, a new solution x_{k+1} is obtained as follows:

$$X_{k+1} = X_k + \lambda_k d_k \quad (4-9)$$

The objective function at x_{k+1} will have a better value than that at x_k , i.e.,

$$f(x_{k+1}) \leq f(x_k)$$

If the termination criteria are not met, go to step 5, and the process repeats.

The following is the list of termination criteria. If either of the criteria is met, the process of seeking better solution terminates, and an optimality is declared.

- If there is no binding constraints, and all the components of the gradient at x_k equal to (or nearly equal to) zero, i.e.,

$$|\nabla f(x_k)| \leq \varepsilon_\nabla$$

- If all the components of the improving direction are found to be zero or near zero, i.e.,

$$|d_k| \leq \varepsilon_d$$

- If the step length along which the solution will move is found to be zero (or near zero), i.e.,

$$\lambda_k \leq \varepsilon_\lambda$$

where $\varepsilon_\nabla, \varepsilon_d, \varepsilon_\lambda$ are pre-defined allowable final lengths of uncertainty.

After the optimal solution is reached, the objective function is calculated at the optimality. The following measures of effectiveness (MOEs) are also calculated for the optimal solution:

- System delay including delay for arterial intersections and for the freeway section
- Throughput of the corridor
- Average delay per vehicle at each intersection. This is a representation of level of service for the intersection.
- Degree of saturation for each movement of the intersection. This is also an indication of level of service at the intersection.

Figure 4-3 illustrates the steps of split/diversion rate optimization procedure.

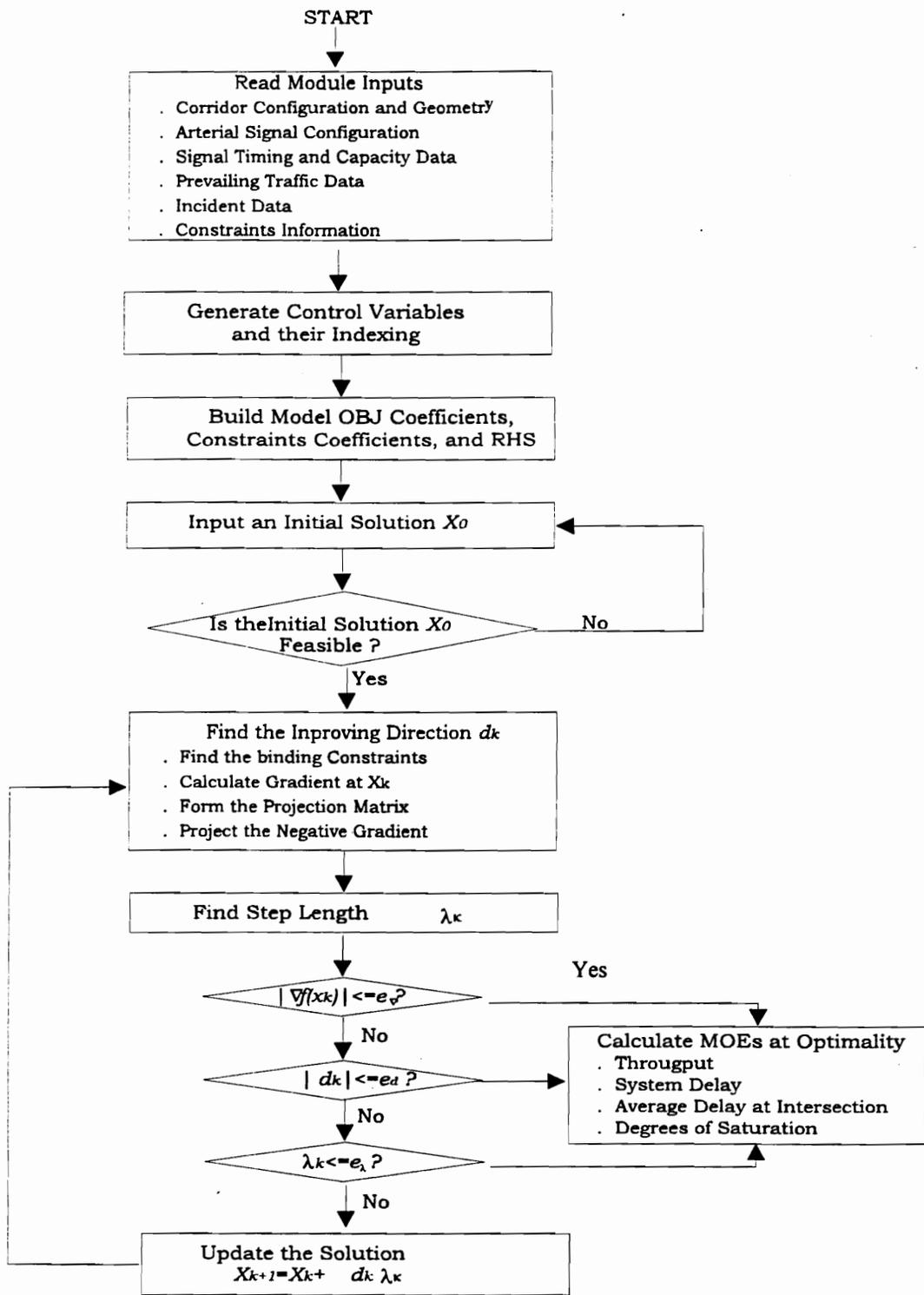


Figure 4-3. Procedures for Optimizing Phase Splits and Diversion Rates

4.4.3 Offset Optimization

The proper setting of offsets at intersections on the arterial reduces traffic stops and delay. Normally, offset optimization is considered only for the arterial main stream traffic that is the through movement. For the scenario with incident and diversion discussed in this study, however, the situation is different. Due to diversion from the freeway, the movements described below may need to be considered:

- The through movement which normally carries high volume of traffic flow
- The turning movement from the off-ramp to the main stream traffic. Due to the diversion, this movement may carry a considerable high traffic volume.

The objective of the offset optimization is to adjust the start time of the green (for the phase which gives the right of way to main stream movement) of the intersection relatively to that of the upstream intersection (or the master controller) in such a way, that the major portion of traffic from the upstream intersection will experience least amount of delay.

One of the most important aspects of traffic flow characteristics is platoon dispersion. As a queue becomes a moving platoon, it tends to disperse, or spread out, the farther downstream it travels. The dispersion reflects the normal tendency of driver to maintain safe headway between vehicles. To choose the offsets of downstream intersections, the platoon dispersion on links between intersections needs to be considered. It should be pointed out that platoon dispersion is not considered in the split/diversion rate optimization model. The reason to use a simple, deterministic traffic model in the split/diversion rate optimization model is to maintain model simplicity and solution algorithm efficiency, and at the macroscopic level of split/diversion rate determination, use of an aggregated average index such as the arrival rate is acceptable. However, to determine the proper offset, considering platoon dispersion factor bears more significance.

4.4.3.1 Platoon Dispersion Traffic Model Formulation

The TRANSYT-7F's platoon dispersion model is used. For each time interval (step), t , the arrival flow at the downstream stopline (ignoring the presence of a queue) is found by the following recurrence equation:

$$q'(t+T) = F*q_t + [(1-F)*q'(t+T-1)] \quad (4-10)$$

where

$q'(t+T)$: predicted flow rate (in time interval $t+T$ of the predicted platoon), where T is defined below;

q_t : flow rate of the initial platoon during step t

T : 0.8 times the cruise travel time on the link

F : a smoothing factor where, $F=1/(1+aT)$, and a is an empirically derived constant, called the platoon dispersion factor(PDF). $a=0.35$ is normally used.

Consider two adjacent intersections i and $i+1$, both intersections have 4 phase in a cycle of same length (with different splits). The cruise travel time on the link between these two intersections is T_{cruise} . According to the platoon dispersion model, the travel time for a platoon on the link from intersection i to intersection $i+1$ will be T , and

$$T=0.8 T_{cruise} \quad (4-11)$$

Suppose phase 1 is for traffic from the link from the freeway off-ramp, and phase 2 is for mainstream through movement on the arterial as shown in Figure 3-1. Depending on T , the offset at intersection $i+1$ with regard to intersection i (this offset can be easily converted to other reference intersection) and the phase lengths for both intersections, the platoon traffic discharged from phase 1 and 2 of intersection i will arrive at stopline of intersection $i+1$ at different stages of different phases. The offset at intersection $i+1$ should be set in such a way that the traffic released from phase 1 and 2, the majority of the traffic from upstream intersection i , will be discharged from intersection $i+1$ in mainstream direction with least amount of delay.

There are a number of scenarios that the platoon of diverted and mainstream traffic from upstream intersection arrives at different stages of the phases due to the offset and platoon travel time between the intersections. The initial departure flow rate (at upstream) changes due to phase change, and the discharge characteristics also changes because of phase changes at intersection $i+1$.

Figure 4-4 shows an example of a specific scenario where the first vehicles of platoon discharged from upstream phase 1, $P1(i)$, arrive at the stopline of intersection $i+1$ during phase 1, $P1(i+1)$. The last vehicles of platoon discharged from upstream phase 1, $P1(i)$, arrive at the stopline of intersection $i+1$ during phase 2, $P2(i+1)$. The first vehicles of platoon discharged from upstream phase 2, $P2(i)$, arrive at the stopline of intersection $i+1$ during phase 2, $P2(i+1)$. The last vehicles of platoon discharged from upstream phase 2, $P2(i)$, arrive at the stopline of intersection $i+1$ during phase 4, $P4(i+1)$. The platoon travel time on the link between the intersections is T . The offset at intersection $i+1$ is D (with reference to intersection i).

Suppose both intersections in Figure 4-4 are upstream of incident location, the traffic from phase 1 and 2 of intersection i will be discharged through phase 2 of intersection $i+1$ as desired. The traffic from Phase 1 and 2 of upstream will have several different arrival and departure patterns at the downstream intersection within certain time intervals due to the phase configurations at both intersections.

The following are the formulations of arrivals, departures, and delay at intersection $i+1$, for this particular scenario. A one-second interval is used for recurrent calculation.

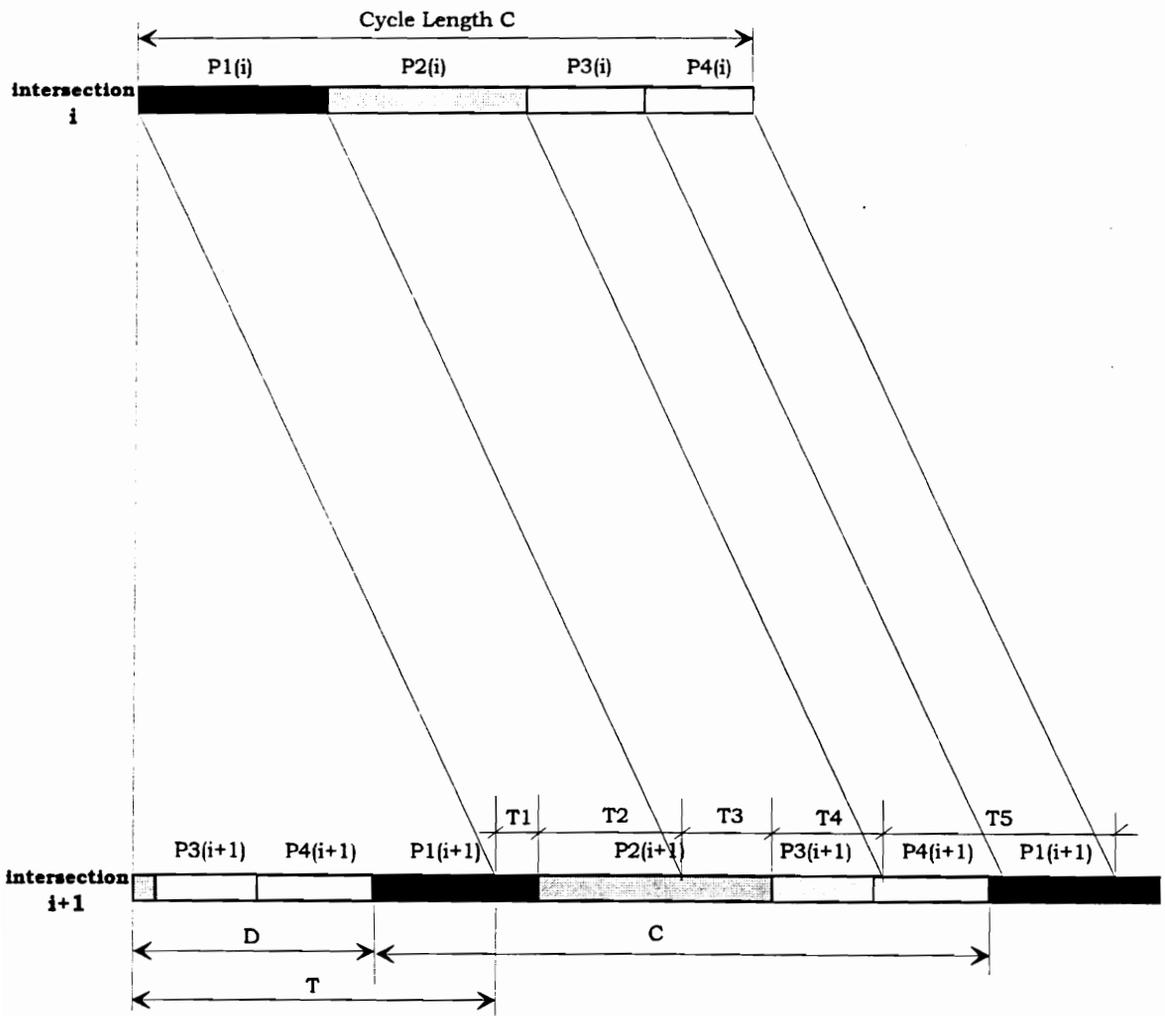


Figure 4-4. A Sample Scenario of Offset, Platoon Travel Time, and Phase Length for a Pair of Adjacent Intersections

Time period T1

In the first time period $[T, T+T1]$: the vehicles discharged from $P1(i)$ arrive at intersection $i+1$ during phase $P1(i+1)$. All the vehicles will stay in the queue until phase $P2(i+1)$ Starts. For this time period, we have:

$$q_t = s * NL_{-1,i} \quad (4-12)$$

$$q'(t+T-1)|_{t=0} = q0 \quad (4-13)$$

where

$q0$ is the arrival rate of previous interval to downstream intersection, i.e., at $t=T-1$, $NL_{-1,i}$ is the number of lanes for the movement that carries left-turn diversion volume from the freeway. The traffic of this movement is given green during phase 1 at intersection i .

By equation (4-10), the arrival rate to intersection $i+1$ can be calculated recurrently.

$$\begin{aligned} q'(0+T) &= F * q_t + [(1-F) * q0] \\ q'(1+T) &= F * q_t + [(1-F) * q'(0+T)] \\ q'(2+T) &= F * q_t + [(1-F) * q'(1+T)] \\ &\dots\dots\dots \\ q'(T1+T) &= F * q_t + [(1-F) * q'(T1-1+T)] \end{aligned} \quad (4-14)$$

In this period, the total arrival volume to intersection $i+1$ that will only be possibly released when $P2(i+1)$ is given green is calculated by

$$v(T1) = \sum_{t=1}^{T1} q'(t+T) \quad (4-15)$$

There is no departure volume for this period since phase $P2(i+1)$ is not given green.

Assume the initial queue length is: $x(0+T)=x0$, then the queue length at time $t+T$ is formulated as follows:

$$\begin{aligned}
x(1+T) &= x(0) + q'(0+T) \\
x(2+T) &= x(1+T) + q'(1+T) \\
x(3+T) &= x(2+T) + q'(2+T) \\
&\dots\dots \\
x(T1+T) &= x(T1-1+T) + q'(T1-1+T)
\end{aligned} \tag{4-16}$$

The delay over this interval is calculated in equation (4-17).

$$Da_{2^{i+1}}(t1) = \frac{1}{2}(x(0+T) + x(T1+T)) + \sum_{t=1}^{T1-1} (x(t+T)) \tag{4-17}$$

Time period T2

In the time period $[T+T1, T+T1+T2]$: the vehicles discharged from $P1(i)$ arrive at intersection $i+1$ during phase $P2(i+1)$. The vehicles arriving and the vehicles in the queue start departing. For this time period, the upstream departure remains the same as period T1 (4-12), and the same recurrent equation as (4-14) is employed to calculate arrivals to intersection $i+1$ for each interval.

In this period, the total arrival volume to intersection $i+1$ is calculated by

$$v(T2) = \sum_{t=T1+1}^{T1+T2} q'(t+T) \tag{4-18}$$

The departure for each interval (1 second) in this period is calculated as follows:

$$vd|_{t+T} = \min \left\{ \begin{array}{l} s * NL_{2_2^{i+1}} \\ x(T+t) + q'(t+T) \end{array} \right. \quad t = T1, T1+1, \dots, T1+T2-1 \tag{4-19}$$

where

$vd|_{t+T}$ is departure volume for interval $[t+T, t+T+1]$,
 $x(T+t)$ is the queue length at $t+T$ calculated as (4-20).

$$x(t+T) = \max \left\{ \begin{array}{l} x(t+T-1) + q'(t+T) - s * NL_{2_2^{i+1}} \\ 0 \end{array} \right. \tag{4-20}$$

The delay over this interval is calculated in equation (4-21).

$$Da_{2^{i+1}}(t_2) = \frac{1}{2}(x(T_1 + T) + x(T_1 + T_2 + T)) + \sum_{t=T_1+1}^{T_1+T_2-1} (x(t + T)) \quad (4-21)$$

Time period T3

In the time period $[T+T_1+T_2, T+T_1+T_2+T_3]$: the vehicles discharged from $P2(i)$ arrive at intersection $i+1$ during phase $P2(i+1)$. The vehicles arriving and the vehicles in the queue (if there is still any) continue being released.

For this time interval, the upstream departure rate is determined by equation (4-22) as follows:

$$q_t = s * NL_{2_2^i} \quad (4-22)$$

Again, the same recurrent equation as (4-14) is employed to calculate arrivals to intersection $i+1$ for each interval. And the arrival, departure, queue, and delay formulations at intersection $i+1$ are as follows:

$$v(T_3) = \sum_{t=T_1+T_2+1}^{T_1+T_2+T_3} q'(t + T) \quad (4-23)$$

$$vd|_{t+T} = \min \left\{ \begin{array}{l} s * NL_{2_2^{i+1}} \\ x(T + t) + q'(t + T) \end{array} \right. \quad (4-24)$$

$$x(t + T) = \max \left\{ \begin{array}{l} x(t + T - 1) + q'(t + T) - s * NL_{2_2^{i+1}} \\ 0 \end{array} \right. \quad (4-25)$$

$$D(t_3) = \frac{1}{2}(x(T_1 + T_2 + T) + x(T_1 + T_2 + T_3 + T)) + \sum_{t=T_1+T_2+1}^{T_1+T_2+T_3-1} (x(t + T)) \quad (4-26)$$

$t = T_1 + T_2, T_1 + T_2 + 1, \dots, T_1 + T_2 + T_3 - 1$ for equations (4-23) through (4-26).

Time period T4

In the time period $[T+T1+T2+T3, T+T1+T2+T3+T4]$: the vehicles discharged from $P2(i)$ arrive at intersection $i+1$ After phase $P2(i+1)$. The vehicles arriving will be queued again. The departure rate of upstream is determined by (4-22). In this period, no vehicles from $P1(i)$, $P2(i)$ are released. Equations similar to (4-14) through (4-17) can be used to calculate arrival rate at each interval, arrival volume of the period, queue length at each interval, and delay of the period.

Time period T5

In this period $[T+T1+T2+T3+T4, T+C]$, there are no vehicles released from $P1(i)$, $P2(i)$, neither does any vehicle discharging from $P2(i+1)$. However, all the vehicles left in the queue at the end of time period T4 will experience additional delay during this period. The delay is calculated as follows:

$$Da_{2,i}(T5) = x(T1 + T2 + T3 + T4 + T) * (C - T1 - t2 - t3 - t4) \quad (4-27)$$

4.4.3.2 Offset and Phase Length Configuration Scenarios

It is shown in 4.4.3.2 that the configuration of the offset and phase lengths for the adjacent intersections determines the arrival and departure pattern, and has significant effect on the delay. If the phase splits are fixed for the adjacent intersections, the delay can be reduced significantly by properly setting the offset. For the scenario shown in Figure 4-4, for example, the delay can be reduced if the offset T is increased slightly.

A number of offset and phase length configuration scenarios are considered in determining optimal offsets. Equations for arrival, departure, queue and delay are different due to different offset and phase length combination scenarios. Equations (4-12) through (4-27) are applicable only to the sample scenario shown in Figure 4-4.

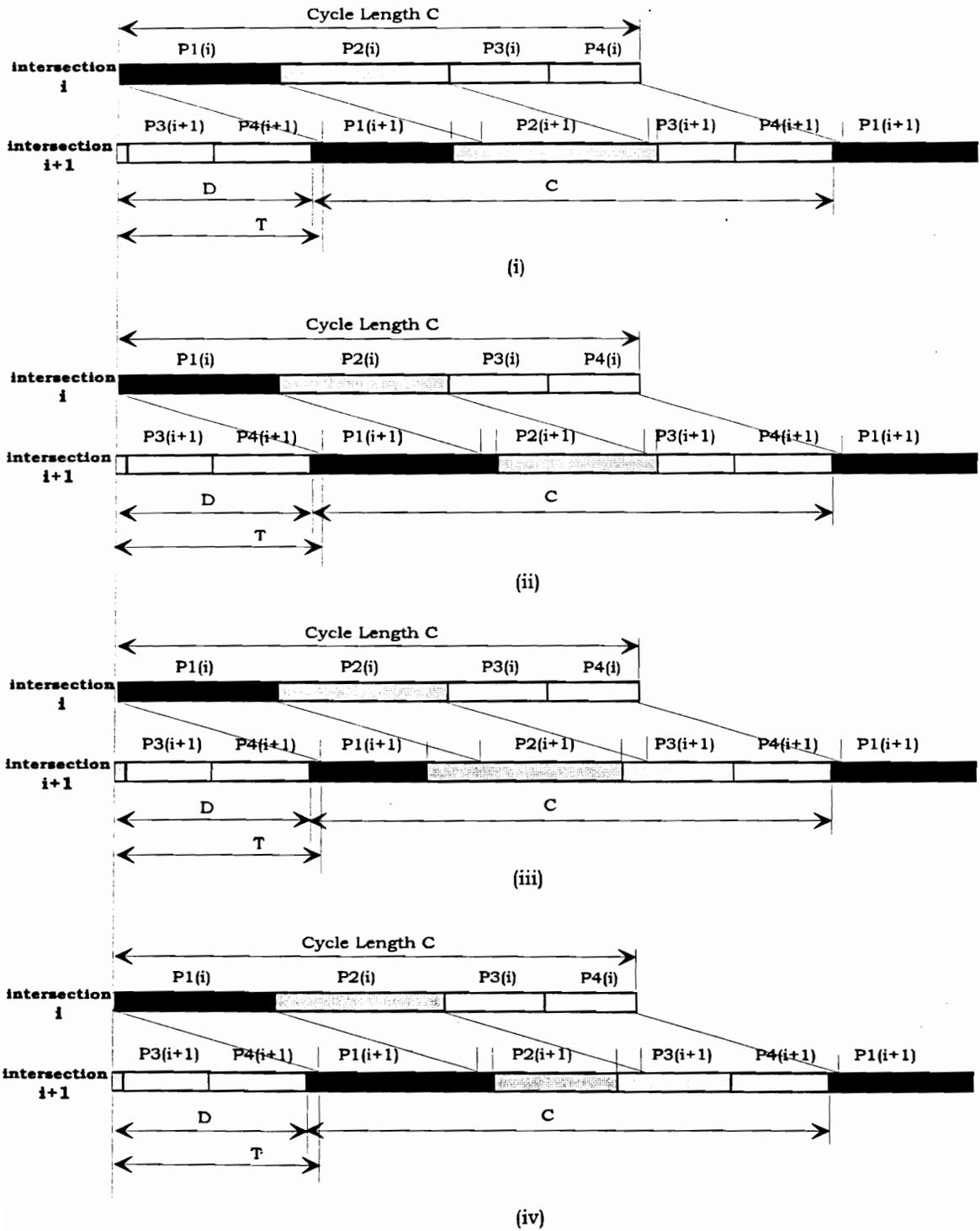


Figure 4-5. Arrival/Departure Patterns for Different Phase Lengths

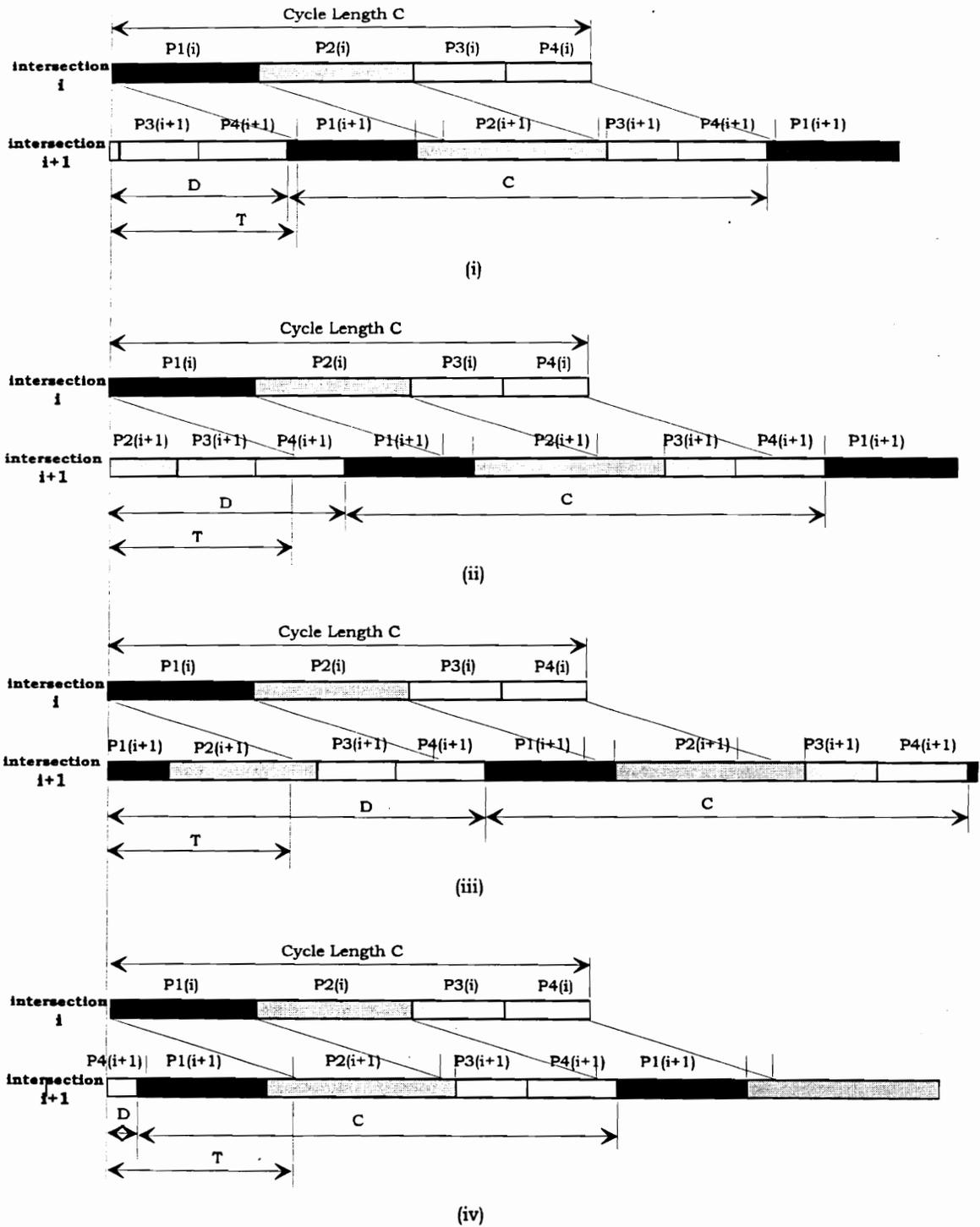


Figure 4-6. Arrival/Departure Patterns for Different Offsets

Figure 4-5 illustrates four typical scenarios of arrival/départure patterns due to the differences in phase lengths at the two adjacent intersections. The offset is the same for all scenarios. Figure 4-6 shows four typical scenarios of arrival/départure patterns due to the differences in offsets at the two adjacent intersections. Phase lengths are the same for all cases.

Sets of equations were formulated for all the possible scenarios.

4.4.3.3 Determining Optimal Offsets

As stated in 4.4.1, offset optimization is carried out after the phase splits for all intersections are optimized. The only variables to be determined are the offsets. Using delay as the objective function, an optimization model can be formulated follows:

$$\begin{aligned} \text{Minimize} \quad & \sum_{\forall \tau} \sum_{\forall \text{cycles in period } \tau} \left\{ \sum_{i=1}^{Nl} \sum_{j=1}^{P_i} \sum_{\forall T} Da_j(T) \right\} \\ \text{Subject to:} \quad & 0 \leq D_i \leq C \end{aligned} \tag{4-28}$$

Where

$Da_j(T)$ is the delay experienced by the vehicles that pass the intersection i when phase j is given green during one of the time interval T during which a persistent arrival/départure pattern prevails in a cycle.

Due to the fact that the phase splits at all intersections are already determined, $Da_j(T)$ is only a function of variable D_i , the offset at intersection i , which is to be optimized in the range between 0 and the cycle length C .

As discussed in 4.4.3.1 and 4.4.3.2, $Da_{j_i}(T)$ takes different equations for different time period during a cycle for different arrival/departure patterns caused by different values of offsets D_i . $Da_{j_i}(T)$ is thus non-differentiable, and (4-28) is a discrete optimization problem. Because of the fact that the delay formulation at intersection i is determined only by the phase splits of intersection i , and $i-1$, and the relative offset between these two intersections, it is possible to separate the optimization problem (4-28) and optimize the offset one intersection at a time.

To determine the optimal offset of the intersection i is now to solve the discrete optimization problem with only one variable D_i , and D_i is bounded between 0 and the cycle length C . Simply by dividing C into some small steps and setting D_i at the values of these points, and evaluating and comparing the resulting delay for the intersection, an optimal offset can be easily determined. The offset value is relative to the upstream intersection, and needs to be converted to the offset value with a common intersection along the arterial that is the reference point.

Suppose the offset of intersection $i-1$ is determined to be D_{i-1} , and the optimal offset of intersection i with regard to intersection $i-1$ is $D_{i-1,i}$, then the optimal offset of intersection i , D_i (to the common reference intersection along the arterial) is determined as follows:

$$D_i = \begin{cases} D_{i-1} + D_{i-1,i} & \text{if } D_{i-1} + D_{i-1,i} < C \\ D_{i-1} + D_{i-1,i} - C & \text{otherwise} \end{cases} \quad (4-29)$$

Figure 4-7 illustrates the steps of determining optimal offsets.

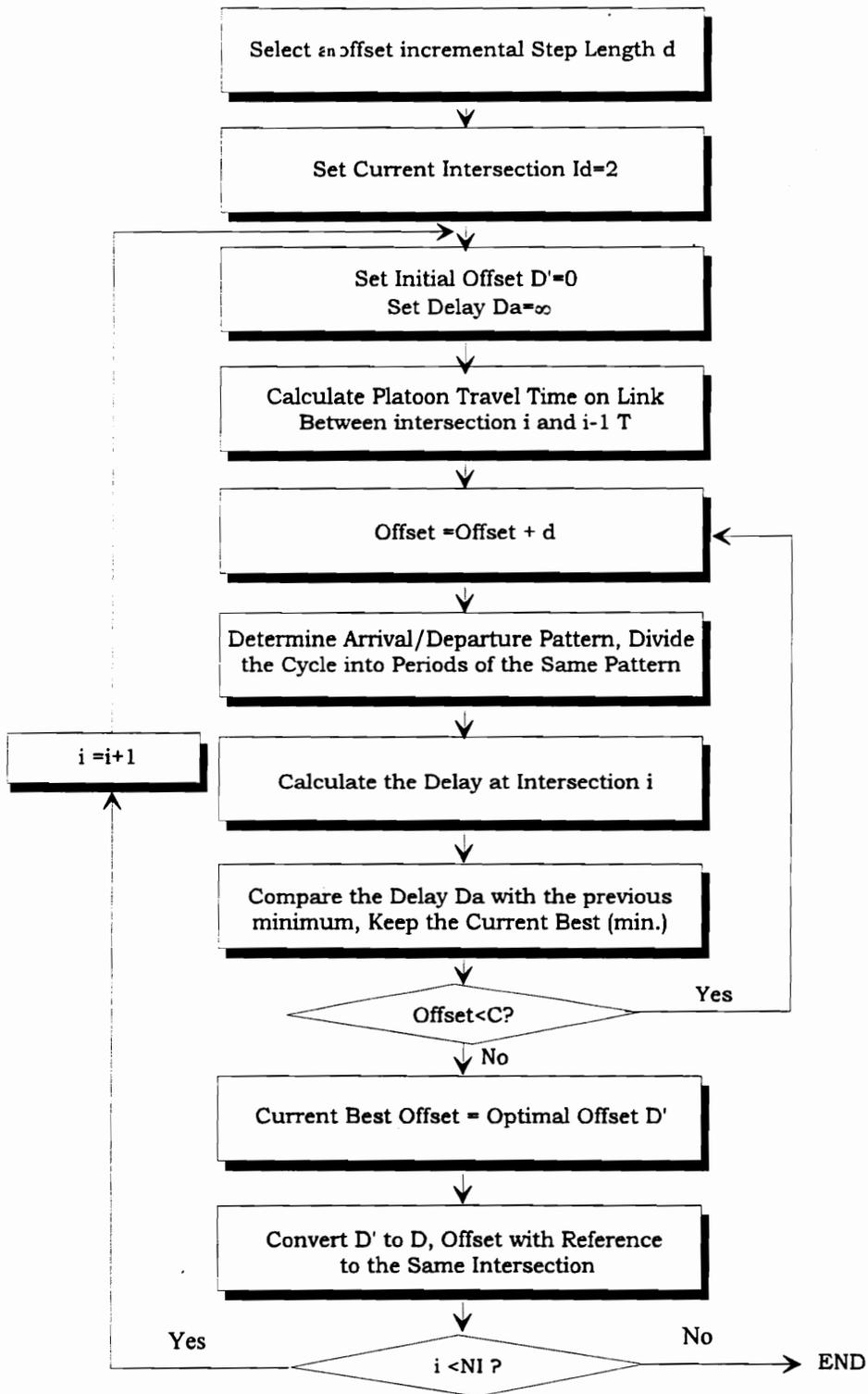


Figure 4-7. The Procedure for Optimizing Signal Offsets

4.4.4 Optimal Cycle Selection

The cycle length plays an important role in affecting the traffic operation of signalized intersections. A small cycle length is considered to be uneconomical since the percentage of effective green time over the cycle length turns to be smaller for shorter cycle length due to the fact that the lost time for each phase is pretty much a constant and independent of the cycle length. With the traffic demand level increasing, longer cycle length is conceived to be more beneficial. However, an extremely long cycle length will result in a great amount of delay since the traffic of the other approaches that are not receiving green has to wait in the queue for a longer time before it is given the green to proceed. A proper cycle length needs to be determined to achieve optimal traffic operation performance at signalized intersections.

As stated in 4.4.1, the optimization of splits and diversion rates is made for a given cycle length for all intersections. The “best” cycle length is determined by evaluating the performance of the optimal timing and diversion plan with a cycle length ranging from minimum cycle length to maximum cycle length at pre-defined step length.

The cycle length is not included as a variable in the basic optimization model because of the fact that with cycle length being added into the model as a variable, the objective function and constraints will no longer have unique formulations, thus lose the differentiability. Also, queue constraints will become nonlinear and non-unique.

It is always nice to make the optimization model as comprehensive and accurate as possible, and to pursue the theoretical global optimum. Nevertheless, with more and more variables being added into the model, the model becomes more and more complicated. Some of the nice features, such as convexity and differentiability, will

no longer exist in the model. An effective and efficient solution algorithm will be hard to obtain. By solving the optimization problem using a layered or sequential optimization approach, fast and effective solution techniques are much easier to obtain by exploiting the special features of the models such as convexity and differentiability. One can always argue that by doing so the original problem has been changed, and the solution to the problem may not be at its global optimum. However, if this is the only way to get a reasonable solution within an acceptable time period, it is a tradeoff we have to make to deal with complicated real world problems.

5. COROPT: A CORRIDOR OPTIMIZATION COMPUTER PROGRAM

5.1 FUNCTIONS OF COROPT PROGRAM

A computer program COROPT (CORridor OPTimization) programmed in C was developed to perform the optimization procedure described in Chapter 4. The program was developed on a UNIX platform.

The COROPT program has several modules and can perform the following options:

1. Data Input

COROPT has a data input module that assists user to enter all the required data items in a correct manner. It guarantees that all the required data is entered, and entered in a correct order and format. All the data entered will be saved to an input data file for future use. When the data input option is chosen, the program will check whether an input file already exists, and ask the user whether overwriting that data file with new data is a desired action.

2. Plan Evaluation

COROPT has the function to do plan evaluation. With all the diversion rates and signal timing plans given, the program can calculate MOEs including delay, and the throughput.

3. Optimization Without Optimizing offsets

This performs the optimization process of determining diversion rates, signal timing plans at signal intersections for each cycle length evaluated, and optimal cycle length. Offsets at all intersections are assumed to be zero and platoon dispersion factor is not considered.

4. Complete Optimization

This performs the optimization process of determining diversion rates, signal timing at intersections, the cycle length, and offsets.

5.2 FEATURES OF COROPT PROGRAM

Corridor optimization is a complicated problem. Model comprehension, complexity, structure of optimization procedures, effectiveness and efficiency of the algorithm, and traffic flow model all play roles in determining whether the approach is meaningful, practical and implementable. A deliberately planned optimization scheme, a carefully selected algorithm and solution procedure, along with the realistic and practical traffic flow modeling, provide the COROPT program with the following nice features.

Comprehensive

The corridor optimization model is comprehensive. The model and approach take into account of the traffic flow of the freeway and the arterial, and solve the diversion problem and signal retiming problem in an integrated manner.

The approach allows for diverting traffic using multiple upstream off-ramps and downstream on-ramps. Both diversion rates and routes are determined by the same model. Also, all key components of the signal timing problem, phase splits, cycle lengths, and offsets, are optimized.

The model was developed for severe incident conditions. It is, however, also suitable for non-incident conditions where either the arterial or the freeway, or both, has a high demand level, and re-assignment of traffic and retiming of arterial intersection signals will improve overall traffic flow conditions of the corridor system.

Flexible

The model and approach have sufficient flexibility to deal with real world scenarios. The following are important issues considered in the model and approach to provide sufficient flexibility for the sake of practical applications.

1). Different time period division

The optimization time horizon is divided into several time periods of prevailing traffic and incident conditions. As time goes by, traffic conditions of demand /capacity level in the corridor system change. The main factors causing major traffic pattern changes are:

- Time-of-day

With the advent or departure of the peak hour, the traffic demand increases or decreases significantly. The prevailing traffic condition changes with time-of-day.

- Clearance process of the incident

Lane blockage often comes with severe incidents and reduces capacity for that link or section of the roadway. The lane blockage situation, however, changes over the time of the incident clearance process. As time goes by, lane(s) that was originally blocked is cleared and reopens to traffic. The available capacity of the roadway is a factor governed by the dynamic lane clearance process

Optimal traffic control strategies are developed for each time period during which the prevailing traffic and incident conditions, thus demand and capacity level, remain unchanged. The COROPT program provides with the flexibility of dividing the optimization time at user's will into multiple time periods according to actual or predicted traffic and incident conditions.

2). Different corridor configurations

The COROPT program also has the ample flexibility to deal with different scenarios of corridor system configurations in the following aspects:

- Different relative location of the parallel arterial

The problem can deal with the cases of (1) having a parallel arterial on the left side of the freeway direction with incident occurring, (2) having a parallel arterial on the right side of the freeway direction with incident occurring, and (3) having two parallel arterials on both sides of the freeway. A symbol lr is adopted in the

COROPT program to indicate the relative position of parallel arterial to the freeway direction of incident occurrence. lr is an input of the program that affecting phasing for diversion traffic.

- $lr=1$: the arterial is to the left of the freeway traffic movement of interest,
- $lr=2$: the arterial is to the right of the freeway traffic movement of interest,
- $lr=3$: there is one arterial on each side of the freeway.

- Different arterial intersection/ramp connection scenarios

Along the arterial parallel to the freeway, there is often a signalized intersection if there is a direct link connecting ramps to the freeway. There are, however, also intersections that are not directly connected to the freeway. The COROPT program handles this problem by employing a symbol δ :

- $\delta = 0$: if there is a ramp directly connects intersection i to freeway, and
- $\delta = 1$: if there is not a ramp directly connects intersection i to freeway.

Figure 5-1 shows the different intersection/ramp connection scenarios. Intersection 3, and 7, as shown in the figure, are not connected to freeway via interchange ramps.

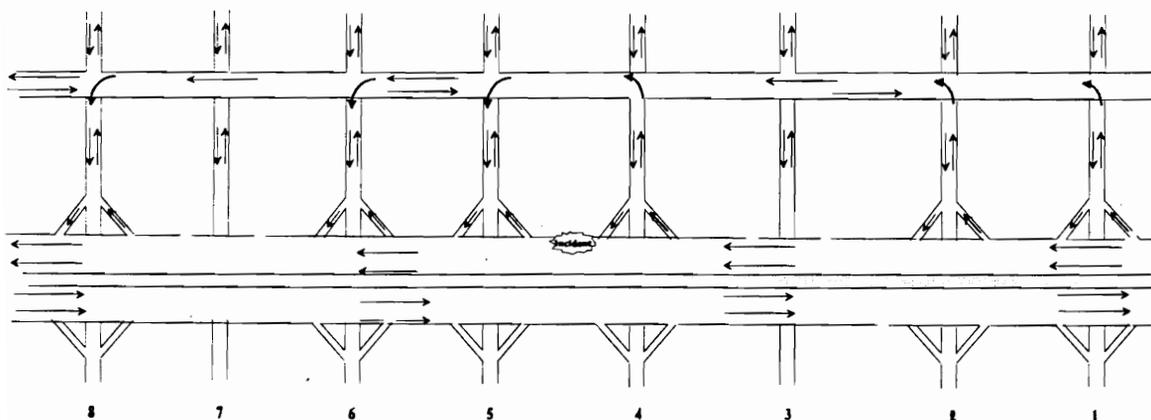


Figure 5-1. Illustration of Intersection/Ramp Connection Scenarios

- Different arterial intersection types and geometric layouts

The basic types of at-grade arterial intersections are the T intersections (with multiple variations of angular approach), the four-leg intersections, and the multi-leg intersections (more than four legs)(AASHTO, 1991). In each particular case the type is determined primarily by the number of intersecting legs, the topography, the traffic pattern, and the desired type of operation.

The COROPT program can handle several intersection types including T intersection, four-leg intersection, five-leg intersection, and six-leg intersection.

- Different lane configurations for arterial streets and ramps

The COROPT program allows for inputting number of lanes for each link and approaches of arterial streets and entrance or exit ramps. However, lane number changes are only allowed at intersections. Adding or dropping lanes in the middle of a link between two intersections can not be accounted for by the program.

- Different corridor length

The length of the corridor in consideration can be represented by number of upstream and downstream intersections. The program can develop optimal control strategies for a corridor of up to twenty intersections on the arterial. It is up to user's decision to include as many upstream and downstream intersections. By including more intersections in the optimization process, it is sometimes more likely to find a better control strategy.

3). Different phasings

The COROPT program can deal with different phasing schemes in the following ways:

- Up to seven phases are allowed in a signal cycle.

- Up to four movements are allowed in each phase at a signalized intersection.
- The program can deal with different phase/lane groups combinations. For example, the 2 commonly used phasing scenarios, shown in Figure 5-2, can be handled by the program. It should be noted that due to the high diversion volume in some movements, phasing *B* probably is a better phasing to use.

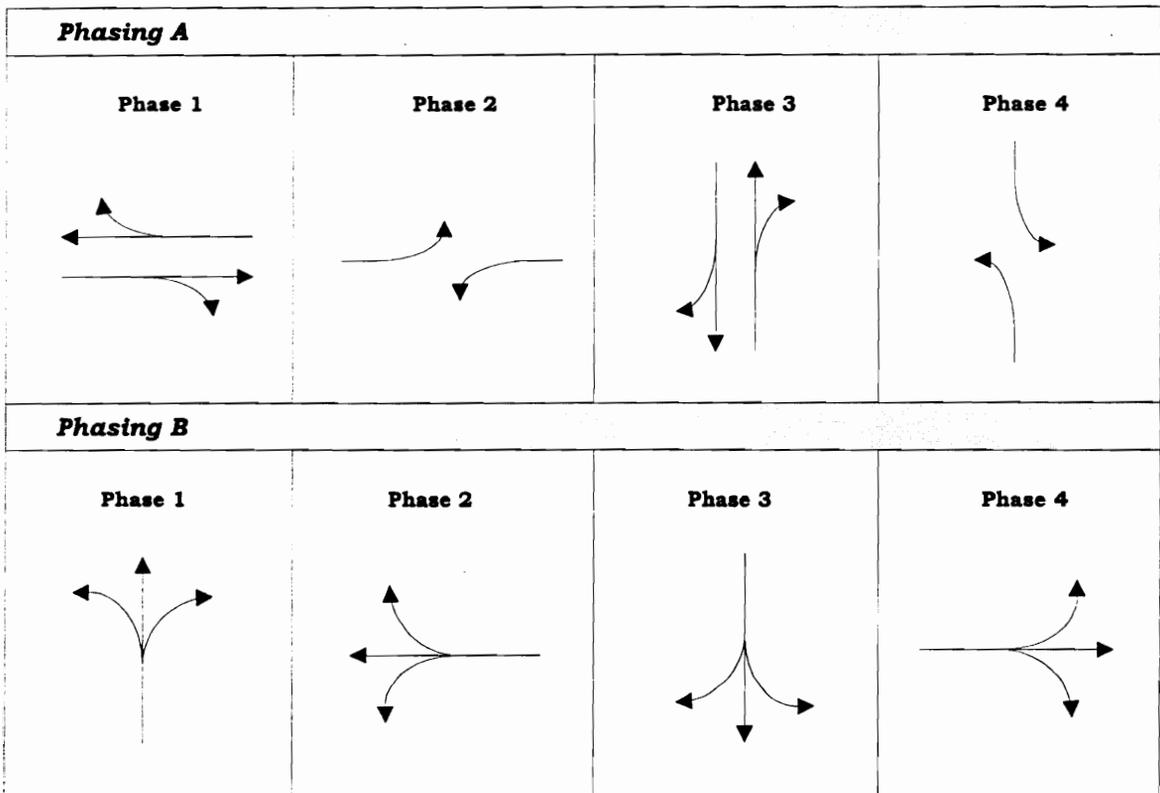


Figure 5-2. Different Cases of Phasings at an Intersection

5.3 DETERMINING PROGRAM PARAMETERS

The COROPT program can be utilized to develop optimal traffic control strategy for a corridor under incident conditions. To use the program, a number of issues and model parameters need to be resolved first.

5.3.1 Determining Optimization Time -- the Incident Duration

The model and COROPT program are developed primarily for corridor traffic control during incidents. Once an incident occurs, the duration of the incident needs to be determined. The special traffic control strategy, part of the incident management effort, is developed for this time period, which should be used as the optimization horizon of the COROPT program run.

The process of incident management consists of four sequential steps: incident detection, response, clearance, and recovery (Figure 5-3). The incident duration is the total time spent on all the steps of the incident management process.

Incident duration is mainly determined by incident type, severity and characteristics, lane blockage, response procedures for the area, response resources availability, traffic demand, and traffic management strategies. Other factors such as incident location, land use, weather, temperature, and time-of-day may also affect incident duration to some extent.

One of the major functional components of the hybrid expert-GIS Wide-Area Incident Management Support System (WAIMSS) (Ozbay et al., 1996) which has been developed at Virginia Tech for incident response and traffic management during roadway incidents is the incident duration estimation module. This module is developed by analyzing historical incident data collected by the Northern Virginia area incident management personnel. Information of over 6000 incidents in NOVA

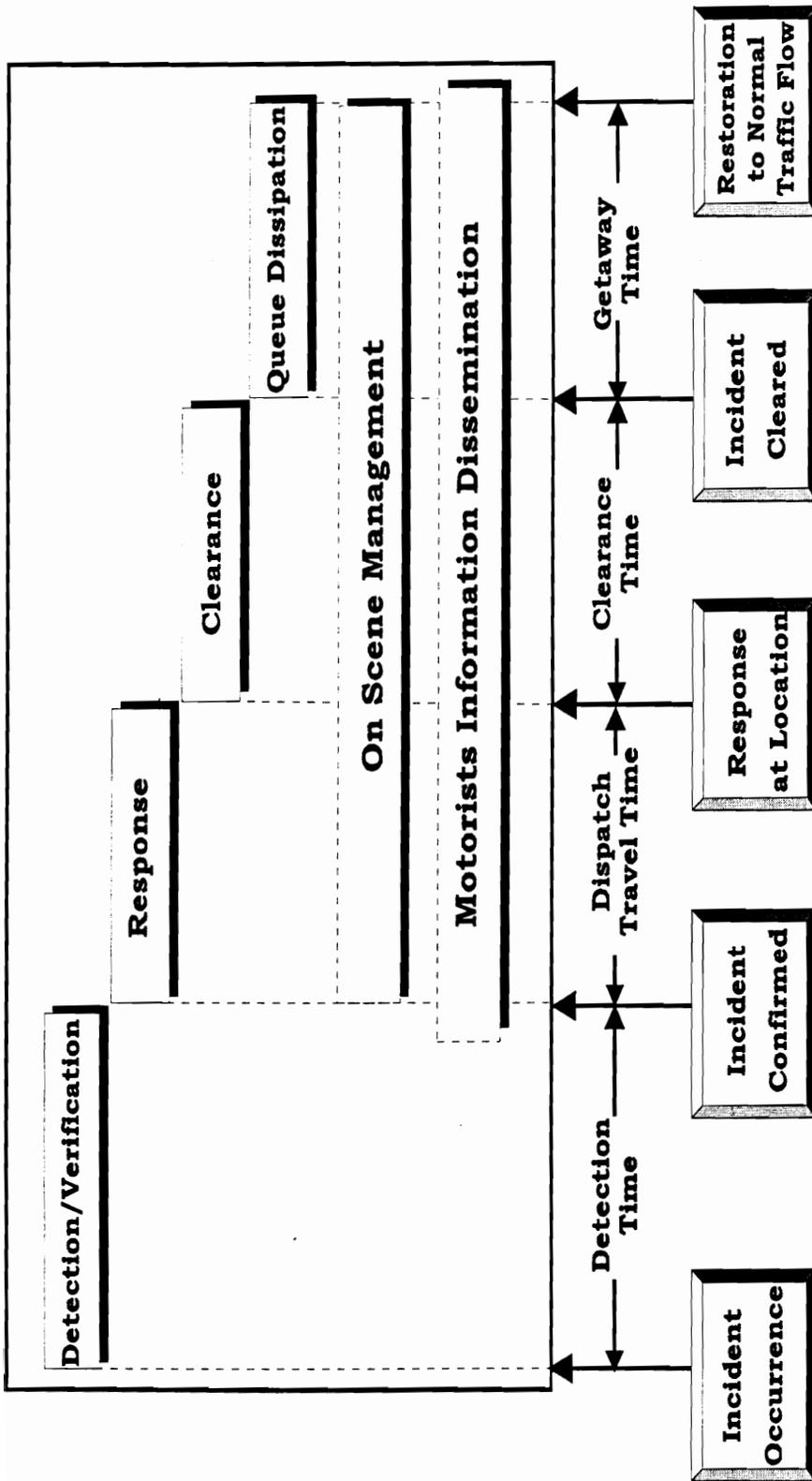


Figure 5-3. Steps of Incident Management Process

was gathered in a two-phase data collection effort. Figure 5-4 shows a sample survey form used for phase 2 data collection (Subramaniam et al. 1994).

A detailed analysis of the data by severity of incidents, operational factors, roadway types, and environmental factors was performed to determine significant factors affecting incident clearance times. Those factors are then used to develop duration estimation/decision trees for each incident type using tree structured regression method (Breiman et al., 1984, Steward, 1996).

Tree Structured Regression Method

For tree structured regression, the response variable is a continuous variable. The tree is constructed by partitioning the data set into a sequence of subsets and terminal nodes. In each terminal node, the predicted response is constant. Within a node k , the response variable y is approximated by its mean value \bar{y}_k :

$$\bar{y}_k = \frac{1}{N_k} \sum_{i \in \text{node } k} y_i \quad (5-1)$$

where N_k is the number of observations in node k .

The Regression tree model approximates the response variable by a step function whose values are the mean values of the response function on the various subsets comprising the tree structure. The cost function associated with a tree is normally taken to be the mean squared error:

$$\text{Cost} = \frac{1}{K} \sum_{k=1}^K \sum_{i \in \text{node } k} (y_i - \bar{y}_k)^2 \quad (5-2)$$

where y_i is the value of observation i .

The Regression tree model is a good choice for the estimation/decision tree development since the model differs from other types of stepwise regression and discriminate analysis models in several ways including:

INCIDENT MANAGEMENT OPERATIONS SURVEY <i>Please do not fill out surveys for disabled cars on shoulders</i>													
Your Last Name: _____ Agency: _____	LIGHT CONDITIONS <div style="display: flex; justify-content: space-around;"> <div style="border: 1px solid black; padding: 2px 10px;">Bright</div> <div style="border: 1px solid black; padding: 2px 10px;">Satisfactory</div> <div style="border: 1px solid black; padding: 2px 10px;">Dark</div> </div>												
Date: _____ Time: _____ Location: _____	TEMPERATURE <div style="display: flex; justify-content: space-around;"> <div style="border: 1px solid black; padding: 2px 10px;"><45 deg</div> <div style="border: 1px solid black; padding: 2px 10px;">45<t<85</div> <div style="border: 1px solid black; padding: 2px 10px;">>85 deg</div> </div>												
INCIDENT TYPE: Select One <table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="padding: 2px;">Road Hazard <input type="checkbox"/></td> <td style="padding: 2px;">Fatal Incident <input type="checkbox"/></td> </tr> <tr> <td style="padding: 2px;">Property Damage <input type="checkbox"/></td> <td style="padding: 2px;">Cargo Spill <input type="checkbox"/></td> </tr> <tr> <td style="padding: 2px;">Personal Injury <input type="checkbox"/></td> <td style="padding: 2px;">Hazmat <input type="checkbox"/></td> </tr> <tr> <td style="padding: 2px;">Disabled Truck <input type="checkbox"/></td> <td style="padding: 2px;">Weather Related <input type="checkbox"/></td> </tr> <tr> <td style="padding: 2px;">Vehicle Fire <input type="checkbox"/></td> <td style="padding: 2px;">Const./Maint. <input type="checkbox"/></td> </tr> <tr> <td style="padding: 2px;">Disabled Car in Travel Lane <input type="checkbox"/></td> <td></td> </tr> </table>	Road Hazard <input type="checkbox"/>	Fatal Incident <input type="checkbox"/>	Property Damage <input type="checkbox"/>	Cargo Spill <input type="checkbox"/>	Personal Injury <input type="checkbox"/>	Hazmat <input type="checkbox"/>	Disabled Truck <input type="checkbox"/>	Weather Related <input type="checkbox"/>	Vehicle Fire <input type="checkbox"/>	Const./Maint. <input type="checkbox"/>	Disabled Car in Travel Lane <input type="checkbox"/>		LANE INFORMATION Total Number of Lanes in the Direction of the Incident → <input style="width: 40px; height: 20px;" type="text"/> (Do not count both directions of the roadway)
Road Hazard <input type="checkbox"/>	Fatal Incident <input type="checkbox"/>												
Property Damage <input type="checkbox"/>	Cargo Spill <input type="checkbox"/>												
Personal Injury <input type="checkbox"/>	Hazmat <input type="checkbox"/>												
Disabled Truck <input type="checkbox"/>	Weather Related <input type="checkbox"/>												
Vehicle Fire <input type="checkbox"/>	Const./Maint. <input type="checkbox"/>												
Disabled Car in Travel Lane <input type="checkbox"/>													
INCIDENT DETAILS # of Cars Involved → <input style="width: 40px;" type="text"/> # of Trucks Involved → <input style="width: 40px;" type="text"/> # of Personal Injuries → <input style="width: 40px;" type="text"/> # of Fatalities → <input style="width: 40px;" type="text"/>	SHOULDER & HOV INFORMATION Is a Right Shoulder Present?..... <input type="checkbox"/> Yes <input type="checkbox"/> No Is a Left Shoulder Present?..... <input type="checkbox"/> Yes <input type="checkbox"/> No Is the Incident on HOV Lane?.... <input type="checkbox"/> Yes <input type="checkbox"/> No (i.e. Is it HOV During Incident Occurrence?)												
RESOURCES USED FOR CLEARANCE (Enter Number of each used for the Clearance) # of Police Vehicles → <input style="width: 40px;" type="text"/> # of Fire Engines → <input style="width: 40px;" type="text"/> # of Ambulances → <input style="width: 40px;" type="text"/> # of Wreckers → <input style="width: 40px;" type="text"/> # of Arrowboards → <input style="width: 40px;" type="text"/> MedVac Helicopter <input type="checkbox"/> Yes <input type="checkbox"/> No Fixed Post Traffic Control <input type="checkbox"/> Yes <input type="checkbox"/> No Alternate Route Established..... <input type="checkbox"/> Yes <input type="checkbox"/> No	LANE AND SHOULDER BLOCKAGE DUE TO INCIDENT Please Cross (X) the Lanes and Shoulder(s) Blocked by the Incident. <table style="width: 100%; border-collapse: collapse;"> <tr> <td style="border-right: 1px solid black; padding: 2px;">Left Shoulder</td> <td style="border-right: 1px solid black; padding: 2px;">Lane 6</td> <td style="border-right: 1px solid black; padding: 2px;">Lane 5</td> <td style="border-right: 1px solid black; padding: 2px;">Lane 4</td> <td style="border-right: 1px solid black; padding: 2px;">Lane 3</td> <td style="border-right: 1px solid black; padding: 2px;">Lane 2</td> <td style="border-right: 1px solid black; padding: 2px;">Lane 1</td> <td style="padding: 2px;">Right Shoulder</td> </tr> </table>	Left Shoulder	Lane 6	Lane 5	Lane 4	Lane 3	Lane 2	Lane 1	Right Shoulder				
Left Shoulder	Lane 6	Lane 5	Lane 4	Lane 3	Lane 2	Lane 1	Right Shoulder						
LAND USE TYPE: Select One <div style="display: flex; justify-content: space-around;"> <div style="border: 1px solid black; padding: 2px 10px;">Open Land</div> <div style="border: 1px solid black; padding: 2px 10px;">Urban/Bldgs</div> <div style="border: 1px solid black; padding: 2px 10px;">Bridge</div> <div style="border: 1px solid black; padding: 2px 10px;">Tunnel</div> </div>	CLEARANCE TIME What was the clearance time for this Incident? <div style="display: flex; justify-content: space-around; align-items: center;"> <div style="border: 1px solid black; width: 40px; height: 20px; display: flex; align-items: center; justify-content: center;"> </div> <div style="margin: 0 10px;">Hour(s) &</div> <div style="border: 1px solid black; width: 40px; height: 20px; display: flex; align-items: center; justify-content: center;"> </div> <div style="margin-left: 10px;">Minute(s)</div> </div>												
PREVAILING WEATHER <div style="display: flex; justify-content: space-around;"> <div style="border: 1px solid black; padding: 2px 10px;">Clear/Cloudy</div> <div style="border: 1px solid black; padding: 2px 10px;">Rain</div> <div style="border: 1px solid black; padding: 2px 10px;">Foggy/Misty</div> </div> <div style="display: flex; justify-content: space-around; margin-top: 5px;"> <div style="border: 1px solid black; padding: 2px 10px;">Snow</div> <div style="border: 1px solid black; padding: 2px 10px;">Sleet/Ice</div> </div>													

Figure 5-4. A Sample Survey Form Used for Incident Data Collection In NOVA

- It is non-parametric and it is not necessary to specify a model form (linear, or Nonlinear), Complex interactions among the variables are easily identified, and
- It is able to include a large number of variables.

Construction of Regression Tree for incident duration estimation

To construct a tree, the following issues need to be addressed:

- A way to select a split at every intermediate node
- A rule for determining when a node is terminal
- A rule for assigning a value to every terminal node

In constructing incident duration estimation/decision trees, splitting of intermediate node was made by minimizing the mean squared error. This decision making process for splitting was also aided with an extensive number of trials of ANOVA (Analysis of Variance) tests and t-tests. When non-binary splits were utilized, non-binary splits of a subset were tested to be significantly different in mean values for each sub-subset using ANOVA. A sample ANOVA table used for non-binary splitting is shown in Table 5-1. The node is considered to be terminal when there are not enough data points in the subset to further split the set and make statistically meaningful conclusion. The average clearance time of the samples in each subset or node is assigned to respective subset/node as the response (or estimation) value.

Table 5-1. Significance of the No. of Police Vehicles Involved on CLT¹ For Property Damage without truck and 1-3 cars involved (ANOVA Result)

SUMMARY						
<i>POLICEVEH</i>	<i>Count</i>	<i>Sum</i>	<i>Average</i>	<i>Variance</i>		
1	267	10062	37.68539	403.9232		
2	54	2466	45.66667	295.5849		
3, 3+	24	1550	64.58333	995.471		
ANOVA (Single Factor)						
<i>Source of Variation</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-value</i>	<i>F crit</i>
Between Groups	17444.58	2	8722.291	20.43091	4.15E-09	3.022123
Within Groups	146005.4	342	426.9164			
Total	163450	344				

1. CLT= Clearance Time, used as a proxy of Incident Duration

Figure 5-5 is a sample estimation/decision tree developed for property damage incidents.

Statistical analysis of data has also shown that the duration of incidents is normally distributed for homogenous sub-groups of incidents that are categorized according to the most significant affecting factors. This property of incidents is then used to estimate the likelihood of an incident to last more than a specific period of time.

The findings of incident duration and distribution study in the Northern Virginia area are found to be very comparable to the results of other studies. The incident duration for several types of incidents was very close to that presented by other previous studies elsewhere in the country (Ozbay et al., 1996). The incident duration results of the Northern Virginia Study can provide good estimation, and the duration can be used as the optimization horizon of the COROPT optimization. It should be pointed out that tuning and validation of the duration model using incident data of other areas is necessary, for incident duration is affected by demand level, incident clearance procedure, and traffic network characteristics, and, thus, site-specific.

Another important aspect of the duration prediction is that due to the employment of new incident management and traffic management strategies, such as diversion and signal retiming, time spent on incident clearance is reduced, the queue length and delay also decrease, the overall incident duration is thus reduced. Proper adjustment needs to be made for incident duration prediction, and for optimization time selection for the COROPT program.

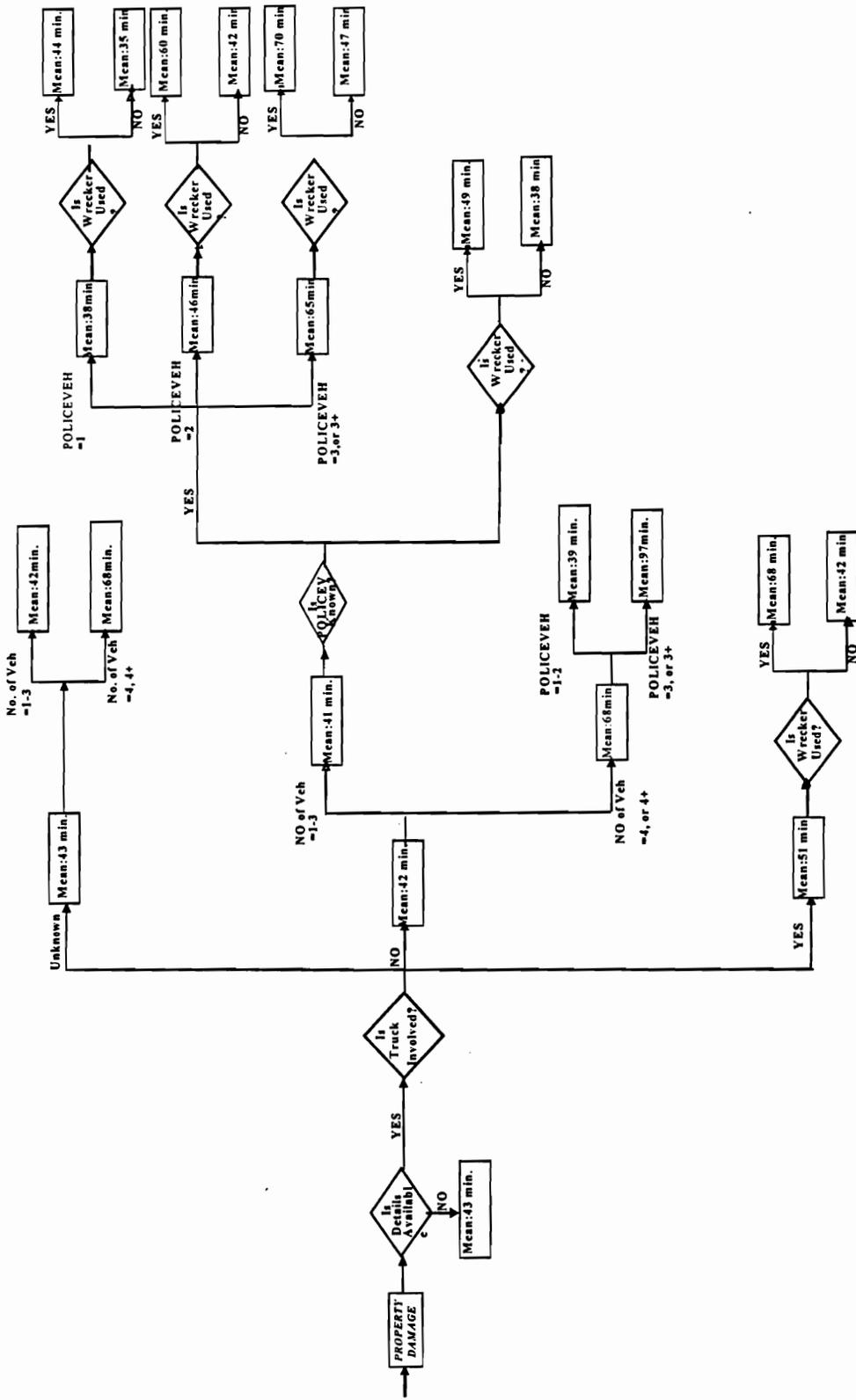


Figure 5-5. A Sample Estimation Tree for Incident Duration Estimation of Property Damage Incidents

5.3.2 Determining Time Periods with Stable Prevailing Traffic and Incident Conditions

The optimization time is divided into several time periods with the same prevailing traffic and incident conditions. Division of the optimization time into time periods is determined by the two major factors:

- Demand changes during different time-of-day
Advent or departure of the peak hour is the major factor of demand level change. The historical traffic data record can be used to predict the demand level pattern changes during different time of day.
- Dynamic Lane Clearance process
The lane blockage and available capacity change with the dynamic lane clearance process. Initially, several travel lanes may be blocked by incident and the available capacity is limited. As the incident clearance progresses, an additional lane that was previously blocked will be cleared and open for traffic after certain time, the available capacity of that roadway section with incident will increase accordingly.

As part of the research of the incident management project at Virginia Tech, a study was performed on the phenomenon of dynamic lane clearance. Usually when incidents block one or more lanes, the entire incident is not cleared at once. Incident management crews often open lanes before the entire incident is cleared in order to reduce delays. To model this phenomenon, a knowledge base was created by surveying incident management personnel in the Northern Virginia region. A sample page of this survey is shown in Figure 5-6 (Mastbrook, et al., 1996). The collected data was analyzed, and the results on duration of blockage of lanes and shoulder, for different lane blockage combinations, different incident type, and different incident duration were reached. The results and approach can be used to determine the available capacity of roadway incident section at different stages of incidents.

**60 Minute Property Damage Incidents(No Injury)
4-Lane Cases**

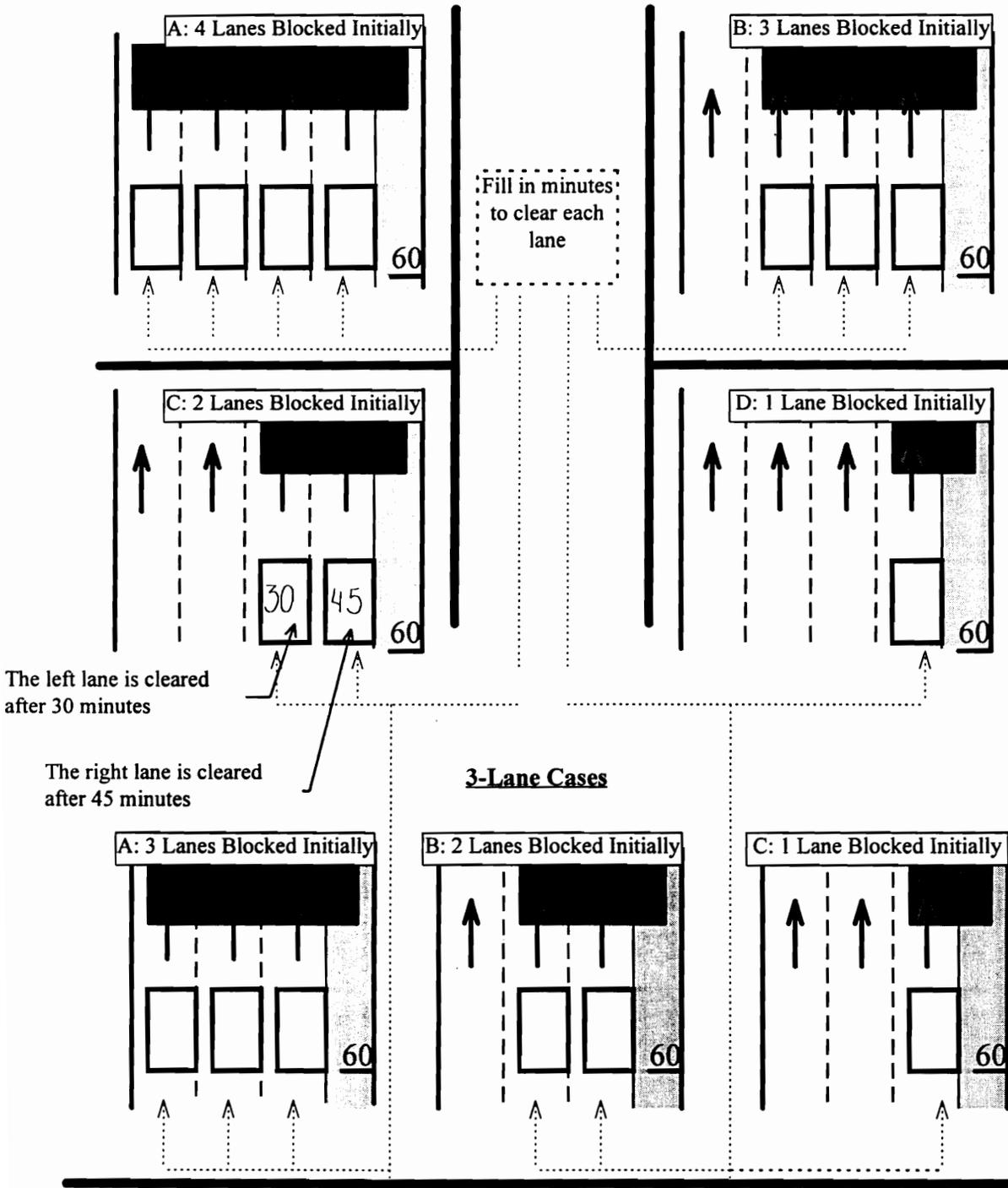


Figure 5-6. A Sample Page of the Survey on Lane Clearance

Knowing the changing pattern of the traffic demand over time, along with the knowledge of lane blockage situations at different stages, we can divide the optimization time into time periods of the same prevailing traffic and incident conditions. As illustrated in Figure 5-7, during the optimization time OT , the demand level first rises to peak, then drops. Meanwhile, the available capacity (for incident section) increases step by steps due to the opening of additional previously blocked lane. The optimization time needs to be divided into eight time periods ($TP1$ through $TP8$) so that each time period has the same traffic and incident condition.

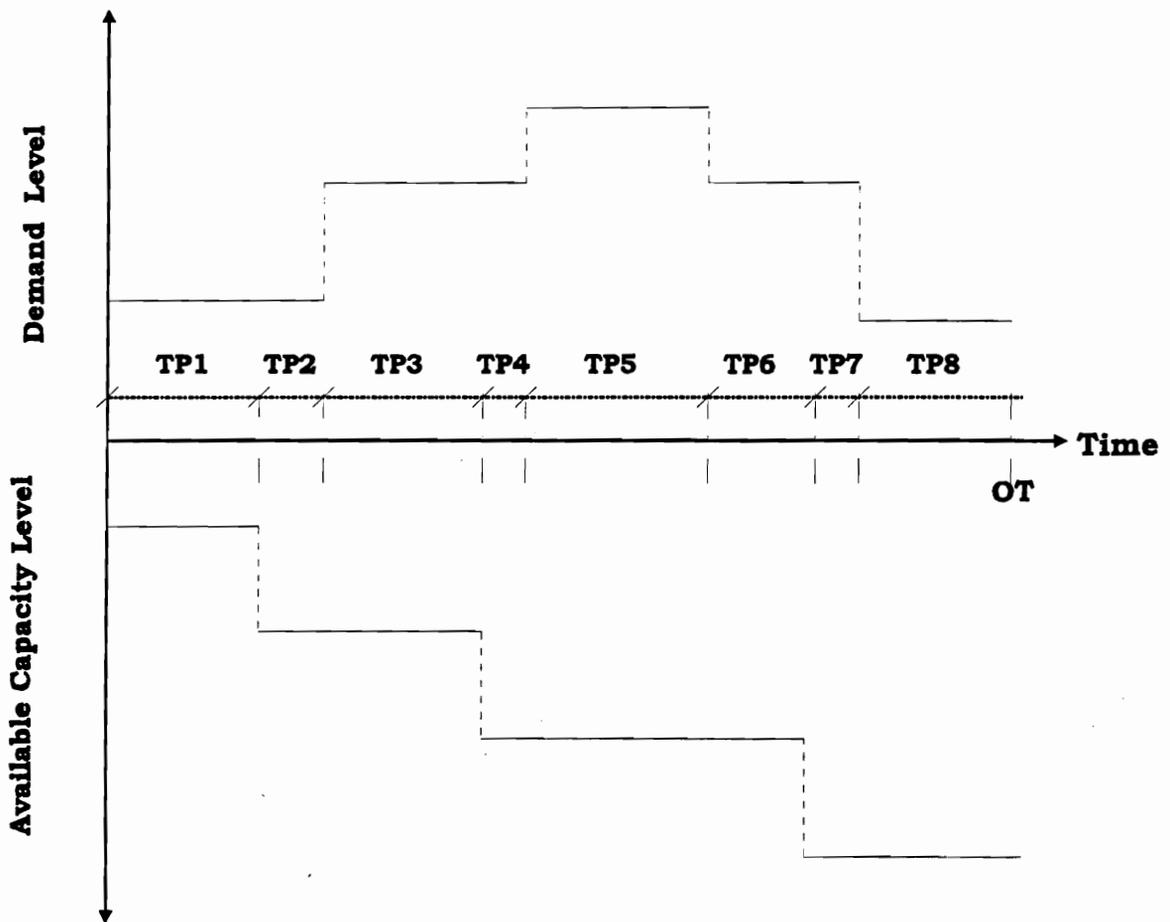


Figure 5-7. Dividing Optimization Time into Time Periods with the Same Prevailing Traffic and Incident Condition

5.3.3 Determining Available Capacity for Incident Section

The available capacity of freeway incident section is determined by the availability of the service of travel lanes and shoulder. A four-lane freeway with two lanes blocked has less than 50% of original capacity since drivers tend to slow down when passing by the incident site, and the available capacity of the section is thus further reduced. A freeway having all lanes open but having shoulder uncleared does not have 100% of its original capacity due to the same reason mentioned.

The available capacity of the freeway section under incident condition can be determined by the following table due to Lindley (1987).

Table 5-2. Fraction of Freeway Section Capacity Available Under Incident Conditions

No. of Lanes Per Direction	Should Disablement	Should Accident	1-Lane Blocked	2-Lane Blocked	3-Lane Blocked
2	0.95	0.81	0.35	0.00	-----
3	0.99	0.83	0.49	0.17	0.00
4	0.99	0.85	0.58	0.25	0.13
5	0.99	0.87	0.65	0.40	0.20
6	0.99	0.89	0.71	0.50	0.25
7	0.99	0.91	0.75	0.57	0.36
8	0.99	0.93	0.78	0.63	0.41

5.3.4 Determining Weights of Objective Function Components

In the optimization objective function formulation (Equation 4-1a), there are weights $\xi_{m_j^i}$, and ψ associated with components of the corridor. $\xi_{m_j^i}$ is the weight for movement m of j th phase at i th intersection. By assigning different values of $\xi_{m_j^i}$, we prioritize the movements in terms of importance, or sacrifice.

In the case of having a severe incident on the freeway and traffic being diverted via the arterial, we will never allow that a large portion of capacity for the movement on

the diversion route is not used. In addition, we do not want that queue spills back and blocks the upstream intersections. A heavy penalty should be associated with the objective function to prevent the cases of either having considerably large portion of unused capacity, or the queue length being extremely long. This is to say that a higher ξ_{m_j} value should be assigned to all movements on possible diversion routes.

On the other hand, a small ξ_{m_j} is normally assigned to any side street movement, or the movement that will not be on the diversion route. These movements have a lower demand level, and some queue buildup is expected. Diversion is always at the cost of sacrificing some travelers' benefit (mainly, travel time).

The values of ψ ξ_{m_j} 's are relative values, as can be seen in the formulation of the objective function. Once one value is fixed, the others are determined to take values relative to the fixed value. A value of 1.0 for all side street movements, and movements that will not be on the diversion routes, and a value of 1.2 for all movements on the possible diversion routes can normally be used in the objective function.

The determination of the value of ψ takes a different approach. For under-saturated traffic flow conditions, the components for arterial movement, and the objective function component for freeway traffic may all represent the square of queue length. However, the queue lengths for arterial movements and queue length on freeway may have huge difference, and taking the square makes the difference even bigger. ψ comes to play a role in determining how much traffic should be diverted. Unlike the situation that ξ_{m_j} values for movements of different levels of importance can be readily seen and selection of relative weights can be easily done, the relationship between ψ and ξ_{m_j} 's is not explicit. It depends on the demand level, as well as number of intersections in the corridor system.

Table 5-3 shows the effect of ψ value on the throughput of the corridor (defined as the total vehicles exited the corridor on the direction of freeway incident link during optimization period), the freeway delay, the arterial delay, and the total delay, as well as the total diversion rate for the optimal solutions. A corridor with eight intersections on arterial is optimized giving different weights of objective value for freeway component. The weights for all traffic movements at arterial intersections, ξ_{m_j} 's, are all fixed at 1.0. Figure 5-8 shows the effect of the objective function weight for the freeway component on the freeway delay, the arterial delay, and the total delay. Figure 5-9 illustrates the effect of objective function weight for the freeway component on the corridor throughput. Figure 5-10 presents the effect of the objective function weight for the freeway component on total volume that is diverted from the freeway, bypasses the freeway incident section via the arterial, and goes back to the freeway.

Figure 5-11 illustrates the effect of the objective function weight for the freeway component on the corridor throughput at three different traffic demand/incident lane blockage levels for the same corridor.

It can be seen from Table 5-3, and Figures 5-8 through Figure 5-11 that the weight of the freeway component of the objective function ψ plays an important role in determining the nature and quality of the resulting optimal solution for corridor control. The effect of ψ on the resulting plan is summarized below:

- When $\psi = 0$, the freeway portion is actually dropped from consideration in the objective function. No traffic will be diverted away from the freeway to bypass the incident section. The freeway delay is at its maximum level since no volume is diverted. On the other hand, the arterial delay is at its minimum level since there is no extra traffic diverted from the freeway. At $\psi = 1$, virtually the same results hold as in case of $\psi = 0$.

Table 5-3. The Effect of the Objective Function Weight for the Freeway Component on Throughput, Delay, and Diverted Volume

Weight	Throughput (Vehicles)	Freeway Delay (Vehicle.Seconds)	Arterial Delay (Vehicle.Seconds)	Total Delay (Vehicle.Seconds)	Total Diversion Rate (Vehicles/Hour)
0	1293	187283	325838	513121	0
1	1293	187214	325883	513097	0
5	1478	89870	404567	494437	683
10	1421	69245	429822	499067	732
15	1517	32184	462331	493615	898
20	1473	30948	470889	501837	925
25	1458	26666	501022	527688	971
30	1485	20836	504268	525104	951
50	1316	9182	586874	596056	1030
100	1261	6315	682214	688523	1048

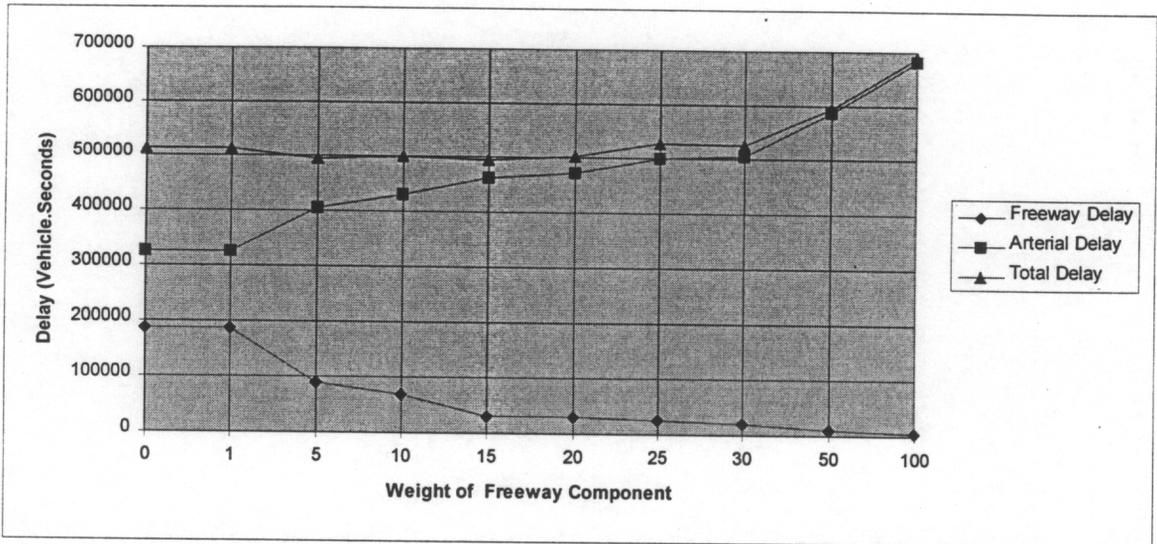


Figure 5-8. The Effect of the Objective Function Weight for the Freeway Component on Delays

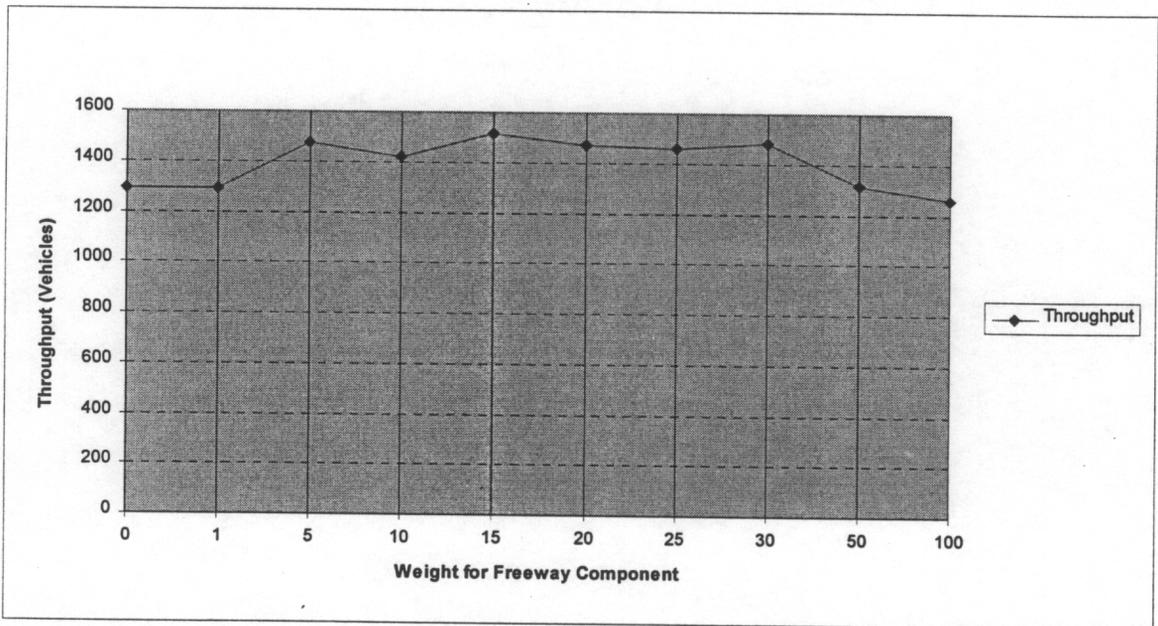


Figure 5-9. The Effect of the Objective Function Weight for the Freeway Component on the Corridor Throughput

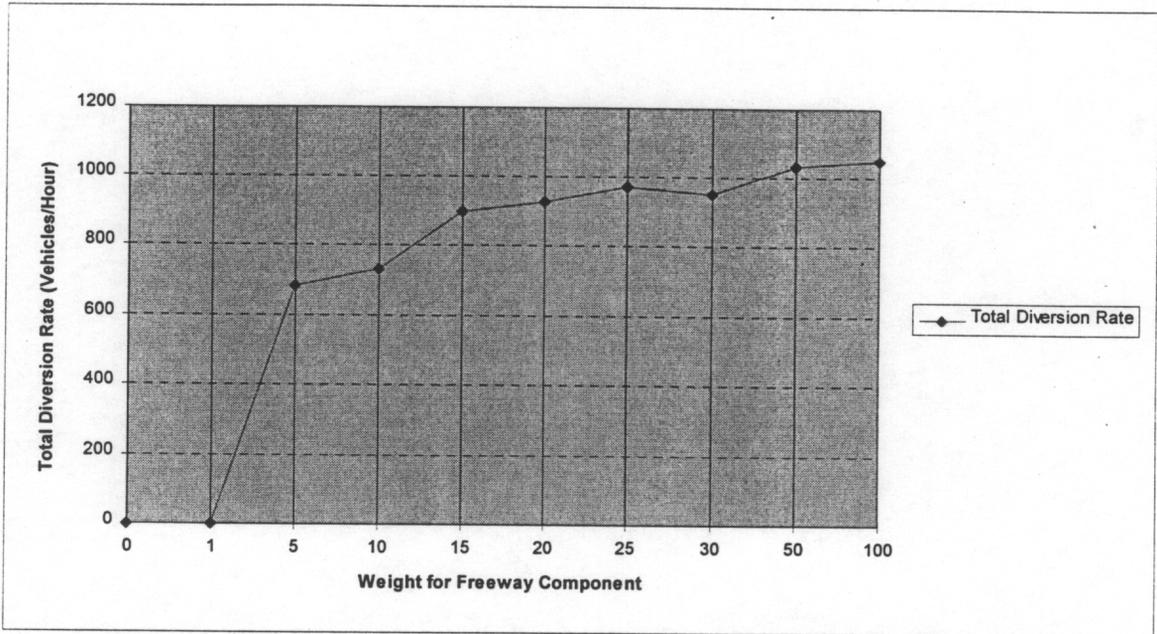


Figure 5-10. The Effect of the Objective Function Weight for the Freeway Component on the Total Diversion Rate

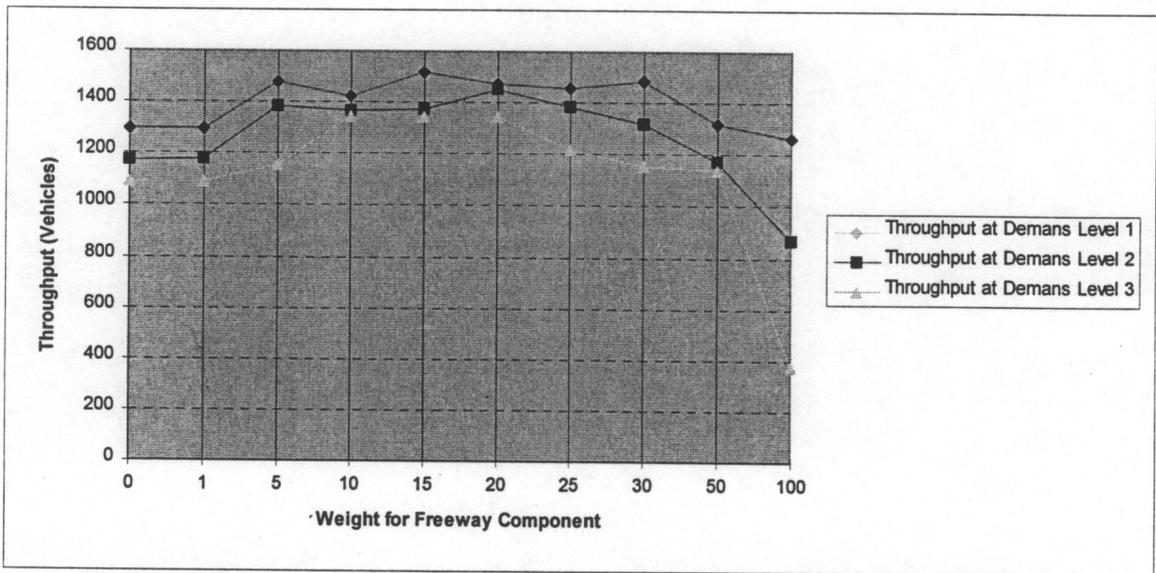


Figure 5-11. The Effect of the Objective Function Weight for the Freeway Component on the Corridor Throughput at Different Traffic Demand Levels

- As the value of ψ increases, more and more traffic is diverted from the freeway (Figure 5-10). Consequently, the delay on the freeway decreases, and meanwhile the delay on the arterial increases due to the extra traffic as the result of diversion (Figure 5-8).
- As the value of ψ increase from 0, the throughput of the corridor starts increasing. However, after ψ reaches certain value (30 to 50 in Figure 5-9), increasing ψ value leads to dropping in the corridor throughput. This is due to the fact that too much traffic is diverted from the freeway, as the consequence, the arterial network is over-congested, and the total throughput decreases.
- As the value of ψ increase from 0, a trend of decreasing total delay can be spotted in Figure 5-8. However, as ψ continue to increase after certain point, the total delay of the corridor increases. This is the result of an extremely large amount of delay on arterial intersections due to congestion caused by diverted traffic.

To increase the corridor throughput, and to reduce the total delay are goals of corridor traffic control. It would be ideal to select a ψ value with which the optimization model would yield a solution with maximum throughput and minimum total delay at the same time. Unfortunately, this is not always achievable. Especially when the normal demand level on the arterial is high, increasing throughput can only be achieved at the cost of increased total delay.

It is shown that when ψ is at the value between 10 and 25, a high level of throughput can be obtained, and a relatively low level of total delay is maintained. Moreover, the diversion volume, delay, and the throughput are not vary sensitive to actual ψ value within this range. Multiple trial results with different demand levels also support this argument (Figure 5-11). For arterial with 2 to 10 intersections, a ψ value in this range yields good MOE's (throughput and total delay).

A default value of $\psi = 20$ is, therefore, adopted for the COROPT program. However, as the corridor physical conditions (including the number of intersections), traffic demand and incident conditions change, $\psi = 20$ may become a not very good choice from time to time. To deal with this situation, the COROPT program provides a function to “select” an optimal ψ value. The selection is achieved by evaluating the throughput values of optimal solutions from different ψ values, and choosing the ψ value that yields the largest throughput. The corridor throughput is used as the criteria of selecting ψ , since obtaining maximum throughput is the primary goal of corridor traffic control. Diversion increases the throughput, but not necessary decreases the total delay due to the increased travel, and extra stop and delay at intersections. This gives the edge to corridor throughput over total delay as the better candidate for selecting ψ value in the optimization model.

5.4 LIMITATIONS OF THE COROPT PROGRAM

There are some restrictions on the corridor size, intersection configurations and signal timing settings in the COROPT program. Some of the restrictions are listed below:

- Maximum number of intersections on the arterial: 20
- Maximum number of phases for a cycle: 6
- Maximum number of movements for each phase: 4
- Maximum number of inequality constraints: 200
- Maximum number of equality constraints: 50
- Maximum number of freeway incidents: 1

However, there is no restriction on the number of time periods over the optimization time, neither is the constraint on the length of optimization time. The same arrays

are used in the program for variables of each time period. The increased number of time periods does not require extra memory in the COROPT program.

6. CASE STUDY AND MODEL VALIDATION

6.1 INTRODUCTION

The previous chapters dealt with the mathematical formulation and subsequent development of solution technique for a nonlinear programming based approach for determining corridor traffic control measures. A computer program developed to implement the model approach was also presented.

This chapter deals with case studies, and the testing and validation of this model. The optimization results of model applications to corridors with different physical configurations, different demand level, and different incident conditions will be presented and analyzed. The validation procedure includes a comprehensive testing of the model's performance, sensitivity to inputs, robustness, and comparative evaluation with the commonly used signal optimization package and traffic simulation/optimization package.

6.2 CASE STUDY

A number of COROPT program runs were conducted to test the performance of the model and program under different combinations of corridor, traffic, and incident conditions.

6.2.1 Case Study for Arterial with Different Number of Intersections Considered for Diversion Purpose

To investigate the advantage of using multiple on-ramps and off-ramps for diversion, cases of 2, 4, 8, and 16 intersections on the arterial of the corridor were optimized using the COROPT program. The results and analysis are presented.

The following data is used for model inputs:

- The number of phases for each arterial intersection: 4
- The number of left-turn lanes: 1
- The number of right-turn and through lanes: 2 for arterial main directions, otherwise
- Average travel time on arterial link between intersections: 160 seconds (assuming intersections are equally spaced at two-mile interval, average speed on the arterial=45MPH)
- Minimum phase length: 10 seconds
- Maximum phase length: 60 seconds
- Minimum cycle length: 60 seconds
- Maximum cycle length: 180 seconds
- Maximum allowable queue length on main directions: 50 vehicles
- maximum ramp metering rate: 1080 vehicles/hour
- Normal rates: 108 vehicles/hour for side street left-turns; 288 vehicles/hour for side street through and right-turns combined; 144 vehicles/hour for arterial left-turns; 720 vehicles/hour for arterial through and right-turns
- Arrival rate to the freeway: 3000 vehicle/hours
- Available Capacity: 1400 vehicle/hour (left lane blocked)
- Optimization time: 1200 seconds

Half of the intersections are upstream of the freeway incident location; the other intersections are downstream of that freeway incident location. Figure 6-1 illustrates a corridor with eight arterial intersections. Figure 6-2 shows the phasing used in all case studies except the case study for phasing where other phasing scheme will also be discussed. Tables 6-1 through 6-4 present the results of optimal diversion rates and signal timing plans for arterial with 2, 4, 8, and 16 intersections respectively; Table 6-5 summarizes the results of total diversion, optimal cycle lengths, and MOEs for all cases with different number of intersections on the arterial.

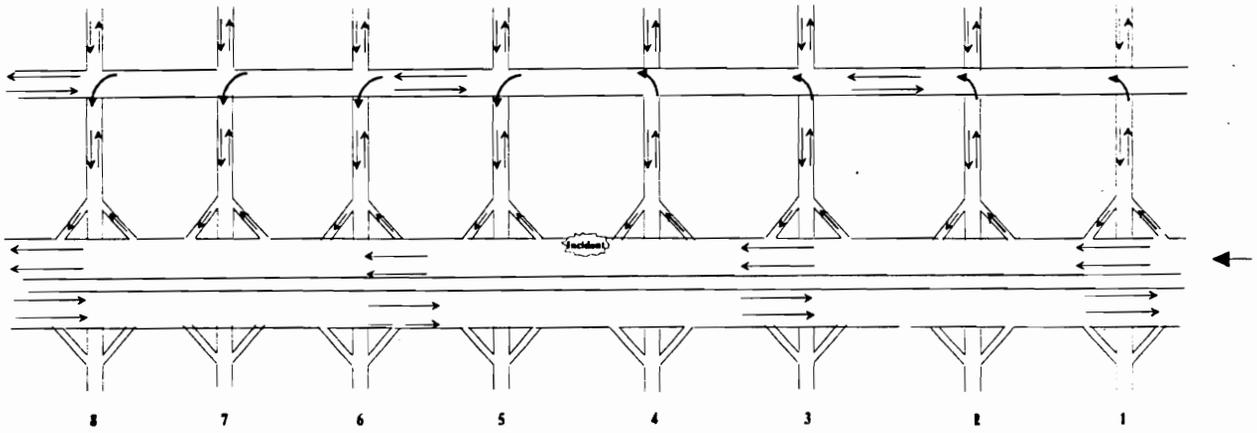
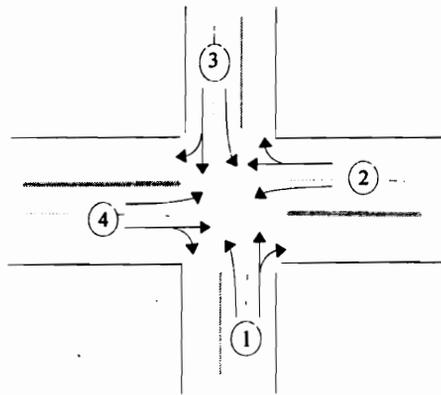


Figure 6-1. Illustration of a Corridor with Eight Intersections



Phasing			
Phase 1	Phase 2	Phase 3	Phase 4

Figure 6-2. Illustration of Phasing Used in Case Studies

Table 6-1. Optimal Diversion Rates and Timing Plans for a Two-Intersection Arterial

	Phase Length (Seconds)				Offset (Seconds)	Div. Rates (Vehicles/hour)
	Phase 1	Phase 2	Phase 3	Phase 4		
Intersection 1	52	24	19	25	-	936
Intersection 1	17	60*	19	24	100	936

* Maximum phase length

Table 6-2. Optimal Diversion Rates and Timing Plans for a Four-Intersection Arterial

	Phase Length (Seconds)				Offset (Seconds)	Div. Rates (Vehicles/hour)
	Phase 1	Phase 2	Phase 3	Phase 4		
Intersection 1	46	27	22	25	-	683
Intersection 2	40	43	15	22	85	691
Intersection 3	18	60*	18	24	0	721
Intersection 4	18	59	19	24	5	652

* Maximum phase length

Table 6-3. Optimal Diversion Rates and Timing Plans for an Eight-Intersection Arterial

	Phase Length (Seconds)				Offset (Seconds)	Div. Rates (Vehicles/hour)
	Phase 1	Phase 2	Phase 3	Phase 4		
Intersection 1	32	25	22	21	-	364
Intersection 2	28	33	18	20	70	368
Intersection 3	26	40	15	19	30	372
Intersection 4	23	47	12	18	35	385
Intersection 5	14	51	15	20	35	678
Intersection 6	16	46	17	21	25	546
Intersection 7	19	37	21	23	80	267
Intersection 8	23	29	25	24	20	0

Table 6-4. Optimal Diversion Rates and Timing Plans for a Sixteen-Intersection Arterial

	Phase Length (Seconds)				Offset (Seconds)	Div. Rates (Vehicles/hour)
	Phase 1	Phase 2	Phase 3	Phase 4		
Intersection 1	34	25	30	31	-	165
Intersection 2	31	30	28	30	90	164
Intersection 3	30	35	27	29	90	173
Intersection 4	29	39	25	27	85	175
Intersection 5	27	44	23	26	80	189
Intersection 6	26	48	21	25	20	196
Intersection 7	24	53	19	23	25	209
Intersection 8	23	58	17	22	30	228
Intersection 9	19	55	20	25	20	642
Intersection 10	20	52	21	26	20	514
Intersection 11	23	44	25	27	20	301
Intersection 12	26	36	29	28	100	41
Intersection 13	27	34	30	29	0	1
Intersection 14	27	34	29	29	90	0
Intersection 15	27	34	29	29	85	0
Intersection 16	27	34	29	29	85	0

Table 6-5. Summary of Optimal Results for Corridors with Different Number of Arterial Intersections

	Two-Intersections	Four-Intersections	Eight-Intersections	Sixteen-Intersections
Optimal Cycle Length (Seconds)	120	120	100	120
Arterial Delay/Intersection (Vehicle.Seconds)	117302	95034	75466	60398
Freeway Delay (Vehicle.Seconds)	132771	45193	21745	20170
Corridor Throughput (Vehicles)	1021	1336	1342	1345
Total Diversion Rates (Vehicles/Hour)	936	1373	1491	1499

Figure 6-3 illustrates the effect of the number of intersections that are considered for diversion on delays; Figures 6-4 and 6-5 show the effect of the number of involved intersections on corridor throughputs and total diversion rates.

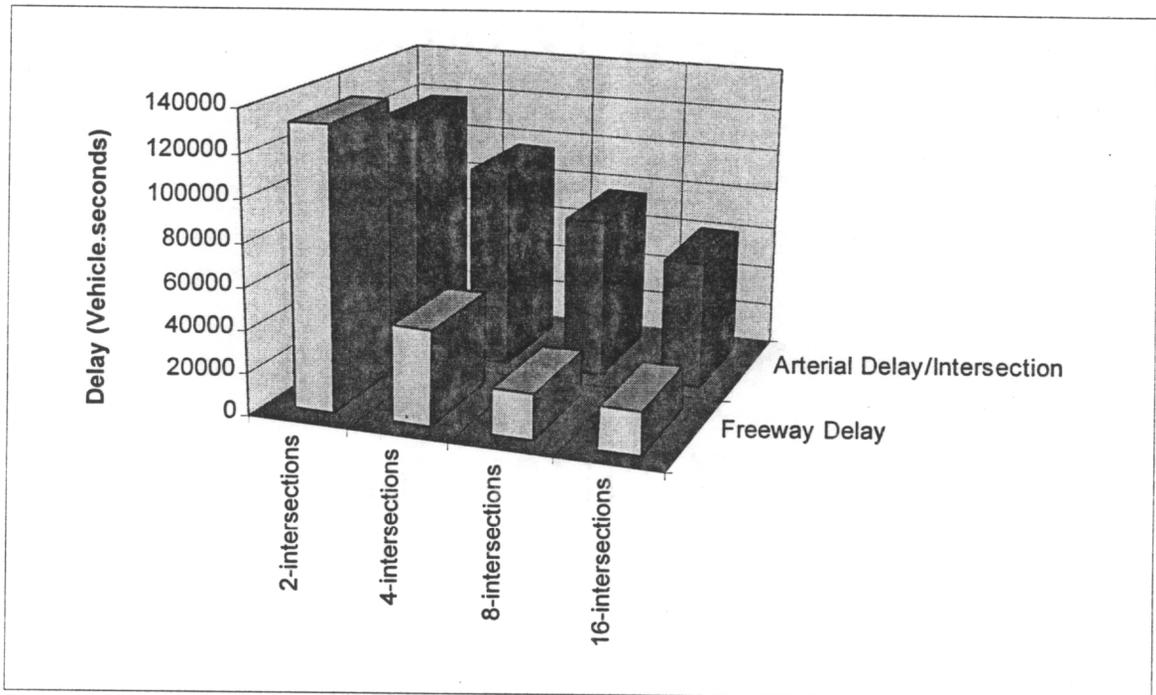


Figure 6-3. The Effect of the Number of Involved Intersections on Delays

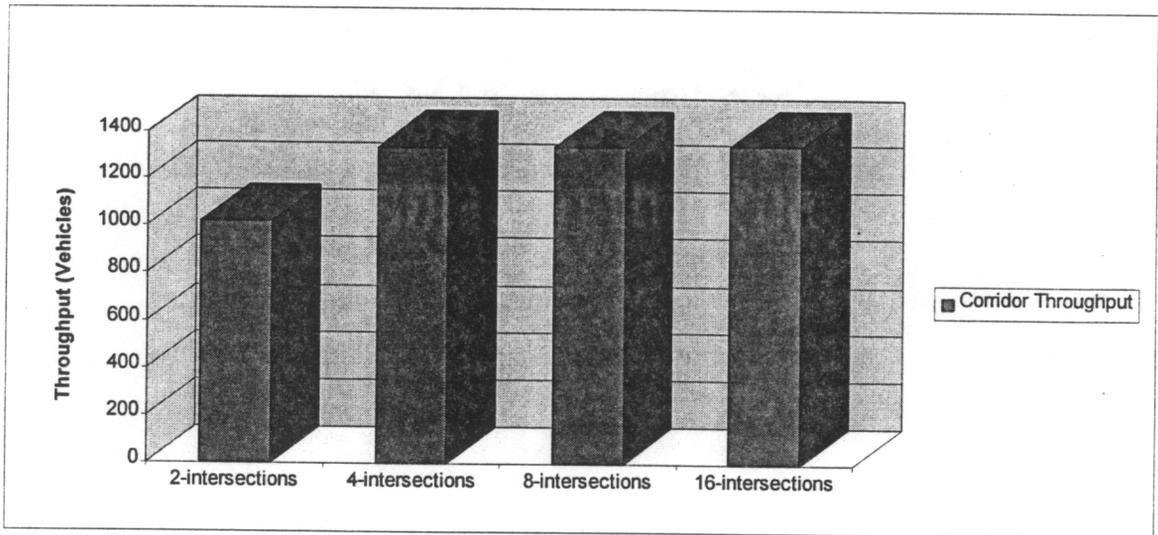


Figure 6-4. The Effect of the Number of Involved Intersections on the Corridor Throughputs

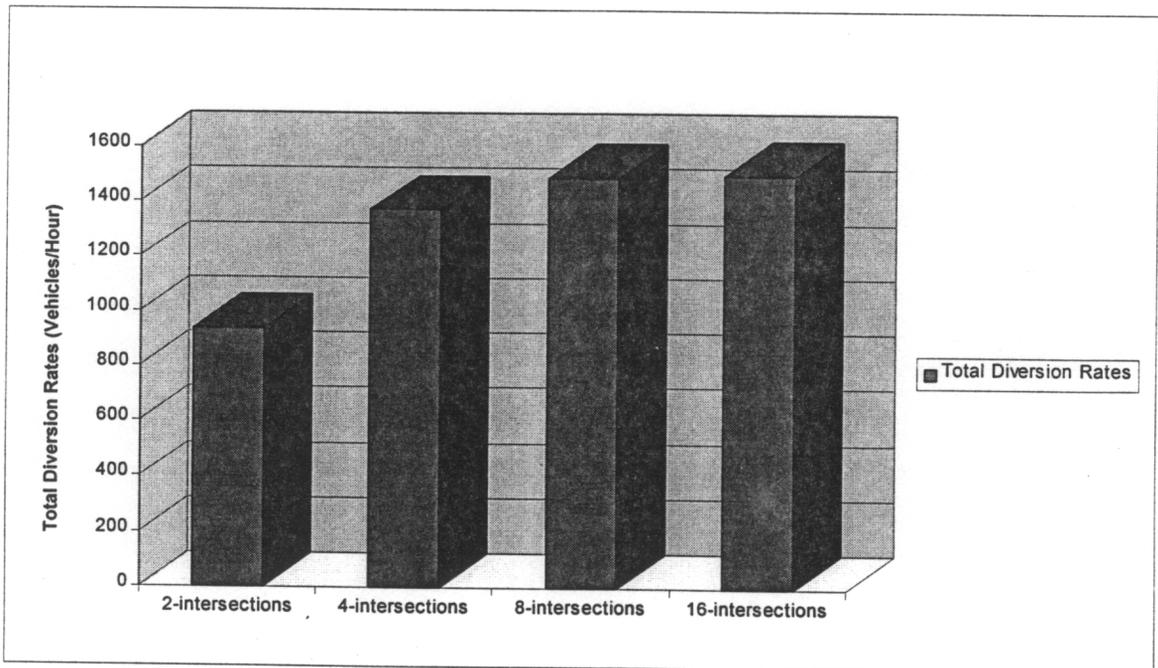


Figure 6-5. The Effect of the Number of Involved Intersections on Total Corridor Diversion Rates

The following can be drawn from results presented in Tables 6-1 through 6-4:

- The diversion volume is split between all the upstream off-ramps, with the ramp closer to the incident location carrying a little more diversion volume.
- For diverting traffic back to the freeway after incident location, the closer the on-ramp is to the incident location, the more diversion volume it carries. This is reasonable because the fact that incident-free freeway links have higher capacity and speed. This nature encourages traffic to take the freeway whenever it is possible.
- The diversion normally cannot be completed with maximum gains using only one on-ramp due to the limitation of the ramp capacity and possible extra delay to other approaches at that intersection. However, the results show that for the arterial that has more than 8 intersections, fewer number of on-ramps are used to divert traffic back to freeway than the number of off-ramps used to divert traffic from freeway to arterial.
- A tendency of allocating more green times to links carrying more traffic is evident.
- If the number of intersections involved in the diversion are small, there are huge differences in phase lengths in a cycle at different intersections. This is not desirable since it means that traffic of some movements will experience intolerantly long delay. In addition, the long phase length may approach its upper limit such as the case shown in Table 6-1 and Table 6-2. On the other hand, if more intersections are used for the diversion purpose, the intersections far away from incident locations normally have well-balanced phase lengths in a cycle. The closer the intersection is to the incident location, the bigger the differences in phase lengths in a cycle. This is because the links closer to incident location carry more diverted volume.

The advantages of using more intersection for diversion and retiming all those intersection signals can be clearly seen from Table 6-5, and Figures 6-3 through 6-5.

- For the case of two-intersection arterial, the maximum phase length limits the allocation of green time for diversion movement, thus limits the diversion rates. It should be pointed out that not only the maximum phase length or minimum phase length could be the limiting factor, but the capacity of the on- and off-ramp could also prohibit diverting more traffic.
- As more intersections are under corridor control (diversion and signal retiming), the average delay at an intersection decreases. It should be noted that average delay at an intersection rather than total arterial delay is used here, for it does not make sense to compare arterial delay among arterials having different number of intersections.
- As more intersections are under corridor control, the freeway delay decreases.
- As more intersections are under corridor control, the total diverted volume (from the freeway, via the arterial, and back to the freeway) increases.
- As more intersections are under corridor control, the corridor throughput increases. This is very important and critical to the corridor traffic flow operation. Increasing corridor throughput is one of the ultimate goals of traffic control for a corridor under incident conditions.

It is shown that when the number of intersections under corridor control increases and exceeds eight (four intersections upstream of the incident location, and four intersections downstream of the incident location), the MOEs of the corridor improves as delay continues to decrease, while the throughput and diverted volume increase. However, the gain is minimal as shown in Figures 6-4 and 6-5. Having more arterial intersections under corridor control is at the cost of having more equipment, more software and hardware requirements, and more traffic management and directing personnel, therefore, it is not economical to try having too many intersections under corridor control.

6.2.2 Case Study for Different Phasing Scenarios

The phasing schemes affect the traffic operation of surface streets. It will also affect diversion rates and the corridor throughput. To investigate the effect of different phasing schemes on corridor traffic conditions, the same corridor with four intersections on the arterial and the same traffic arrival rates as in 6.2.1 was optimized using the COROPT program for two different phasing schemes: phasing A and phasing B as shown in Figure 6-6.

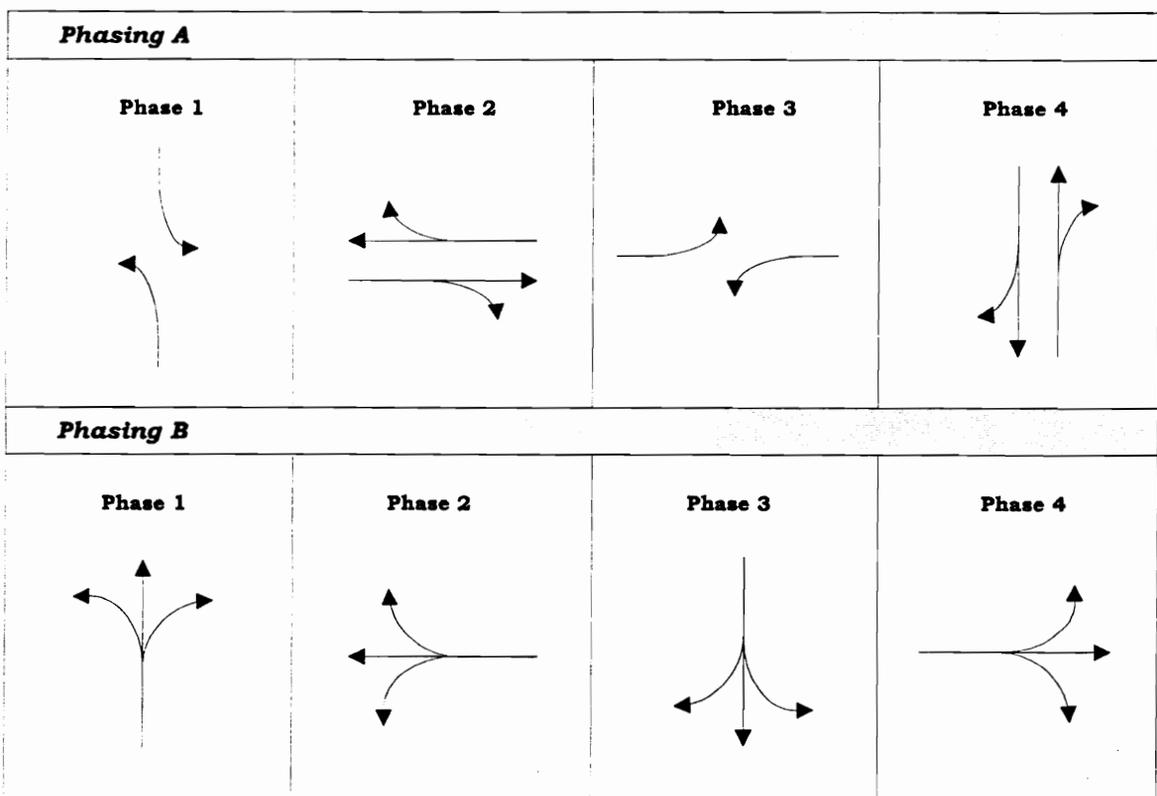


Figure 6-6. Two Phasing Schemes Used to Determine the Effect of Phasing on Corridor Traffic Flow Conditions

Table 6-6 presents the optimization results for the two phasing scenarios. It clearly shows that phasing B yields better performance in terms of the corridor throughput,

diverted volume, and the freeway delay. Due to less diverted volume from the freeway, the arterial delay is less for phasing A. Overall, phasing B is a better phasing scheme for it better uses the corridor capacity and moves more traffic through the corridor. The same phasing (B) is used for all the other case studies and model validation.

Table 6-6. The Effect of Phasing on Corridor Traffic Conditions

	Phasing A	Phasing B
Optimal Cycle Length (Seconds)	120	120
Arterial Delay (Vehicle.Seconds)	307786	475170
Freeway Delay (Vehicle.Seconds)	121620	45193
Corridor Throughput (Vehicles)	987	1336
Total Diversion Rates (Vehicles/Hour)	992	1373

6.2.3 Case Study for the Corridor with Different Demand Level and Incident Conditions

To study the performance of the COROPT program under different traffic demand conditions, a corridor with 8 intersections on the arterial (as shown in Figure 6-1) was optimized using the COROPT program, for three different demand levels on the arterial.

Case 1: The demand level is light

Normal rates: 72 vehicles/hour for side street left-turns; 180 vehicles/hour for side street through and right-turns; 108 vehicles/hour for arterial left-turns; 360 vehicles/hour for arterial through and right-turns combined

Case 2: The demand level is moderate

Normal rates: 108 vehicles/hour for side street left-turns; 288 vehicles/hour for side street through and right-turns combined; 144 vehicles/hour for arterial left-turns; 720 vehicles/hour for arterial through and right-turns

Case 3: The demand level is high

Normal rates: 144 vehicles/hour for side street left-turns; 360 vehicles/hour for side street through and right-turns; 180 vehicles/hour for arterial left-turns; 900 vehicles/hour for arterial through and right-turns combined

The other data is the same as those in case of the eight-intersection corridor in section 6.2.1.

The COROPT program optimization results for those three cases are presented in Table 6-7.

Table 6-7. Summary of Optimal Results for the Corridor with Different Demand Level on the Arterial

	Light Demand	Moderate Demand	High Demand
Optimal Cycle Length (Seconds)	100	100	100
Arterial Delay/Intersection (Vehicle.Seconds)	26968	75466	102771
Freeway Delay (Vehicle.Seconds)	166	21745	63405
Corridor Throughput (Vehicles)	1394	1342	1212
Total Diversion Rates (Vehicles/Hour)	1597	1491	1281

It can be seen that as the demand level on the arterial increases, the delays on both the freeway and the arterial increase. In addition, the corridor throughput decreases as demand increases. The reason for that is that the high demand level on surface streets limits the amount of traffic that can be diverted from the freeway, the total diversion volume and the corridor throughput is thus reduced.

6.2.4 The Effect of Cycle Length on Corridor Traffic Conditions

The COROPT program was used to optimize the same corridor with eight intersections as in 6.2.1. Cycle lengths ranging from 60 seconds to 180 seconds, at the interval 10 seconds, are fixed for each optimization run, aiming to investigate the effect of the cycle length on corridor traffic conditions. Two demand levels were considered.

1. Light to moderate demand level

Normal rates: 90 vehicles/hour for side street left-turns; 230 vehicles/hour for side street through and right-turns; 130 vehicles/hour for arterial left-turns; 330 vehicles/hour for arterial through and right-turns combined

2. Moderate to high demand level

Normal rates: 120 vehicles/hour for side street left-turns; 540 vehicles/hour for side street through and right-turns; 160 vehicles/hour for arterial left-turns; 810 vehicles/hour for arterial through and right-turns combined

The results are shown in Table 6-8, and Figures 6-7, and 6-8 as well.

Table 6-8. The effect of the Cycle Length on Corridor Traffic Conditions

Cycle Length	Light to Moderate Demand		Moderate to High Demand	
	Total Delay (Vehicle.seconds)	Throughput (Vehicles)	Total Delay (Vehicle.seconds)	Throughput (Vehicles)
60	220181	1198	1342167	435
70	234530	1169	1247940	925
80	228928	1212	1217175	1013
90	210826	1168	1254125	1058
100	221048	1219	1218023	1098
110	241034	1220	1217209	1103
120	265520	1242	1278850	1147
130	292245	1236	1285874	1163
140	315711	1195	1295579	1143
150	363737	1096	1340017	869
160	395252	1023	1337688	679
170	442040	892	1339441	885
180	483557	802	1374653	514

It can be clearly seen that small cycle lengths lead to high delay, this is especial true when the demand level is relatively high as shown in Figure 6-6. In addition, small cycle lengths are not good either for increasing the throughput of the corridor. As the cycle length increases, a tendency of drop in the total delay and rise in the throughput is shown in Figure 6-6 and 6-7. However, as the cycle length keeps increasing, the delay will re-increase, and the throughput will decrease considerably after some point.

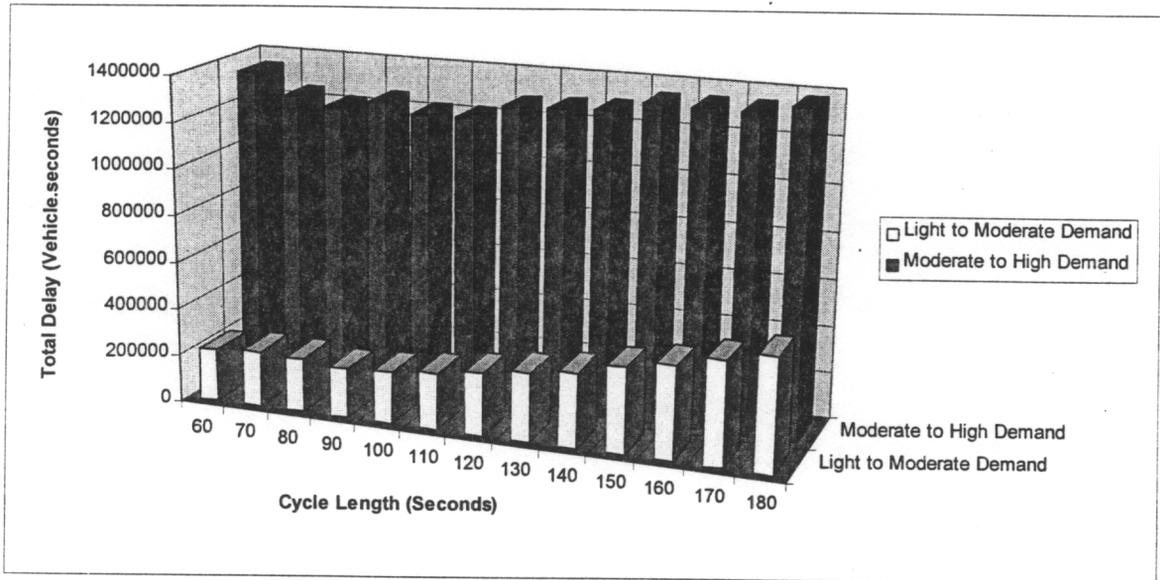


Figure 6-7. The Effect of the Cycle Length on the Total Delay

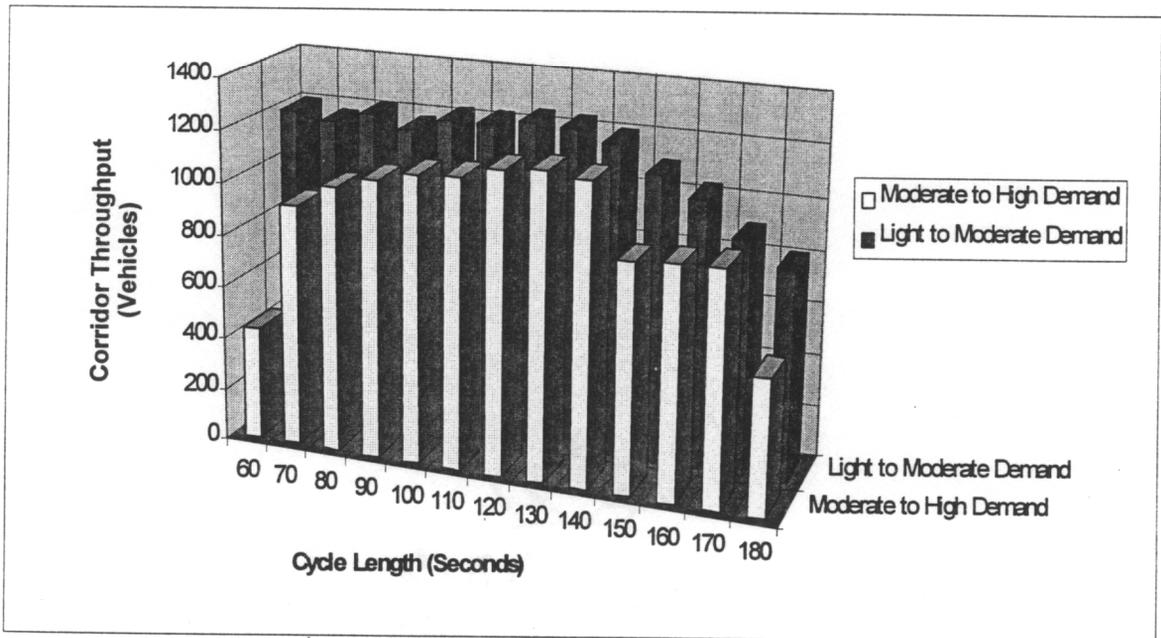


Figure 6-8. The Effect of the Cycle Length on the Corridor Throughput

Figures 6-7 and 6-8 also show that the maximum throughput and the minimum total delay are not achieved at the same cycle length. For the cases illustrated above, the minimum delay is achieved at cycle length about 80 seconds, while a cycle length of about 120 seconds is ideal if maintain the maximum throughput of the corridor is the primary goal.

6.3 MODEL VALIDATION

6.3.1 Test Example

The same hypothetical corridor as shown in Figure 6-1 was used for model execution and validation. The corridor consists of one freeway and a parallel arterial. Four off-ramps upstream of the incident location and four on-ramps downstream of the incident location are considered for diversion purposes.

Three optimization periods with stable prevailing traffic and incident conditions were considered. The following is the summary of input data:

Period 1 (0-1200 seconds)

Normal Arrival rates:

Arterial: through+right turning: 900 vehicles/hour, left turning: 180 vehicles/hour

All side streets: through+right turning: 360 vehicles/hour, left turning: 180
vehicles/hour

on- or off-ramp metering rate for each ramp: 180 vehicles/hour

Freeway: 3600 vehicles/hour

Available capacity of freeway incident section: 1800 vehicles/hour

Period 2 (1200-2400 seconds)

In period 2, the arrival rate for arterial through and right turning movement is 720 vehicles/hour, arrival rate to freeway is 3430 vehicles/hour, and the available capacity for the freeway incident section is 1980 vehicles/hour.

Period 3 (2400-3600 seconds)

In this period, the available capacity of the freeway incident section is 2340 vehicles/hour, and the rest of the arrival rates remain the same as those for period 2.

The results of the COROPT optimization are listed in Tables 6-9 and 6-10 for diversion rates and timing plans, respectively. A cycle length of 120 seconds was found to be the optimal cycle length for each period.

It can be seen from Table 6-9 that the diversion volume is split between all the upstream off-ramps with the ramp closer to the incident location carrying a little more diversion volume. For diverting traffic back to the freeway after incident location, the closer the on-ramp is to the incident location, the more diversion volume it carries. This is reasonable because the fact that incident-free freeway links have higher capacity and speed encourages traffic to take the freeway. The diversion cannot be completed using only one on-ramp due to the limitation of ramp capacity and possible extra delay to other approaches at the intersection. From Table 6-10, a tendency of allocating more green time to links carrying more traffic is also evident.

Table 6-9. Summary of Optimization Results (Diversion Rates)

Optimization Period	Off-Ramp Rate (vehicles/hour)				On-Ramp Rate (vehicles/hour)			
	1	2	3	4	5	6	7	8
1	489	501	527	566	841	687	394	180
2	509	513	517	534	765	659	468	180
3	405	408	410	420	603	514	348	180

Table 6-10. Summary of Optimization Results (Green Time for Each Phase in Seconds)

Intersection No.	Period 1				Period 2				Period 3			
	Phase1	Phase2	Phase3	Phase4	Phase1	Phase2	Phase3	Phase4	Phase1	Phase2	Phase3	Phase4
1	35	30	25	30	37	32	26	25	36	29	29	26
2	32	38	21	29	34	39	23	24	34	34	27	25
3	30	46	17	27	31	46	19	24	32	39	24	24
4	28	54	13	25	28	53	16	23	30	44	22	24
5	18	55	19	28	20	55	21	24	23	48	24	25
6	21	49	21	29	22	50	23	25	24	44	25	26
7	25	38	26	31	25	43	26	26	27	38	28	27
8	28	31	31	30	29	32	31	28	29	32	31	28

6.3.2 Validation of Optimal Timing Plans

TRANSYT-7F was used to validate the signal timing optimization mechanism of the COROPT program. By superimposing diverted traffic from COROPT results into the normal arrivals to the arterial intersection, a TRANSYT-7F optimization run was performed to obtain the optimal timing plan, and the comparison of TRANSYT-7F results with the COROPT optimal timing results was made. Table 6-11 is the results comparison for optimization period 1, with a cycle length of 120 sec.

The table shows that the differences in green times from the two methods are very small.

6.3.3 Evaluating the Effectiveness of the Optimal Diversion Strategy Combined with Arterial Retiming

The diversion strategy obtained from COROPT was evaluated with multiple INTEGRATION simulation runs. Its operational performance of the corridor was evaluated and compared with other scenarios using INTEGRATION simulation. The following three cases are considered:

- Case 1: No diversion, no arterial signal retiming
- Case 2: Implementing user equilibrium-based routing and signal retiming provided by INTEGRATION
- Case 3: Implementing diversion and arterial signal retiming strategies developed by COROPT

By specifying time-varying routing path trees, INTEGRATION run was performed to simulate each case. Table 6-12 summarizes the results of INTEGRATION simulations for evaluating the effectiveness of different diversion and signal retiming strategies.

Table 6-11. Comparisons of TRANSYT-7F and COROPT Timing Plans (Green Times in Seconds)

Intersection No.	Phase 1			Phase 2			Phase 3			Phase 4		
	CO*	7F*	Dif*									
1	35	33	-2	30	32	2	25	23	-2	30	32	2
2	32	30	-2	38	40	2	21	19	-2	29	31	2
3	30	27	-3	46	49	3	17	15	-2	27	29	2
4	28	27	-1	54	56	2	13	11	-2	25	26	1
5	18	17	-1	55	56	1	19	17	-2	28	30	2
6	21	20	0	49	47	-2	21	20	-1	29	32	3
7	25	24	-1	38	38	0	26	26	0	31	32	1
8	28	27	-1	31	32	1	31	30	-1	30	31	1

7F*: Green Time from TRANSYT-7F Optimization

CO*: Green Time from COROPT Results

Dif*: Dif=(7F-CO)

Table 6-12. Simulation Results for Different Diversion and Signal Retiming Strategies

MOE	Period 1			Period 2			Period 3		
	ND*	UE*	CO*	ND	UE	CO	ND	UE	CO
Total Network Travel (Veh-km)	8150	9326	8826	13671	11616	14074	18632	17868	18663
Average Network Speed (km/h)	25	23	24	17	16	16	14	10	12
Average Network Stops (%)	52.9	48.6	52.4	57.5	55.0	56.2	59.4	50.1	57.7
Corridor Throughput (veh)	712	671	817	1252	850	1351	1631	1292	1764

ND: No diversion

UE: Using user equilibrium based routing trees and signal timing optimization provided by INTEGRATION

CO: Simulation run using the diversion paths and optimal signal timing from COROPT as Inputs

It can be seen from comparing the results that the diversion and arterial signal retiming strategy contributes to improvement of overall corridor performance. The total network travel and corridor throughput (defined as the total vehicles that exit the corridor on the direction of the freeway incident link) is increased by 2.7% and 9.4%, respectively. While the network travel increases, the average network speed and average network stops remain pretty much unchanged. It is interesting to notice that the user equilibrium-based routing and signal retiming methods implemented in INTEGRATION do not provide an advantage over the no-diversion case in terms of overall corridor performance. User equilibrium-base assignment and retiming approach implemented in INTEGRATION sometimes gives better performance in network stops percentage. This is very good if the demand level is light or moderate. However, having the fewest number of stops is not a feature to which the highest priority should be given for the corridor under sever incidents. Therefore, the COROPT program yields the best corridor traffic condition for this case.

7. CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE RESEARCH

This dissertation presents an approach of developing an integrated optimal control model for a corridor with incidents on the freeway. A nonlinear programming model was formulated and a solution procedure was developed. A corridor traffic control computer program (COROPT) was developed and coded in C. The COROPT optimization results were compared and validated using TRANSYT-7F and INTEGRATION packages.

Conclusions on the model and recommendations for future enhancement are presented below.

7.1 CONCLUSIONS

In an innovative way, the developed optimization model incorporates corridor control measures including diversion and signal retiming in an integrated manner. The objective function of the optimization model considers the interaction among the corridor components, and clearly reflects the primary goals of corridor traffic control under freeway incident conditions: to divert as much traffic away from the freeway as possible, not to over-congest the arterial and surface streets; and properly re-set the signal timing plans at all intersections to accommodate the changed traffic demand levels and patterns.

By limiting queue lengths in constraints and penalizing long queues in the objective function, the optimization model has a special feature of congestion prevention that is crucial for traffic control under high demand level. The model is well-suited for near-saturated or saturated traffic conditions, while most of the current traffic models and optimization programs have not shared much success.

The gradient projection method is used to solve diversion and signal retiming control measures simultaneously. By using a specifically developed simple and realistic traffic flow model and employing a sequential optimization approach, the computer program COROPT can obtain optimal traffic control strategies quickly and effectively. An optimal solution for a corridor of eight to sixteen intersections can normally be obtained using COROPT program in a short period of several minutes. This opens the door for real time implementation.

The model employs an off-line approach. It does not require any expensive traffic detecting equipment nor any other expensive hardware installations. Nevertheless, the fast performance of the model and program is suitable for real time applications provided on-line data is available.

The COROPT program has the flexibility to deal with various corridor configurations, different size of the corridor system, and different timing phasing. It addresses the time-varying factor of traffic flow, and can handle changing traffic and incident conditions over the time.

In addition to its flexibility and fast performance, the COROPT program also yields good and realistic results. Its robustness to handle different network and traffic conditions has also been shown.

The testing and validation using available traffic simulation and signal optimization packages showed that the COROPT strategy provides advantages over the existing traffic control models and methodologies and improves the overall performance of the corridor, especially, the corridor throughput.

The model and program developed can be used to assist incident management personnel to develop better and proper incident response plans, especially diversion

plans and arterial signal retiming plans.

Though the model and program were developed for severe incident conditions, they are also suitable for non-incident conditions where either the arterial or the freeway, or both have a high demand level, and re-assignment of traffic (diversion) and thus retiming are beneficial and necessary.

7.2 RECOMMENDATIONS FOR FUTURE RESEARCH

Future research will concentrate on model improvement, including enhancing the flexibility and versatility of the model.

Currently the model can only deal with the standard cases of having one or two arterials parallel to the freeway. In real world scenario, corridors or transportation networks can be much more complicated. To apply the model concept to any corridor condition, a network flow based traffic flow model has to be developed.

Although the program can deal with different basic phasing schemes, it is unable to handle various special traffic control measures such as stop sign or yield sign that may present at some intersections. It can not deal with special signal phasing such as permitted left-turns. Signal control at an intersection can be very complicated and much needs to be done in the model and the program to make them versatile.

The testing and validation of the model were performed using macroscopic optimization and simulation models. To acquire more reliable and convincing results, more model validation using microscopic simulation is also needed.

Finally, the implementation issues need to be address in the future research.

REFERENCES

- A Policy on Geometric Design of Highways and Streets*, American Association of State Highway and Transportation Officials (AASHTO), 1994
- Bang, K-L and Nilsson, L-E, "Optimal Control of Isolated Traffic Signals," *Control in Transportation Systems*, Proceedings of the IFAC/IFIP/IFORS Third International Symposium, Columbus, Ohio, August 9-13, 1976
- Bang, K.L. and Nilson, L.E., "Optimal Control of Isolated Traffic Signals," *Proceeding of Australian Road Research Board*, Vol. 8, 1976
- Bazaraa, M.S., Jarvis, J.J., and Sherali, H.D., *Linear Programming and Network Flows*, New York: Wiley, 1990
- Bazaraa, M.S., Sherali, H.D., and Shetty, C.M., *Nonlinear Programming: Theory and Algorithms*, New York: Wiley, 1993
- Boillo, F., Blosserville, J.M., Lesort, J.B., Motyka, V., Papageorgiou, M., and Sellam, S., "Optimal Signal Control of Urban Traffic Networks," *Proceedings of the Sixth International Conference on Road Traffic Monitoring and Control*, Venue, April 28-30, 1992
- Chang, E.C.P., Cohen, S.L., Liu, C., Chaudhary, N.A., and Messer, C.J., "MAXBAND-86: Program for optimizing left-Turn Phase Sequence in Multiarterial Closed Networks," *Transportation Research Record* 1181, Transportation Research Board, Washington D.C., 1988
- Chang, G., Wu, J., and Cohen, S. L., "Integrated Real-Time Ramp Metering Control for Non-Recurrent Congestion: Modeling and Strategy Development," *Proceedings of IVHS America Annual Meeting*, Atlanta, 1994
- Chang, G., Wu, J., and Lieu, H., "Real-Time Incident-responsive Corridor Control: A Successive Linear Programming Approach," *Proceedings of IVHS America Annual Meeting*, Atlanta, 1994
- Chaudhary, N. A. and Messer, C. J., "PASSER IV: A Program for Optimizing Signal Timing in Grid Networks," *Transportation Research Record* 1287, Transportation Research Board, Washington D.C., 1993
- Clowes, D.J., "Real-Time Wide Area Traffic Control - The User's Viewpoint," *Traffic Engineering and Control*, April, 1982

Codelli, D. K., Niedringhaus, W.P., Wang, P.T.R., and Young, B.S., *The Traffic and Highway Object for REsearch Analysis, and Understanding (THOREAU) IVHS Model*, Volume I, MITRE Corporation, McLean, Virginia, 1992

Codelli, D. K., Niedringhaus, W.P., Wang, P.T.R., and Young, B.S., *User's Manual for Traffic and Highway Object for REsearch Analysis, and Understanding (THOREAU)*, Volume II, Revision 1, MITRE Corporation, McLean, Virginia, 1993

Cohen, S.L., "Concurrent Use of MAXBAND and TRANSYT Signal Timing Programs for Arterial Signal Optimization," *Transportation Research Record*, 906, 1983

Cremer, M. and Schoof, S., "On Control Strategies for Urban Traffic Corridors," *Control, Computers, Communications in Transportation*, IFAC/IFIP/IFORS, 1989

Daniel, B.F., Chaudhary, N.A., and Garza, R.U., *A report on the Users Manual for the Microcomputer Version of PASSER III-88*, Texas Transportation Institute, September, 1988

Drew, D.R., *Traffic Flow Theory and Control*, McGRAW-Hill, 1969

Elahi, S. M., Radwan, A. E., and Goul, K. M., "Knowledge-Based System for Adaptive Traffic Signal Control," *Transportation Research Record*, No. 1324, 1991

FHWA, *Traffic Control Systems Handbook*, Report FHWA-IP-85-11, U.S. Department of Transportation, 1985

Gallivan, S. and Heydecker, B., "A Continuous-Time formulation for Traffic-Responsive Signal Control," *Proceedings of the eleventh International Symposium on Transportation and Traffic Theory*, Yokohama, Japan, July 18-20, 1990

Gallivan, S. and Heydecker, B., "Optimizing the Control Performance of Traffic Signals at Single Junction," *Transportation Research-B*, Vol.22B, No.5, pp 357-370, 1988

Gartener, N. H., "OPAC: A Demand-Responsive Strategy for Traffic Signal Control," *Transportation Research Record*, 906, 1983

Gartner, N.H., Assmann, S.F., Lasaga, F., and Hou, D.L., "A Multi-Band Approach to Arterial Traffic Signal Optimization," *Transportation Research-B*, Vol. 25B, No.1, pp55-74, 1991

Gartner, N.H. and Reiss, R.A., "Congestion Control in Freeway Corridors: The IMIS System," *Flow Control of Congested Networks*, Springer-Verlag, 1987

Grafton, R.B. and Newell, G.F., "Optimal Policies for the control of an Undersaturated Intersection," *Proceedings of the 3rd International Symposium on Theory of Traffic Flow*, New York, 1967

Hadi, M. A. and Wallace, C. E., "Hybrid Generic algorithm to Optimize Signal Phasing and Timing," *Transportation Research Record 1287*, Transportation Research Board, Washington D.C., 1993

Henry, J.J., and Farges, J.L., "PRODYN," *Control, Computers, Communications in Transportation*, IFAC/IFIP/IFORS, Paris, France, 1989

Highway Capacity Manual (special report 209), U.S. Department of Transportation, Federal Highway Administration, Washington D.C., 1991

Hobeika, A.G., *Transportation and Land Use*, Class Notes, Virginia Tech, 1992

Hunt, P.B. and Holland, T.R., "The Effect of an Incident in a SCOOT System - in Simulation and On-Street," *Traffic Engineering and Control*, February, 1985

Hunt, P.B., Robertson, D.I., and Bretherton, R.D., "The SCOOT On-line Traffic Signal Optimization Technique," *Traffic Engineering and Control*, April, 1982

Improta G., and G.E. Cantarella, "Control System Design for an Individual Signalized Junction," *Transportation Research-B*, Vol.18B, No.2, pp147-167, 1984

Improta, G. and Sforza, A., "Optimal Offsets for Traffic Signal Systems in Urban Network," *Transportation Research-B*, Vol. 16B, No.2, pp.143-161, 1982

Jezequel, R., "Traffic Control at an Intersection by Means of a Minicomputer," *Control in Transportation Systems*, Proceedings of the IFAC/IFIP/IFORS Third International Symposium, Columbus, Ohio, August 9-13, 1976

Kessaci, A., Farges, J.L., and Henry, J.J., "On Line Estimation of Turning Movements and Saturation Flows in PRODYN," *Control, Computers, Communications in Transportation*, IFAC/IFIP/IFORS, Paris, France, 1990

Larry H., Pitu, B. M., and Dennis S., "Hierarchical Framework for Real-Time Traffic Control," *Transportation Research Record 1360*, Transportation Research Board, Washington D.C., 1992

Larry H., Pitu, B. M., and Dennis S., "Hierarchical Framework for Real-Time Traffic Control," *Transportation Research Record 1360*, Transportation Research Board, Washington D.C., 1992

Lindley, J. A., "Urban Freeway Congestion: Quantification of the Problem and Effectiveness of Potential Solutions," *ITE Journal*, VOL. 57, No1, pp27-32

Linkenheld, J.S., Benekohal, R.F., and Garrett, J.H., "Knowledge-Based System for Design of Signalized Intersections," *Journal of Transportation Engineering*, Vol. 118, No.2, Marh-April, 1992

Luk, J.Y.K., "Two Traffic-Responsive Area Traffic Control Methods: SCAT and SCOOT," *Traffic Engineering and Control*, February, 1985

MAcGowan, C.J. and Lum, H.S., "SIGOP or TRANSYT," *Traffic Engineering*, April, 1975

Manzur E. A., Essam R., and K. Michael G., "Knowledge-Based System for Adaptive Traffic Signal Control," *Transportation Research Record*, No. 1324, 1991

Mastbrook, S., *A Framework For Predicting Urban Freeway Incident Delays In Real Time*, Master's thesis, Blacksburg, Virginia, 1996

Mauro, V. and Taranto, D. D., "UTOPIA," *Control, Computers, Communications in Transportation*, IFAC/IFIP/IFORS, Paris, France, 1989

Miller, A.J., "A Computer Control System for Traffic Networks," *Proceedings of the 2nd International Symposium on Theory of Road Traffic Flow*, Paris, 1965

Miller, A.J., "A Computer Control System for Traffic Networks," *Proceedings Of The Second International Symposium of The Theory of Traffic Flow*, London, 1963

Nathan H. G., "OPAC: A Demand-Responsive Strategy for Traffic Signal Control," *Transportation Research Record*, No. 906, 1983

Nathan H. G., Chronis S., and Phillip J. T., "Development of Advanced Traffic Signal Control Strategies for IVHS: A Multi-Level Design," *Transportation Research Board 74th Annual Meeting*, , Washington D.C., January 22-28, 1995

Nathan H. G., Philip J. T., and Christina M. A., "Evaluation of optimized Policies for Adaptive Control Strategy," *Transportation Research Record*, No. 1324, 1991

OCED Road Research Group, *Traffic Control in Saturated Conditions*, 1981

Özbay K., Hobeika A.,and Zhang Y., "A Study for Predicting Incident Clearance Duration in Northern Virginia," Paper submitted to 1997 TRB meeting, Blacksburg, VA, 1996

Paolo D., and Pitu M., "REALBAND, An Approach for Real-Time Coordination of Traffic Flows on Networks," *Transportation Research Board 74th Annual Meeting*, Washington D.C., January 22-28, 1995

Papageorgiou, M., "Modeling and Real-Time Control of Traffic Flow on the Boulevard Peripherique in Paris," *Control, Computers, Communications in Transportation*, IFAC/IFIP/IFORS, 1989

Peterson, A, Bergh, T. and Steen, K., "LHOVRA - A new Traffic Signal Control Strategy for Isolated Junctions," *Second International Conference on Road Traffic Control*, 15-18 April, 1986

Peterson, A, Bergh, T. and Steen, K., "LHOVRA - A new Traffic Signal Control Strategy for Isolated Junctions," *Traffic Engineering and Control*, April, 1982

Pooran, F. J. and Lieu, H. C., "Evaluation of System Operating Strategies for Ramp Metering and Traffic Signal Coordination," *Proceedings of IVHS American annual Meeting*, Atlanta, 1994

Reljic, S. S., and M. Stojic, "Another Way of Optimal signal Plan Calculation at an Isolated Intersection," *Transportation and traffic theory*, Proceedings of the eleventh International Symposium on Transportation and Traffic Theory, Yokohama, Japan, July 18-20, 1990

Research and Management for Advanced Traffic Signal Systems, *Transportation Research Circular*, No. 431, July 1994

Robertson, D.I. and Bretherton, R.D., "Optimum Control of an Intersection for any Known Sequence of Vehicle Arrivals," *Traffic Control and Transportation Systems*, Proceedings of the 2nd IFAC/IFIP/IFORS Symposium, Monte Carlo, 16-21, September, 1974

Rogness, R. O. and Messer, C. J. "Heuristic Programming Approach to Arterial Signal Timing," *Transportation Research Record*, 906, 1983

Ross, D.W., *A Computer Control Scheme for Critical Intersection Control in an Urban Network*, Stanford Research Institute, 1969.

Sen, S., "Coordinated Optimization of Phases (COP)," July, 1993

Sen, S. and Head, K. L., "Controlled Optimization of Phases (COP) at an intersection," November, 1994

Subramaniam, S., Zhang, Y., Sundar, S., and Masbrook, S., *Progress Report of Wide-Area Incident Management Project*, Virginia Tech Center for Transportation Research, July, 1994

TRANSYT 7F User's Manual, University of Florida Transportation Research Center, Gainesville, Florida, June, 1984

Tsay, H.S., "A Microcomputer-Based On-line Traffic Control System," *ITE Journal*, December, 1989

Van Aerde, M., Voss, J., Ugge, A., and Case, E. R., "Managing Traffic Congestion in Combined Freeway and Traffic Signal Networks," *ITE Journal*, February 1989

Van Aerde and M., Yagar, S., "Dynamic Integrated Freeway/Traffic Signal Networks: Problems and Proposed Solutions," *Transportation Research - A*, Vol. 22A, No. 6, 1988

Vincent, R.A. and Young, C.P., "Self-optimizing Traffic Signal Control Using Microprocessors - The TRRL MOVA Strategy for Isolated Intersections," *Second International Conference on Road Traffic Control*, 15-18 April, 1986

Vincent, R.A. and Young, C.P., "Self-optimizing Traffic Signal Control Using Microprocessors - The TRRL MOVA Strategy for Isolated Intersections," *Traffic Engineering and Control*, April, 1982

Walmsley, J., "The practical Implementation of SCOOT Traffic Control System," *Traffic Engineering and Control*, April, 1982

Webster, F.V., *Traffic Signal Setting*, Road Research Technical Paper 39, H.M.S.O. , London, 1958

Weinberg, M.F., "Digital-Computer-Controlled Traffic Signal System for a Small City," *NCHRP 29*, 1966

Yagar, S., "Integrated Networks, Traffic Monitoring and Control Systems," *Proceedings of Engineering Foundation Conference*, Henniker, New Hampshire, 1983

Zijverden, J.D., Kwakernaak, H., "A New Approach to Traffic-Actuated Computer Control of Intersections," *Proceedings of the 4th International Symposium on Theory of Traffic Flow*, Bonn, Germany, 1969.

VITA

Yunlong Zhang was born on August 3, 1964 in Kunshan, Jiangsu Province, China. He received his B.S. in civil engineering in 1984, and M.S. in civil engineering (transportation) in 1987 from Southeast University, Nanjing, China. From 1987 to 1992, he taught in the Department of Transportation Engineering at Southeast University. He also served as the vice director of the Highway Engineering Division in the Department from 1990 to 1992. He worked as a research Associate at Virginia Tech Center for Transportation Research from May through October of 1996.

His professional interests include ITS, incident management, transportation modeling and simulation, traffic signal optimization, and operations research and computer applications in transportation. He has been extensively involved in the wide-area incident management expert-GIS system project sponsored by FHWA/VDOT.

Upon his graduation he will join the ITS group of Kaman Sciences Corporation in Colorado Springs, Colorado.

A handwritten signature in black ink, reading "Yunlong Zhang". The signature is written in a cursive style with a large, sweeping flourish at the end of the name.