Development and Applications of Multi-Objectives Signal Control Strategy during Oversaturated Conditions

By

Zaeinulabddin M Adam

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

Montasir M. Abbas, Chair
Antoine G. Hobeika
Hesham A. Rakha
Raghu Pasupathy

August 10, 2012
Blacksburg, VA

Keywords: Signal system, Oversaturation control strategy, Critical route, Pareto front, Queue management, Wavelet analysis, Multi-objective control

Development and Applications of Multi-Objectives Signal Control Strategy during Oversaturated Conditions

Zaeinulabddin M Adam

ABSTRACT

Managing traffic during oversaturated conditions is a current challenge for practitioners due to the lack of adequate tools that can handle such situations. Unlike under-saturated conditions, operation of traffic signal systems during congestion requires careful consideration and analysis of the underlying causes of the congestion before developing mitigation strategies. The objectives of this research are to provide a practical guidance for practitioners to identify oversaturated scenarios and to develop a multi-objective methodology for selecting and evaluating mitigation strategy/ or combinations of strategies based on a guiding principles. The research focused on traffic control strategies that can be implemented by traffic signal systems. The research did not considered strategies that deals with demand reduction or seek to influence departure time choice, or route choice. The proposed timing methodology starts by detecting network's critical routes as a necessary step to identify the traffic patterns and potential problematic scenarios. A wide array of control strategies are defined and categorized to address oversaturation problematic scenarios. A timing procedure was then developed using the principles of oversaturation timing in cycle selection, split allocation, offset design, demand overflow, and queue allocation in non-critical links. Three regimes of operation were defined and considered in oversaturation timing: (1) loading, (2) processing, and (3) recovery. The research also provides a closed-form formula for switching control plans during the oversaturation regimes. The selection of optimal control plan is formulated as linear integer programming problem. Microscopic simulation results of two arterial test cases revealed that traffic control strategies developed using the proposed framework led to tangible performance improvements when compared to signal control strategies designed for operations in under-saturated conditions. The generated control plans successfully manage to allocate queues in network links.

DEDICATION

I dedicate this work to the soul of my mother, Hashimia Zein-elabddin Omer

And

To my father Dr. Mohamed Ahmed Khadam

ACKNOWLEDGEMENTS

I take this opportunity to express my gratitude to those who have been instrumental in the successful completion of this dissertation through their encouragement, guidance, and generous support. I would like to show my greatest appreciation to Dr. *Montasir Abbas* for his guidance and persistent inspiration throughout my Ph.D. study at Virginia Tech. I can't say thank you enough for his tremendous support and help. Without his encouragement and guidance this research would not have materialized.

I am also indebted to the assistance of the other committee members: Dr. *Raghu Pasupathy*, Dr. *Hesham Rakha*, and Dr. *Antoine G. Hobeika*. Each of them provided me with valuable suggestions and tremendous support that expedited my research. I immensely benefited from the classes that I studied with them and from the insightful guidance, advice, and feedback they provided on my doctoral research. In addition, they have always provided me with the support I needed. I would like also to extend my thanks to *Mr. Eric Nelson, P.E.*, of Advanced Traffic Solutions, LLC for his help in data collection and software services support.

Very special thanks go to my colleagues at Signal Control and Operation Research Lab with whom I forged very special bonds. I am greatly thankful to *Asmita Ghrat* and *Milos Mladenovic* for their immense help and support. I thank *Bryan Higgs, Linsen Chong, Yatish Kasaraneni, Dr. Pengfei Li, Lakshmi Rajshekar, Sudeeksha, Pooja Dwivedi, and* Yan *Li* for their help, generosity and friendship.

My deepest appreciation is to all my family members, especially my brother *Dr. Ibrahim* and my friends for their unlimited encouragement and support. I also owe particular gratitude to the Via Department of Civil and Environmental Engineering and to *Charles E. Via Endowment Program* for their generous support of my studies at Virginia Tech. Finally, yet importantly, I would like to express my thanks to Mrs. *Lindy Cranwell* and Mrs. *Merry Gayle-Moller* who have provided me with great and valuable help and genuine hospitality that I am truly grateful for them.

ACKNOWLEDGMENT OF SPONSORSHIP

This work was sponsored by American Association of State Highway and Transportation Officials (AASHTO), in cooperation with the Federal Highway Administration (FHWA), and was conducted in the *National Cooperative Highway Research Program* **(NCHRP),** which is administered by the Transportation Research Board of the National Academies. This research was performed under NCHRP Project 03-90 by Kimley-Horn & Associates, with Virginia Tech University, University of Minnesota, and Dowling & Associates. Kimley-Horn & Associates was the primary contractor and Dr. *Douglas Gettman* was the Project Director and Principal Investigator. *Dr. Montasir Abbas*, Associate Professor of Civil Engineering at Virginia Tech was co-Principal Investigator, *Dr. Henry Liu*, Associate Professor of Civil Engineering at the University of Minnesota and co-Principal Investigator.

TABLE OF CONTENTS

1 INTRODUCTION 1
1.1 Research Background and Motivation 1
1.2 Research team 2
1.3 NCHRP 2
1.4 Research Objectives 3
1.5 Dissertation Layout 3
2 Literature Review 5
2.1 Traditional Signal Control Objectives 5
2.2 Congestion Diagnosis & Estimation 6
2.2.1 Definitions of Congestion and Level of Service 7
2.2.2 Definitions based on Queue Length 9
2.3 Measures of oversaturation based on Delay/Stops/Speed/Travel Time 10
2.4 Oversaturation Control Strategies 13
2.4.1 Isolated Intersections 13
2.4.2 Models for Queue Interactions between Closely-Spaced Intersections 17
2.4.3 Diamond Interchanges 17
2.4.4 Reduced Cycle Times 19
2.4.5 Metering/Gating 19
2.4.6 Recovery from Congestion 21
2.4.7 Dynamic Optimization Algorithms 22
2.5 TRANSYT-7F 23
TRANSYT-7F under congested conditions 25
2.5.1 SYNCHRO 25
2.6 UTC systems 26
2.6.1 UTC Operational Strategies 28
2.7 Adaptive Control Algorithms 28
2.7.1 SCOOT 29
Cycle optimizer 30
Offset optimizer 30
2.7.2 SCATS 33
2.8 Multi-Objective Control 35
2.9 Origin-Destination Matrix Estimation Methods 36
2.10 Wavelet Analysis 38
3 Research Methodology and Organization 40
3.1 Introduction 40

3.2 Wavelet Filtering 40
3.3 Simulation 40
3.4 Multi-objective evaluation 41
3.5 Dissertation Organization 41
3.5.1 Task1: Literature review 41
3.5.2 Task 2: Critical routes determination and scenarios development 42
3.5.3 Task 3: Development of oversaturation control strategies 42
3.5.4 Task 4: Data collection and Simulation modeling 42
4 Operational of Signal System in Oversaturated conditions 43
4.1 Introduction 43
4.2 Congestion definitions 43
4.2.1 Problematic Scenarios 44
4.3 High-Level System Dynamics 46
4.4 A Catalog of Mitigation Strategies 47
4.4.1 Local strategies 47
4.4.2 Arterial Strategies 49
4.4.3 Network level strategies 49
4.5 Summary of control strategies 51
4.6 Proposed timing framework during oversaturated conditions 53
5 Critical Routes Estimation Using Wavelet Filtering 54
5.1 Introduction 54
5.2 Background 54
5.2.1 WAVELET'S APPROXIMATIONS AND DETAILS 55
5.2.2 WAVELET COEFFICIENTS 56
5.2.3 DAUBECHIES WAVELET 56
5.3 METHODOLOGY (WAVELET ANALYSIS AND TIME PLAN) 57
5.3.1 Proposed Wavelet Filtering Framework 57
5.4 Critical Route Volume Determination 59
5.4.1 Experiment and Analysis 60
5.4.2 Wavelet Analysis Results 61
5.4.3 Critical routes volume profile 66
6 Timing Framework for Oversaturated Arterial Network 68
6.1 Introduction 68
6.2 Cycle Length Determination 69
6.3 Green Splits Design 72
6.4 Oversaturation offset design 73

6.4.1 Spillback-avert offset 73
6.4.2 Starvation-avert offset 74
6.5 Splits-offsets calculation 75
6.6 Summary and Conclusion 76
7 Multi-objective Timing Tool for Oversaturated Arterial Network 77
7.1 Introduction 77
7.2 OTS features 77
7.2.1 Timing based on movements' volumes profiles 77
7.2.2 Congestion matrix 79
7.2.3 Multi-objective control 79
7.3 Delay Calculation 80
7.4 Throughput calculations 81
7.5 Control problem formulation 82
7.5.1 Parameters Definitions 83
7.5.2 Objective function 84
7.5.3 Problem formulation 84
7.5.4 Control variables: 84
7.5.5 Queue management control 85
7.6 Optimal control switching points problem 85
7.6.1 Optimal switching points Problem formulation 86
Problem Inputs 87
Control variable 87
Objective functions 87
7.6.2 Optimization solver 89
7.7 OTS optimizer improvement: Initial solution 90
7.8 Illustration example: Initial solution impact on optimal solution 93
7.9 Model verification and validation 95
7.10 OTS Validation: Control delays calculations 96
8 Case Study: Arterial network (The Reston Parkway, Herndon, VA) 99
8.1 Introduction 99
8.2 Reston Parkway Network 99
8.2.1 Calibration data 100
8.2.2 Traffic pattern in Reston parkway 101
8.2.3 Building Volume scenarios 102
8.2.4 Timing Plans developments 108
8.3 Simulation Result and Pareto front Analysis 110

8.4 Statistical analysis 114
8.6 Lessons learnt and guidance from Reston Parkway Case Study 116
8.7 Summary and Conclusion 116
9 Case Study: Post Oak Network, Houston, TX 118
9.1 Introduction 118
9.2 Study network: Post Oak Network 118
9.3 Step 1: Development of Critical Route Scenarios 119
9.4 Step 2: Control Strategy Development 121
9.5 Step 3: Control Plans Development 122
9.6 Simulation Experiment 125
9.7 Performance measures evaluation 125
9.7.1 Network overall performance measures 126
9.7.2 Queues analysis 126
9.7.3 Network reserve capacity percentage 129
9.7.4 Throughput analysis 132
Definition: 132
9.7.5 Comparison of network loads and Throughput flows 137
9.7.6 Travel time analysis of results 138
Travel time profiles analysis 140
9.8 Summary and Conclusion 147
10 Conclusions and Future Directions 150
10.1 Introduction 150
10.2 Oversaturation Definitions 150
10.3 Critical Routes Identification 150
10.4 Oversaturation Control Strategies Development 151
10.5 Test Case 1 151
10.6 Test Case 2 152
10.7 Research Conclusion Summary 152
10.8 Recommendations for Future Research 154
References 156
Appendix 162

TABLE OF FIGURES:

Figure 1: Vehicle Miles Traveled Versus Lane Mileage (Source: FHWA; "Finding Out What America Thinks", Publication No. FHWA-OP-01-018) 1
Figure 2: A Hierarchy of Traffic Conditions (Source: Draft report from FHWA signal timing under saturated conditions project) 9
Figure 3: Idealized cumulative arrivals and departures for under- and over-saturated conditions (Source: Dion, F., *et al.*, 2004) 11
Figure 4: Random and Overflow Delay Models Compared (Source: Roess et al., 2004)) 12
Figure 5: Approach spillback (de facto red) 44
Figure 6: Storage bay spillback 45
Figure 7: Storage blocking 45
Figure 8: intersection starvation 45
Figure 9: Cross-Blocking effects 46
Figure 10: Loading, Oversaturation, and Recovery Phases of Operation 46
Figure 11: Control strategies operational objectives 51
Figure 12: Signal filtering process (Source: MATLAB Wavelet Toolbox™ 4 User's Guide) 56
Figure 13: Wavelet Multi level decompositions (Source: MATLAB Wavelet Toolbox™ 4 User's Guide) 56
Figure 14: Daubechies' compactly support family scaling function wavelet set (Source: MATLAB Wavelet Toolbox 4 User's Guide 57
Figure 15: The critical routes identification framework 57
Figure 16: Critical Routes Volume Estimation Framework 59
Figure 17: Critical movements on Reston Parkway network in Northern Virginia 61
Figure 18: Critical movements count profiles on Reston Parkway network 61
Figure 19: Critical Movements' Detail 2 Coefficients 62
Figure 20: Critical Movements' Detail 3 Coefficients 63
Figure 21: K-means clustering for movements' detail 3 coefficients 64
Figure 22: K-means clustering for movements' approximation coefficients 64
Figure 23: Correlated critical movements and the corresponding routes 65
Figure 24 : Route (A) Volume during P.M. Peak Period 66
Figure 25: Route (B) Volume during P.M. Peak Period 66
Figure 26: Route (C) Volume during P.M. Peak Period 66
Figure 27: Timing Framework for oversaturated conditions 68
Figure 28: Minimum delay cycle length as function of intersection degree of saturation 69
Figure 29: Shockwave at Signalized Intersection 70
Figure 30: Maximum cycle length that prevents spillback based on approaches g/C ratio and link length 71
Figure 31: Maximum cycle length that prevents starvation based on R/C approaches R/C ratio and flow rate 72
Figure 32: (a) link spillback and upstream intersection blockage (b) Spillback Avert Offset 73
Figure 33: (a) starvation at downstream intersection (b) Starvation-avert offset 74
Figure 34: Offset Values Feasible Region 74
Figure 35: Two conflicting movements' volume profiles and their corresponding v/c ratios 78
Figure 36: Demand profile during oversaturation condition 78
Figure 37: Timing tool congestion matrix, critical movement (1) and non-critical movements (0) 79
Figure 38: Intersection saturation rate 82
Figure 39: Oversaturation timing Framework 86
Figure 40: Control Plan: optimal time plans, control objectives, and optimal switch points 89
Figure 41: illustration of the initial solution feasible surface 90
Figure 42: comparison of random & "v/c-based and scaled" generated initial splits solutions 93
Figure 43: comparison of random & "v/c-based and scaled" generated initial splits solutions 94
Figure 44: General verification, validation and calibration framework, Rakha, et al., [91] 95

Figure 45: HCM's example; control delay components outputs 96
Figure 46: OTS outputs for HCM "multi period delay calculation" example 97
Figure 47: OTS's control delays outputs 97
Figure 48: OTS control delay error percentage 98
Figure 49: Reston Parkway Network.....100
Figure 50: Reston Parkway P.M. Northbound Traffic.....101
Figure 51: Reston Parkway P.M. Southbound Traffic.....102
Figure 52: The Reston Parkway Network Critical Routes Scenarios 103
Figure 53: Problematic Scenario 5 critical movements and congestion diagnosis 104
Figure 54: Scenario 5 control strategies 1, Phase re-service at Sunset Hill Rd. intersection.....106
Figure 55: Scenario 5 control strategy 2, Metering at Dulles Toll road EB off- ramp intersection.....107
Figure 56: Metering intersection, gated link and the storage capacity in Reston Parkway Network.....107
Figure 57: Oversaturation offsets on Reston Parkway 109
Figure 58: Scenario 5 with Phase Re-service: Network performance measures of the optimal control strategies 111
Figure 59: Scenario (5) with Metering: Network performance measures of the optimal control strategies 112
Figure60: Scenario 5 Optimal Control Strategies Time-of-peak period for performance measures (Delay, Stop, and Throughput) 113
Figure 61: I-610 loop and US-59 interchange, Houston, TX 118
Figure 62: Houston Galleria, Uptown District, skyline view 119
Figure 63: Post Oak parking lot facilities and highways accesses points 120
Figure 64: Network exits & routes to the highways interchange ramps 120
Figure 65: critical routes pass through Richmond Ave. 121
Figure 66: Control Plan 1, 2 and 3's optimal timing plans and their optimal switch points 124
Figure 67: Richmond Ave. WBT P-ratios during the analysis period (3:30-6:30 P.M.) for the base and tested control plans 127
Figure 68: P-ratios on EBT P-ratios during the analysis period (3:30-6:30 P.M.) for the base and tested control plans 128
Figure 69: Network reserve queuing capacities percentages 129
Figure 70: Network reserve queuing capacities percentages 130
Figure 71: Network reserve queuing capacities percentages 130
Figure 72: Network reserve queuing capacities percent 130
Figure 73: The P-ratios of critical/non-critical links of the base timing and tested control plans TP1, TP2, and TP3 131
Figure 74: Traffic loading 132
Figure 75: Network traffic load, throughput reduction %, and control objective for the tested control plans....133
Figure 76: Control plan1 held back demand profiles 134
Figure 77: Control plan 2 held back demand profiles 135
Figure 78: Control plan 2 held back demand profiles 136
Figure 79: Comparison of network traffic loading 137
Figure 80: Comparison of network total loading reduction %.....137
Figure 81: Total network held back demand 137
Figure 82: Breakdown of held back demand 138
Figure 83: route (A), WBT on Richmond Ave. 138
Figure 84: Network critical routes travel times (max, average, and min).....138
Figure 85: critical route (B), EBT on Richmond Ave., ...139
Figure 86: Route B travel times (max, average, and min) 139
Figure 87: Route (H), WBT on Richmond Ave. then SLT on Rice Ave. 139
Figure 88: Route H travel times (max, average, and min).....139

Figure 89: Route (G), EBT on Richmond then NLT n Sage Ave. 140
Figure 90: Route G travel times (max, average, and min) 140
Figure 91: Route (A) travel time profile for (a) Control Plan 1, (b) Control Plan 2, and (c) Control Plan 3 141
Figure 92: Route (B) travel time profile for (a) Control Plan 1, (b) Control Plan 2, and (c) Control Plan 3 143
Figure 93: The impact of control objective switching time on route (A) travel time 144
Figure 94: The impact of control objective switching time on route (B) travel time 145
Figure 95: Network critical routes (A, B, G, and H) travel time distributions 146
Figure 96: Network critical routes (J and K) travel time distributions 146
Figure 97: Relationship between congestion level and reliability 147
Figure 98: Oversaturation timing spreadsheet flow logic; input sheets, output sheets, and macros 162
Figure 99: OTS volume parameters input cells 162
Figure 100: OTS volume parameters input cells 163
Figure 101: OTS shockwaves parameters input cells 163
Figure 102: OTS Delay calculation parameters input cells 164
Figure 103: OTS queue management parameters input cells 164
Figure 104: OTS optimizer (objective function) parameters input cells 165
Figure 105: OTS optimizer (multi-period) parameters input cells 165
Figure 106: OTS optimizer initial solution parameters input cells 166
Figure 107; OTS timing control constraints input cells 166
Figure 108: OTS timing control constraints (barrier relaxation) input cells 167
Figure 109: OTS control strategy (flare-the-green) parameters input cells 167
Figure 110: OTS offset design parameters input cells 168
Figure 111: OTS offset design parameters input cells 168
Figure 112: OTS EBT offsets time-space diagrams outputs 169
Figure 113: OTS timing plans evaluation parameters input cells 169
Figure 114: OTS optimal timing plans evaluation parameters input cells 170
Figure 115: OTS control plans selection & switching parameters input cells 170
Figure 116: OTS control plans selection & switching parameters input cells 171
Figure 117: OTS control plans selection & switching output diagram 171

TABLE OF FIGURES:

Table 1: SCCOT's cycle length increments 30
Table 2: SCOOT levels of optimization 33
Table 3: Congestion Attributes Taxonomy.....44
Table 4: Network critical movements level 2 detail coefficients (D2) 62
Table 5: Network critical movements level 3 detail coefficients (D3) 62
Table 6: Oversaturation phases and the recommended control objectives 80
Table 7: OTS and HCM delay control and the corresponding error.....98
Table 8: Reston Parkway timing plans during P.M. peak period 100
Table 9: Reston Parkway network critical routes scenarios 103
Table 10: Reston Parkway network Scenario 5 oversaturation attributes.....105
Table 11: The operational objectives of the applied control strategies.....108
Table 12: scenario 5 timing plans and applied control strategies 109
Table 13: Scenario 5 control strategies Total Improvement Percentage during Peak Period 114
Table 14: t-Statistical test for the phase re-service control strategies' performance measures.....115
Table 15: t-Statistical test for the metering control strategies' performance measures 115
Table 16: Post Oak network oversaturation attributes.....121
Table 17: Post Oak applied mitigation strategies 122
Table 18: Oversaturation control plans objectives and schedules 122
Table 19: Enumeration of optimal timing plans in the switching problem.....123
Table 20: Optimal Control plans description.....123
Table 21: Optimal TP for the developed control plans (1, 2, and 3) 125
Table 22: Overall network performance for the base timing and the developed control plan 126

1 INTRODUCTION

For the first time in history, at unknowable moment in 2008, the majority of the world's human populations now live in urban areas. Cities have absorbed nearly two-third of global population growth since 1950. As traffic congestion, accidents, and other transportation-related problems continue to increase, improvement to transportation infrastructure through the traditional means of capacity enhancement has proven ineffective. Building more roads becomes undesirable for economic, social, and environmental concerns. Also, many studies have concluded the fact that building more capacities would eventually induce more unforeseen vehicular demand. The provision of an efficient traffic signal control system has become extremely important in urban areas due to the recent public awareness of the impacts of the high congestion levels on urban environment and quality of life.

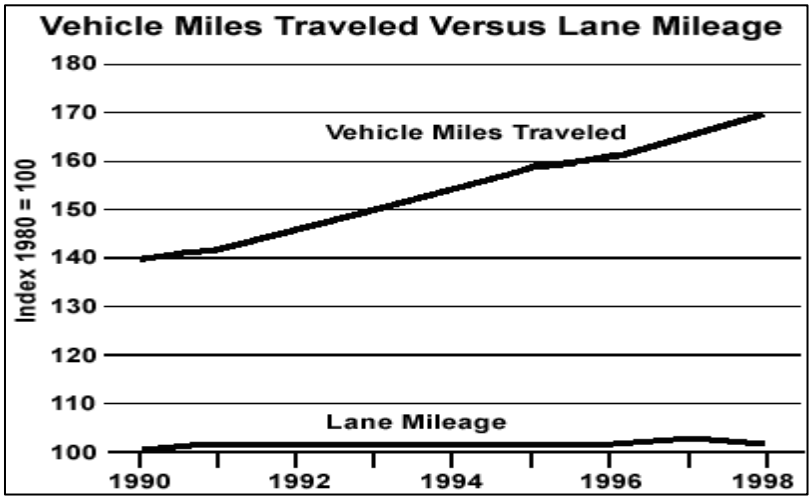


Figure 1: Vehicle Miles Traveled Versus Lane Mileage (Source: FHWA; "Finding Out What America Thinks", Publication No. FHWA-OP-01-018), used under fair use, 2012

1.1 Research Background and Motivation

Operation of signal systems during congestion regimes, unlike under saturated conditions, is a challenging task. As intersections capacities drop from processing all arriving demands, queues start building up. Subsequently, links spillbacks, storages bays are blocked, and starvations prevent intersections from functioning effectively. Mitigation of oversaturated conditions involve trade-offs between the storage of queues from the oversaturated routes to other less saturated routes. In oversaturation, typical traffic *control strategies* do not work as efficiently, particularly since the objectives need to be decidedly differently when mobility is restricted. Counter-intuitively or perhaps paradoxically, the same control strategy that provides user-optimal delay minimization in under-

saturation can in some cases work not in favor of the minimizing total delay when system become oversaturated.

Many of traffic engineering experts have concerns about developing a precise definition of the saturation condition. However, most experts agree that, in oversaturation conditions, demand exceeds capacity. Some practitioners considers at least one 15-minute period where the queue was persistent as a stipulation to call a facility oversaturated. Several of the practitioners view the oversaturation based on the ability of the traffic facilities to store the excess queues. Another perspective focuses on the growth rate of the residual queue as an important indicator of oversaturation. The common practice in dealing with congestion is to implement strategies (where possible) to prevent oversaturation from occurring in the first place, rather than reacting to the issues after the fact. Another common strategy for dealing with oversaturation is to start making simple changes to splits or phase sequence before moving to more complex approaches. Another strategy is to select a "loser" intersection along the arterial and store traffic on the approaches to this location, to the benefit of the other intersections in the system. However, most of the practical methods in use are just a "tweak" to the pre-calculated splits and offsets values in the field based on observed conditions that does not rely on any software tools, particularly when residual queues are prevalent.

1.2 Research team

The development of optimal control strategies during oversaturated conditions was a task of a National Cooperative Highway Research Program (NCHRP) project 03-90. NCHRP was created in 1962 to conduct research in critical problem areas that affect highway planning, design, construction, operation, and maintenance nationwide. The research project was a collaborative effort between researchers at: Kimley-Horn & Associates, University of Minnesota, University of California-Berkeley, and Virginia Tech, with a goal to develop a "practitioners Guidance" for operation of traffic signal systems in oversaturated conditions". Upon completion of the studied problem by this research, the problems and recommended solutions are presented in an NCHRP report. Information contained in these reports is considered to be the most current, nationally recognized data on the topic presented. The information contained in these reports is usually adopted in subsequent issuances of the design manuals that host the subject topic.

1.3 NCHRP

"*Systematic, well-designed research provides the most effective approach to the solution of many problems facing highway administrators and engineers. Often, highway problems are of local interest and can best be studied by highway departments individually or in cooperation with their state universities and others. However, the accelerating growth of highway transportation develops*

increasingly complex problems of wide interest to highway authorities. These problems are best studied through a coordinated program of cooperative research. In recognition of these needs, the highway administrators of the American Association of State Highway and Transportation Officials initiated in 1962 an objective national highway research program employing modern scientific techniques. This program is supported on a continuing basis by funds from participating member states of the Association and it receives the full cooperation and support of the Federal Highway Administration, United States Department of Transportation. The needs for highway research are many, and the National Cooperative Highway Research Program can make significant contributions to the solution of highway transportation problems of mutual concern to many responsible groups. The program, however, is intended to complement rather than to substitute for or duplicate other highway research programs."[1]

1.4 Research Objectives

The dissertation objective is to define a multi-objective methodology to develop and evaluate mitigation strategy (or combinations of strategies) for oversaturated conditions. The goal is to provide a practical guidance for practitioners to identify oversaturated scenarios and apply appropriate strategies. The research focused on identifying traffic control strategies that can be implemented by traffic signal systems to handle certain types of oversaturated conditions on surface streets

There are four objectives of this research, and can be summarized as follow:

1. Define oversaturation attributes taxonomy, identify operational objectives during oversaturation phases, catalog, and classify mitigation control strategies for isolated intersection, arterial network, and grid network.
2. Develop a wavelet-based method to assist identifying network critical routes during oversaturated conditions.
3. Develop a framework for arterial network signal timing during oversaturated conditions to generate optimal timing plans and then evaluate the optimal timing plans with a multi-objective approach.
4. Develop signal timing tool for oversaturated arterial network to address some of the limitations in the existing timing software's and apply the developed tool to two congested network case studies.

1.5 Dissertation Layout

Chapter 1, the introduction, includes background, research motivation, and objectives of the research. In chapter 2, an extensive literature review was documented. The literature review includes a critical review of existing and past research in the identification of oversaturation and performance measures

and strategies for addressing oversaturated conditions of various extents. Chapter 3 presents researches methodologies used in the dissertation. Chapter 4 presents the attribute of the oversaturation; congestion definitions, problematic scenarios, control objectives, and control strategies. Chapter 5 introduces a proposed method that investigates the potential wavelet filtering techniques in determining the network critical routes for control purpose. Chapter 6 presents a comprehensive timing framework for oversaturated networks. The framework consists of three major components: control strategy development, timing optimization, and timing plans evaluation. The development of oversaturation timing tool is presented in Chapter 7. Two cases studies are presented in chapters 8 and chapter 9. Finally, chapter 10 provides a summary of the findings and the conclusions of the research effort and future recommendations.

2 Literature Review

2.1 Traditional Signal Control Objectives

The main purpose of installing a traffic signal on a roadways intersection is to provide safe and equitable right-of-way to a number of competing movements. In general, the intersection control objectives can be classified into two basic categories, delay minimization and progression bandwidth maximization. Delay-oriented control approaches requires minimizing delay, stop, and queue length. While bandwidth-oriented approaches target maximizing network progression by optimizing four signal-timing parameters (e.g., cycle length, green split, offset, and phase sequence). Webster's Delay minimization model represents the fundamental equation of signal timing for an individual intersection. Webster [2] developed a model to estimate intersection delay using a deterministic queuing analysis and empirical result from simulation.

$$d = \frac{c(1-\lambda)^2}{2(1-\lambda x)} + \frac{x^2}{2q(1-x)} - 0.65\left(\frac{c}{q^2}\right)^{1/3} x^{(2+5\lambda)} \qquad [2\text{-}1]$$

Where:

d = average overall delay per vehicle (seconds),

C = cycle length (seconds),

λ = proportion of the cycle that is effective green for phase under-saturation (*g/C*),

v = arrival rate (vehicles/hour),

c = capacity for lane group (vehicles/hour),

q = flow rate (veh/hour)

S : saturation flow (veh/hour)

x : degree of saturation

The first term of Webster delay formula is an expression for delay when vehicles arrive at uniform rate. The second term of the equation considers the random nature of the arrivals. While, the third term in the formula is purely an empirical term that is aimed to bridge the gap between theoretical and practical results. According to Webster's theory for isolated intersection, the minimum (optimal) cycle length of an under-saturation intersection is calculated as follow:

$$C_m = \frac{1.5L+5}{1-Y} \qquad [2\text{-}2]$$

Where:

Cm: minimum cycle length

L : Total loss time per cycle

Y : sum of flow (volumes/ saturation flow) ratios for all critical phases

Highway Capacity Manuel (HCM) calculates intersection delay based on Webster uniform delay equation, Akcelik incremental delay equation, and initial queues delay equation [3]. Thus, the HCM delay equation consists of three terms that expresses: uniform arrival delay, incremental delay that capture random arrivals and overflow delays, initial queue delay as the following equation shows.

$$\text{Delay} = F.\ d_1 + d_2 + d_3$$

$$d_1 = \frac{0.5 \times C \times \left(1 - \frac{g}{C}\right)^2}{1 - \left[min\left(1, \frac{v}{c}\right) \times \frac{g}{C}\right]} \qquad [2\text{-}3]$$

$$d_2 = 900 \times 0.25 \times \left[\left(\frac{v}{c} - 1\right)\sqrt{\left(\frac{v}{c} - 1\right)^2 + \frac{4 \times \left(\frac{v}{c}\right)}{0.25 \times c}}\right] \qquad [2\text{-}4]$$

$$d_3 = \frac{1{,}800 \times Q_b \times (1+u) \times t}{C \times T} \qquad [2\text{-}5]$$

Where:
d_1: Uniform delay
d_2: incremental delay
d_3: delay due to preexisting queues (for under saturation $d_3 = 0$)
F : progression factor
C: cycle length
c: capacity
v: approach volume
g: effective green
Q_b: size of initial queue
T: analysis period, hour
t: duration of oversaturation within T, hour
u: delay parameter

Roess et al. [4] defined the bandwidth as *"The time between the first and the last vehicle that pass through the entire arterial system without stopping".* The bandwidth-oriented approaches in signal timing seek to obtain the maximum bandwidths for arterial both directions. Little et al. [5] developed mixed-integer linear program (MILP) model based on the mathematical bandwidth maximization model. The developed model generates signal timings parameters that determine the followings:

- The largest possible bandwidth of equal size for both directions of an arterial
- The favorable direction of an arterial with a large bandwidth

Based on maximum progression theory, Little *et al*.,[5] developed a model called "MAXBAND" which could yield a global optimal solution, finding cycle time, offsets, and NEMA phase sequence patterns to maximize the weighted combination of the bandwidths in both arterial directions.

2.2 Congestion Diagnosis & Estimation

Generally, an oversaturated intersection is defined as one in which the demand exceeds the capacity. Therefore, a network of intersections would become oversaturated when the system is overloaded with heavy demand which exceeds the total capacity of the network. Based on this definition of oversaturation, the v/c ratio (v is demand and c capacity) is thereby used in theoretical formula to identify whether an intersection or a network is congested. Analytically, the v/c ratio can be estimated directly by dividing the demand flow rate by the capacity. According to the relationship between saturation flow rate and capacity, v/c ratios for each lane group are calculated by using the following equation:

$$\boldsymbol{x_i = \frac{v_i}{c_i} = \frac{(v/s)_i}{(g/C)_i}} \qquad \textbf{[2-6]}$$

Where, X_i is the degree of saturation (v/c ratio) for lane group (i); v_i and c_i are demand flow rate and capacity for lane group (i), respectively; $(v/s)_i$ and $(g/C)_i$ are flow and green ratio for lane group (i) respectively. A lane group with (v/c > 1) is identified as oversaturated. For a single intersection with two competing demands, Gazis [6] expanded this concept to diagnose oversaturation by proposing the following model:

$$\frac{q_1}{s_1} + \frac{q_2}{s_2} > 1 - \left(\frac{L}{C}\right) \qquad [2\text{-}7]$$

Where, q_1 and q_2 are arrival rates for two directions; $s_1$and q_2 are saturation flow rates for two directions; L is the total lost time and C is the cycle length. Since Equation [2-8] only fits for intersections with fixed cycle length and lost time, Green [7] modified it to Equation [2-8] , which was called "absolute" oversaturation to deal with the situation when C is not a fixed value.

$$\frac{q_1}{s_1} + \frac{q_2}{s_2} > 1 \qquad [2\text{-}9]$$

Direct application of the above models is difficult, partly because of the uncertainty of the capacity and saturation flow and partly due to the difficulty to measure the arrival flow using current data collection systems especially under congested situations (the very conditions that we are trying to identify). Alternatively, some characteristics of oversaturation can be utilized to diagnose this condition.

2.2.1 *Definitions of Congestion and Level of Service*

Longley [8] identified two types of urban traffic congestion; a *primary* congestion that is caused by the development of queues at signalized intersections, and a *secondary* form of congestion that is caused by the blockage of un-signalized intersections by *primary* congested traffic. Longley presented a procedure for controlling congested controlled networks, when primary congestion is unavoidable. The basic premise of Longley's procedure is to manage queues so that a minimum number of

secondary intersections are blocked. This distinction between primary and secondary congestion and the relationship to signalized and un-signalized intersections is not useful in our work here as we are primarily concerned about signalized intersections.

Pignataro, et al., [9] defined traffic operations in controlled network based on congestion levels as:

a) *Uncongested Operations*: The situation where there is no significant queue formation. Traffic performance may range from very low demand per cycle to conditions where the demand is a significant fraction of the capacity value. Short queues may occasionally occur, but do not last for any length of time (presumably this "any length of time" was meant to indicate "not more than a few cycles").

b) *Congested Operations*: Refers to the entire range of traffic operations which may be experienced when traffic demand approaches or exceeds the road and/or intersection capacity. Furthermore, congested operations can be divided into two sub-categories: *saturated* and *over-saturated* operations.

i) Saturated Operations: is a term that describes that range of congestion wherein queues form, but their adverse effects on the traffic in terms of delay and/or stops are local.

ii) Oversaturated Operations: is characterized as a situation wherein a queue exists, and it has grown to the point where upstream traffic operations are adversely affected.

NCHRP 3-38 defined congestion (saturated) traffic conditions as follows:

- ***Local congestion***: it occurs when more vehicles face a cycle failure that does not result in damaging or excessive queues.
- ***Extended congestion***: cycle failures repeat such that queue extend damagingly through upstream intersections, causing the capacity of the intersections to be reduced.
- ***Regional congestion***: This occurs when the queue from critical intersection joins or influences the queue at upstream critical intersections.
- ***Intermittent congestion***: This occurs as a natural result of stochastic traffic arrival. Even at light volumes timed using the Poisson method, one expects cycle failures 5% of the time.
- ***Recursive (cyclical) congestion***: This is congestion at an intersection with insufficient capacity that occurs predictably as a result of foreseeable demand patterns.
- ***Prolonged congestion***: congestion at queuing intersections create such inefficiencies that demand must fall below the reduced capacity for extended periods to permit residual queues to clear.

Intuitively, the "congestion level" or LOS (Level of Service) can be utilized to recognize an oversaturated condition. These different descriptions for congestion are qualitative at best, and

cannot be measured directly. This taxonomy for congested conditions does little to help us map palliative strategies to saturated conditions and thus we explore more quantitative measures of previous research, such as queue development, travel delay/time, travel stops/speed, flow-density/occupancy relationship, and green time utilization for diagnosis of oversaturation.

2.2.2 *Definitions based on Queue Length*

Due to heavy demand or insufficient capacity, oversaturation has been characterized to relate to the growth of queues such that: "*a stopped queue cannot be completely dissipated during a green cycle*" (Gazis) [6].

This is also commonly known as a "*cycle failure"*. Another common definition of oversaturation is based on the concept of a queue such that "*traffic queues persist from cycle to cycle either due to insufficient green splits or because of blockage*" (Abu-Lebdeh & Benekohal) [10].

Similar definitions have also been clearly described by Roess et al. [4] in *Traffic Engineering* such that: "*the oversaturated environment is characterized by unstable queues that tend to expand over time with potential of physically blocking intersections (blockage, spillback), thus slowing queue discharge rates…*". Thus, it is evident that overlong queues (longer than a link length) indicate the oversaturated situation. However, for some cases, even when the maximum queue length is shorter than the link length, frequent residual queues could be considered as oversaturation. Such queuing could be due to insufficient green splits, downstream blockages, and/or heavy pedestrian flows. The FHWA study on Signal Timing under Saturated Conditions (DRAFT REPORT 6-16-07) defined different types of traffic conditions based on queue information (associated with an individual intersection).

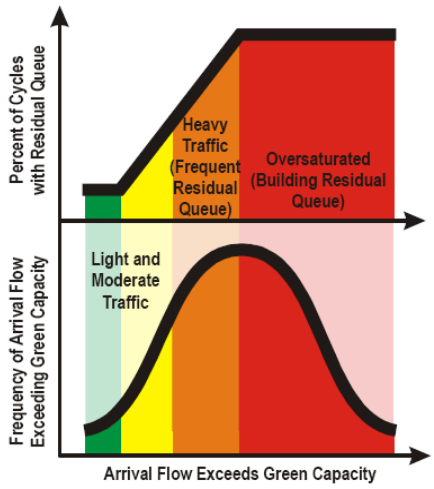


Figure 2: A Hierarchy of Traffic Conditions (Source: Draft report from FHWA signal timing under saturated conditions project), used under fair use, 2012

- **Light traffic:** Characterized by the expectation of minimized cycle failure. In such condition, the signal can fully serve arrival queues; and cycle failures are expected to be infrequent at less than 25% of all cycles.
- **Moderate traffic:** Characterized by the expectation of "fair" operation. At these intersections, drivers expect the operation to be "fair", which means that the degree of saturation on each approach is approximately the same. Some cycle failures are to be expected but do not necessarily violate the expectations of motorists. No approaches have queues that are growing disproportionately to other approaches.
- **Heavy traffic**: Characterized by frequent cycle failures, but with a residual queue that ebbs and flows without growing uncontrollably. The flows exceed capacity up to about half time due to stochastic arrivals.
- **Oversaturated operation:** Characterized either by excessive residual queues that grow (seemingly, at least) without control, therefore causing more widespread damage to the operation of a network.

Figure 2 graphically illustrates that the relationship between such traffic conditions (including light, moderate, heavy and oversaturated) and the residual queue information. When average demand equals capacity, the percent of cycles with residual queues will be 100%; while with demand exceeding capacity, a growing residual queue is inevitable. Similarly to the NCHRP 3-38 taxonomy, these qualitative descriptors are useful to guide thinking about oversaturation, but cannot be used directly to map palliative strategies to oversaturated conditions.

Queue length as well as residual queuing (queues that persist after switching from green to red) for a given phase are the preferred indicators for diagnosis of oversaturation. However, accurate queue estimation typically depends on the arrival flow information, i.e. *input*, which requires installation of upstream detectors. For most agency standard technology deployments at intersections, this requirement cannot be satisfied. Typically, advance detectors for vehicle actuation purposes are installed within a few hundred feet from the stop line, depending on the speed of approach traffic. If the queue length spills over the advance detector location, most traditional input-output approaches to estimating queue length will not work appropriately. Therefore, either additional upstream detectors need to be installed or alternative methods to estimate queue length based on a typical, current layout of detector placements needs to be developed.

2.3 Measures of oversaturation based on Delay/Stops/Speed/Travel Time

Along with the presence of long queues during oversaturated time periods, overflow delay (as part of the queuing delay), becomes one of the major components of total delay at an intersection. Figure 3

(Dion, *et al.*, 2004) compares typical delay accumulation curves for both under-saturated and over-saturated conditions. The bottom component of Figure 3 shows that the overflow delay becomes, over time, a more and more significant component of over-all intersection delay. This component is extremely difficult to measure.

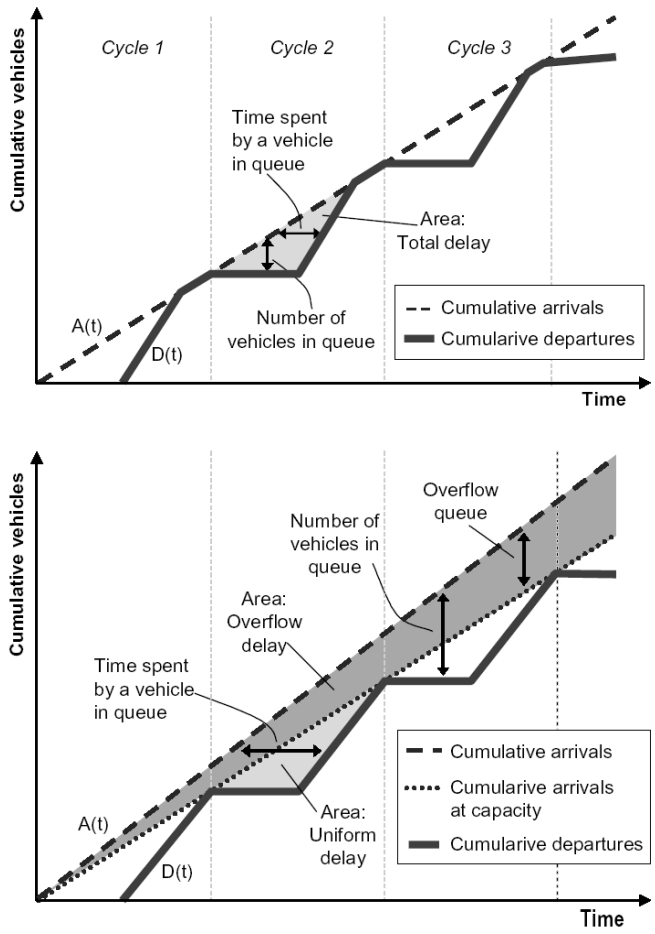


Figure 3: Idealized cumulative arrivals and departures for under- and over-saturated conditions (Source: Dion, F., *et al.*, 2004), used under fair use, 2012

Different delay models for oversaturated intersections have been proposed and compared with traditional delay models which were designed to measure delay during under-saturated conditions [11-15]. Figure 4 compares Webster's random delay model, which only applies for saturation values below 1,and theoretical overflow delay models which can extend to V/C ratios above 1.0 [4]. As illustrated by the representative figure, the estimated delays at oversaturated intersections significantly increase compared with random delays of under-saturated intersections and most research has hypothesized that this growth rate in the average delay per vehicle varies linearly as the V/C ratio increases. These approaches provide models for estimating delays for V/C ratios higher than 1, but they have little value in practice because of the difficulties in the measurement of the input

components of the model – particularly the demand volumes. Any placement of a fixed detection location, at some point, cannot capture the demand volume. This is true because as the demand grows, the point of arrival to the back of the queue continues to grow further upstream. Thus, to meet our goals to estimate oversaturated conditions with real-world detection systems, we cannot utilize delay as the primary estimator of oversaturation.

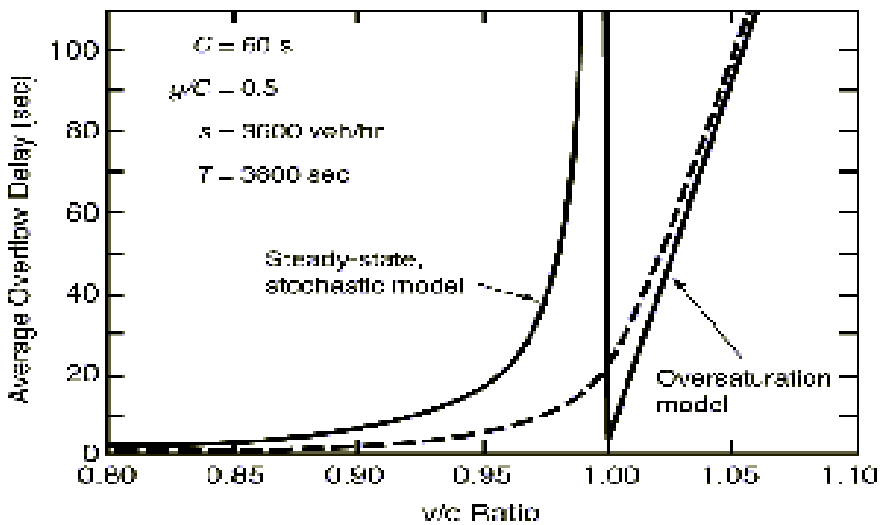


Figure 4: Random and Overflow Delay Models Compared (Source: Roess et al., 2004))

Other performance measures, such as speed (space mean and other various forms), number of stops and travel time, also change significantly during oversaturated conditions. For example, vehicles repetitively accelerate and decelerate in congested conditions leading to a significant increase of number of stops. Travel times also increase significantly as speed decreases. Speed in and of itself is not sufficient to identify oversaturation in arterial traffic systems because detected speed on arterials drops to zero when any queue forms over a detector when the traffic signal indication is red. Based on the placement of the detector, this may not indicate that an oversaturation condition is present. Most models which are designed to estimate speed, stops, and travel time performance during oversaturation are based on queues at intersections. For example, the model developed by Cronje [16, 17] is based on the expected queue size at the beginning of each cycle. Therefore, we conclude that queue estimation is the crucial component in the diagnosis of oversaturated conditions.

Oversaturation Control Strategies

A wide variety of traffic signal control strategies for management of congested conditions have been developed and proposed by various researchers over the past 30+ years. The following is a review of signal control strategies basically grouped by categories based on the extent of application of the strategy.

2.4 Oversaturation Control Strategies

2.4.1 *Isolated Intersections*

Dunne and Potts [18] developed an algorithm for controlling isolated intersection based on closed-loop control concept. Green phases were changed according to the variation in traffic demand. The concept was further improved by Gazis and Potts [6] for traffic signal control in oversaturation conditions. Gazis and Potts's control philosophy was to minimize total system delays by maximizing intersection productivity while at the same time keeping queues' development within acceptable levels. A maximum green time is given to the approach with the higher saturation flow until the queue is dissipated, while other approaches receive the minimum green time.

Gazis and Potts made some assumptions and constraints:

- Traffic demand is known and is not affected by the control decisions
- The cycle length is fixed
- The green times vary between two predetermined values (i.e., lower and upper limits)

Michalopoulos [19] used reservoir analogy to improve Gazis policy by developing an algorithm that determines the switch-over points in oversaturation conditions. ("Switch-over" refers to the time at which the green time of a certain approach alter from maximum green to minimum green allowing other approach's green time to change from minimum to maximum and vice versa [6]. Michalopoulos and Stephanopoulos [20] proposed an efficient two-stage timing method, termed bang-bang control, in attempt to find an optimal switch-over point during the oversaturated period. Later, Tang-Hsien et al., [21] show that Michalopoulos and Stephanopoulos' continuous delay model is inadequate to allocate the optimal cycle length since stops penalty is not considered in the model's formulation.

Khorso [22] demonstrate that the optimal control of an intersection will not be achieved simply by switching of green times from maximum to minimum, but rather by allowing the cycle length itself to vary within specified upper and lower limits. Khorso's improved algorithm contains subroutines that optimally determine cycles' length that result in minimizing delay as well as limiting the development of queues.

Tang-Hsien et al., [21] improved Gazes' model further by developing a discrete optimization algorithm that uses a *bang-bang-like* control method to determine the switch-over point. Unlike in Michalopoulos and Stephanopoulos continuous model, the proposed model can allocate the switching-over points exactly at the end of a cycle. The proposed model minimizes a performance index that incorporates both stops and delays over the entire oversaturated period. The discrete operation provides a smooth, regular, and ordered transfer of control and reliable delays calculations. The bang-bang like control operates alternatively and sequentially with given minimum and maximum green time boundaries. This algorithm can activate a switchover point at the end of a cycle.

According to Chang and Lin's study, such control strategies significantly outperform a policy to increase the cycle time and give each direction the same amount of green time. The performance of Chang and Lin's model is rather robust even when the input data have some measurement error. The algorithm produces optimal cycle length and timing split as its output for a trajectory of given, known input volumes.

The maximal and minimal green times are typically regulated by laws in many agencies (e.g., pedestrian crossing time). This regulation can be considered as constraints in the mathematical formulations. In general Gazis' policy, minimizing delay via split switching, is not an on-line highly responsive policy, but rather an optimization policy based on advanced mathematics and detailed knowledge of demand [9].

Longley [8] introduced a real-time "queues proportionality" concept. He suggested that green times should be adjusted to balance the queues at intersection's approaches during oversaturated conditions. The objective is to minimize the number of blocked intersections by critical intersection's queues. This strategy involves the use of sampled data feedback loops at each intersection. Longley considered the dynamic behavior of an isolated intersection first, and then the analysis was extended to a network of multiple intersections. Longley suggested that the stability of the network may be improved by the use of coordinated signals, and by establishing stability criteria for controller's parameters. The policy, however, requires exact measurements of queues.

Miller [23] suggested the use of actuated control to minimize delays in critical intersections. He suggested that the green time of each approach be extended to a maximum value, to accommodate the additional traffic demand. The proposed policy does not depend on queue measurement (i.e., number of queued vehicles), but rather on attaining a certain threshold value that triggers the actuation. The queue-actuated control was claimed to be effective in high volume levels and intersections with high turn-ins. Miller's concept was later expanded by Weinberg [24] to include downstream delays and compute the total delay. This inclusion provided coordination between critical intersection and its neighboring intersections to ensure that the reduction in critical intersection delay would not come in cost of other intersections. Weinberg's model was modified and evaluated later by Ross, et al. [25] via simulation. The simulation results indicated the policy is effective in saturated conditions in term of delay reduction. The policy implementation required special treatment regarding detectors' allocation.

For under-saturated traffic conditions, typically the offsets between two intersections are determined based on the objective of maximizing the progression band. As congestion increases and intersections become more saturated, residual queues (and/or queues from side-street turning traffic) start to disrupt movement at upstream intersections and the progression band concept does not apply

any longer. If oversaturation continues for a considerable time period, standard fixed-timing plans are likely to aggravate movement disturbance caused by the "spillback" of queues into upstream intersections since they do not take into consideration the actual traffic state [26]. This means that conventional procedures for optimizing fixed-time signal control, such as TRANSYT [27], deteriorate rapidly when severe congestion persists and alternative strategies are therefore required to manage queues along the length of an arterial.

Gordon [28] was the first to introduce the concept of utilizing the storage capacities of approaches to store queues at saturated intersections instead of minimizing the total delay. Gordon's idea was to make efficient use of the upstream link's storage capacity and thereby delay a particular approach as long as possible since it could handle the backup of upstream traffic.

Pignataro et al., [9]classified measures designed to tackle urban traffic congestion into minimal response, highly responsive signalized policies, and non-signal treatments in a signalized network. These methods are explained as follows:

1. Signalized control measures:
   - Minimal response measures:
     - (a) Intersection: Cycle length, Block length, splits, and extra phases
     - (b) System: Phases adjustment for progression, equity offsets, splits to allocate storage available.
   - High responsive measures:
     - (a) Intersection: maximum queue policy
     - (b) System: spread from intersection
2. Non-signalized treatment
   - Regulatory measures:
     - (a) Enforcement (e.g., double parking)
     - (b) Prohibition (e.g., parking, turning)
   - Operations measures:
     - (a) Turning (e.g., left bays, right bays, dual turning lanes, right-turn-on-red)
     - (b) Lane arrangements (e.g., one-way street, reversible lane)
     - (c) Disruption (e.g., pedestrian, bus stop, parking/ no-parking, mid block)

Pignataro, et al., [9]) suggested that in oversaturated conditions a more effective one-way progression scheme is to use "reverse progression" offsets rather than planning for a progression band of unimpeded flow. In a reverse progression, the offsets are set so that the green of a downstream

intersection starts before the upstream green in order to flush the residual queue. In addition, Pignataro et al. recommended that the following principles should be used to determine the signal split at the upstream intersection on an oversaturated arterial where a "reverse progression" offset is being applied:

- Reduce minor-street green time to ensure fewer turn-in vehicles will take a disproportionate amount of the storage in the oversaturated link. When restricting the turning phase time, care should be taken that this does not limit the cross-street through movement. Pignataro et al. recommended physical capacity improvements (e.g. turn lane bays and signalization) in this case.
- In allowing such cross-street movements during oversaturation, it should be recognized that the through arterial movement needs only as much green as it can effectively use at the downstream intersection. Any additional green is, in fact, wasted but could be allocated to the arterial to keep it from the cross turning movement (if signalization by movement does not exist).
- If possible, turn prohibitions should be considered on the cross street so that the crossing through movement is not impacted by queued turning vehicles on the side street.

Pignataro, et al., [9] suggest a queue management strategy using the term "equity offsets". This strategy, based on the principle of reverse progression, seeks to provide equitable treatment of competing flows at a congested intersection located upstream of an oversaturated link. Pignataro et al. identified two main cases when determining the intersection splits:

- Where there are negligible turn-ins from the cross (minor) streets at the upstream intersection, the split at the intersection can be as commonly determined (e.g. Webster's delay minimization formula)
- Where there are substantial turn-ins form the cross streets at the upstream intersection, the cross street traffic should be allowed just enough green to put its "fair share" of vehicles into the oversaturated link. Pignataro et al. did not specify in detail what a “fair share” of green time would be.

Pignataro et al., [9] did not provide mathematical equations to cover all approaches combinations and volume/turn combinations. However, they offered a set of principles that can be used to govern the formulations of the most appropriate strategy. Rathi [29] described a procedure that uses simultaneous offsets along the arterial to control signals and manage queues and negative offsets and flaring of green along cross roads.

Recent research by Beaird, et al [30] and Smaglik, et al [31] describe simple controller logic for truncating a phase early when faced with a downstream restriction of flow. While not explicitly

known (from downstream detection) that the downstream point is restricted, the logic assumes that no flow on a detector when the light is green indicates that there is nowhere for the vehicle to go, and thus it is better for overall intersection operations to move on to another phase.

2.4.2 Models for Queue Interactions between Closely-Spaced Intersections

Rouphail and Akcelik [32] presented an analytical model for predicting the effects of queue interaction on delays and queue length at closely-spaced signalized intersections. Queue interaction refers to the model's ability to predict a reduction in upstream saturation flow rate when interference from the downstream queue is significant. It was shown that the presence of downstream queues has a strong influence on the performance of the system with limited space for queuing. The queue interaction effect may alter the location of a critical intersection.

Prosser and Dunne [33] analyzed the paired-intersection problem by presenting a procedure that explicitly considers queue-blocking effects for determining the capacities of movements at closely-spaced intersections. They employed a graphical technique to estimate reduction in the effective green time. The model assumes no vehicles at upstream intersection will discharge when queue spillback occurs downstream.

Messer [34] extended the Prosser-Dunne model to incorporate a wider range of operating conditions and developed an algorithm based on this extended model to determine the effective green time, phase capacity for the two intersections, and the relative offset between the two signals. The extended model does not stipulate the downstream intersection to be oversaturated, as the blockage due to spillback may occur during under-saturated conditions.

2.4.3 *Diamond Interchanges*

Diamond interchanges operate as critical links between freeway and surface-street roadway facilities. During peak periods, inefficient operation of a diamond interchange and neighboring traffic signals may cause the system to become a bottleneck, degrading not only the capacity of the interchange but also that of the arterial and, in some cases, even the capacity of the freeway due to spillback on the exit ramps [35]. Kovalli, et al., [35]developed an extension to PASSER III to consider oversaturated conditions in optimization of diamond interchange timings. By modeling the effects of spillback, the authors claim to outperform the signal timing approach in PASSER III. The most important factor to be considered in the oversaturated diamond interchange is the management of queue formation on external approaches. Local agencies engineers tend to favor strategies that alleviate queue formation on the arterial cross street in order to enhance progression, Meanwhile state highway official prefer strategies that control queue length on the off-ramp such that vehicles do not back onto the mainline

freeway facilities and impede their operation, possibly at the expense of the arterial street network [36]. Analysis of signalized diamond interchanges must take into account the following:

a) Traffic patterns at signalized interchanges have a high percentage of turning volumes. This requires special attention to the origin-destination patterns at the diamond.

b) Queue spillback considerations are extremely important at compressed and tight urban diamonds.

Signal control at diamond interchanges traditionally has been provided by either a three-phase signal sequence or the four-phase, two-overlap signal phase sequence (TTI-4-phase). Basically, the three phase sequence favors progression for the arterial through movement, with the cross roads and arterial left-turn traffic normally stopped. Three-phase control requires sufficient queue storage space for these movements to avoid queue spillback. Four-phase with overlaps (TTI-4-phase) is commonly seen when spacing between the two intersections of the interchange is short and thus queue spillback would be a major concern [37].

Texas Urban Diamond Signal control was developed by Texas Transportation Institute [36]. This strategy incorporates both three-phase and four-phase sequences with overlaps. This control strategy seeks to maximize the benefit of both control strategies as traffic demand fluctuates. The Three-phase control tends to favor the progression for the arterials' through movements, and is typically preferred when the ramp traffic is balanced and the arterial left turn traffic volumes are low. The Four-phase-with-overlap control, on the other hand, is commonly used when the storage spacing between ramps is limited and queue spillback affects the upstream intersections. The Texas Urban Diamond Signal Control strategy switches between the two control operations mentioned above offline based on time of the day.

The "Arlington approach" was developed by the City of Arlington, Texas, for diamond interchange control under saturated conditions. The Arlington approach uses dynamic phase selection process based on detection of the critical movements (i.e. interior left-turns) to activate two-phase, three-phase, or four-phase control based on cycle-by-cycle detector information. The control approach offers flexible operation to the interior left-turning movements by incorporating protective/ permissive left-turn operation for arterial traffic. During low traffic volume, two-phase operation is applied. As the traffic demand increases, either the three or four-phase control is activated based on the demand level. Messer and Chaudhary [38] indicate that the additional phases initiated by activation of queue detector (that could be accomplished by detector switching and/or delay settings) may quickly result in "explosive" cycle lengths. The increase in cycle length and increase in the number of phases both yield increases in delay. As a result, the control strategy becomes ineffective if

traffic demand simultaneously increases to the point of saturation on several approaches to the interchange.

Kim and Messer [39] developed a dynamic model as a mixed integer linear programming problem to provide an optimal signal timing plan for diamond interchanges during oversaturated conditions. The model maximizes system productivity (throughput), minimizes system delay, and controls queue lengths on external approaches to a diamond interchange. This technique is sensitive to the knowledge of the traffic demand profile. Therefore, any significant variation in the actual traffic demand will render the signal timing plans established by this optimization approach to be ineffective.

2.4.4 *Reduced Cycle Times*

Long cycles reduce the overall proportion of time lost during phase changes, and in general, increase the capacity of an intersection. There is some debate, however, that longer cycles may not be as appropriate in oversaturated conditions. Quinn (1992) speculates that short cycles have the following advantages over long cycles:

- Short cycles allow a high saturation flow to be maintained throughout the green period (the saturation flow falls if the exit side is blocked, which is less likely for short cycles since more vehicles can be stored in the downstream link if it is red);
- Short cycles are useful for clearing intersections blocked by turning traffic; and
- Short cycles provide more frequent opportunities for pedestrians to cross.

Since a reduced cycle time decreases the capacity of the intersection as a whole during free flow conditions, it must be allowed to revert to its original value as soon as it has achieved its objective to manage the oversaturated condition (Quinn, 1992).

2.4.5 *Metering/Gating*

In the context of queue management, there are many strategies to be considered at the micro-level. Such measures attempt to "meter" or impede traffic input at a suitable point or points upstream to prevent the demand from reaching critical levels at critical downstream locations. These strategies can be applied either on a local level, to protect a particular intersection, or on an area-wide level. The practitioners must determine suitable links with enough spacing to store (or gate) the metered traffic. The gated links are those links which have been designated to store the queues which would otherwise block the bottleneck link. When the bottleneck link is congested the green time is reduced on the gated links.

Rathi [40] discussed queue management control strategies that manage (or meter) the rate of flow into and within high-traffic density networks. He broadly categorized metering control strategies as “internal” metering and “external” metering approaches. Lieberman, et al. [41] developed a real-

time traffic control policy for congested arterials. A control policy, known as RT/IMPOST (real-time/internal metering policy to optimize signal timing), is designed to control queue development on every saturated approach by suitably metering traffic to maintain stable queues. The control objectives are to (a) maximize system throughput, (b) fully use storage capacity, and (c) provide equitable service. Consistent with this approach, bounds on queue lengths and signal offsets are determined using a mixed-integer linear program (MILP). A simulation study was conducted by Lieberman to compare four signal's timing tools, including RT/IMPOST, PASSER, TRANSYT, and SYNCHRO. The result showed that RT/IMPOST policy yielded improved network travel speed and delay during oversaturated conditions. RT/IMPOST considers the intersections' turning movements and precisely controls each phase's duration at every cycle length to ensure that downstream intersections queue lengths lie within the levels defined earlier in the optimization process. Hence, detector "blackout" effects (where the queue occupies the detector) should be limited by carefully locating the advance detectors. However, the procedure has a limitation that it does not support lagging left-turn phases and still requires detailed knowledge of approach volumes over time.

Girianna and Benekohal [42] developed a GA-based procedure that produced signal coordination timing for grid-networks of oversaturated one-way arterials. The algorithm provides signal timings that are responsive to traffic demand variations. The proposed procedure applies an on-line load-balancing mechanism to protect critical intersections from becoming oversaturated. Therefore, positive and/or negative progression offsets strategies can be employed depending on the location of the critical intersections (i.e., entry or exit points). The developed algorithm adopts a two-stage strategy where queues are dissipated first before progression is achieved. The duration of the first stage depends on the location of the critical intersection. When critical intersections are located at exit points of the network, all upstream signals' entering traffic volumes are metered by setting lower green times. The metering strategy is combined with setting negative offsets at the exit signals. Later, the offsets are gradually set to positive values as the algorithm promotes forward green bands. When critical signals are located at entry points, the negative offsets are maintained for longer duration at the entry signals. This is done to ensure that all local queues are cleared before more vehicles arrive at downstream signals. The common cycle is changed depending on arrival volumes and the optimal effective green times for the queue-dissipation process. The algorithm requires an efficient use of the green time and de-facto red avoidance. The efficient use of green can be achieved by determining the ideal offset between intersections. De facto red exists when the signal is green but traffic cannot proceed because of backed-up traffic on a receiving street [43]. The de facto red can be avoided by allocating less green time to upstream intersections than downstream intersections. The main issue with the Girianna and Benekohal's algorithm is that it does not explicitly accommodate

turning movements from side streets. In addition, the de facto red constraint makes it difficult to be applied in a two-way-street network. Apparently, the procedure would be ineffective if congestion occurs in both entering and exiting points, since the algorithm assumes that congestion will occur in either the entering or exiting points. In a later effort, Girianna and Benekohal [44] extended the work with a procedure for dissipating queues on two-way arterials. While the method is intriguing, its reliance upon known volumes is problematic for application in practice.

Internal Metering strategies encompass the Critical Intersection Control, arterial strategies that control the flow along congested arterial roads, and grid strategies that control the flow along major arterials and along minor cross-streets in order to prevent "gridlock" conditions. External metering strategies, on the other hand, meter traffic along the periphery of a control area to limit the inflow of traffic, while servicing demand at an acceptable level to improve the overall quality of traffic flow within the control area. That way, the overall performance of the affected traffic should be improved. Rathi and Lieberman [40] indicated that the external metering control strategies have the potential to improve traffic operations within and on the approaches to a congested control area [45].

2.4.6 *Recovery from Congestion*

Most, if not all, traffic signal control strategies were developed to prevent, or at least prolong, the onset of congestion. Some strategies have been developed to manage situations of oversaturated conditions. Even fewer control strategies have been developed specifically to speed recovery from congested conditions. Many standard traffic management strategies can increase capacity and hence prevent or postpone the onset of congestion. However, once secondary congestion occurs, specific methods are required to cope with it and eventually, to disperse it. Here, the principle of queue management becomes very important. In terms of recovering as quickly as possible from severe congestion, Quinn identified the following approaches: 1) modify the control system to disperse the critical queues, 2) provide reserve capacity to relieve congested links, and 3) reduce temporarily the level of demand [45]. These recommendations are not followed by detailed analytical algorithms to achieve these objectives.

Daganzo [46] introduced a cell transmission model (CTM) to capture traffic flow dynamics. Having the network divided into small cells representing the distance that vehicle can travel at free flow time; CTM is used as a numerical approximation of hydrodynamics theory (LWR 1955). Lo, et al. [47-49] applied the CTM to control oversaturated networks by introducing a Dynamic Intersection Signal Control Optimization algorithm (DISCO) that uses the CTM as the calculation engine. DISCO produces timing plans that untie the gridlock in a few cycles. Due to its dynamic nature, DISCO supports fixed green splits in fixed-cycles or fixed-time plans as well as timing plans with

variable green splits, , completely removing the concept of a ‘‘cycle.’’ A genetic algorithm is used to find a near-optimal solution. However, the DISCO model is, like most past work, sensitive toward the quality of traffic volume data. Another limitation for practical implementation of DISCO is the substantial computational power needed to solve the optimization problem for a large network.

Chang and Lin [50] presented a dynamic control method that incorporates a bang-bang-like model to improve TRANSYT-7F’s performance in dealing with oversaturated signalized networks. Their method considers all the over-saturated period until all intersections became under-saturated. The cycle length of every oversaturated intersection in the network is assigned to be equal to the cycle length of a “pivot” intersection, which is defined as the most congested intersection at a certain cycle period. A branch and bound search method is utilized to look for progression routes that maximize the throughput of the network considering the progression priority. The proposed pivot search method then moves from one intersection to another as intersections change from over-saturated to under-saturated. When a new pivot intersection is identified, a fresh timing plan is generated on the basis of the new pivot intersection’s conditions. Finally, a smoothing procedure was introduced to minimize the effects of transitioning between timing plans.

2.4.7 *Dynamic Optimization Algorithms*

A number of strategies have been developed to improve the signal operation in oversaturated conditions (Abu-Lebdeh et al. 1997; 1999; 2000; 2003; 2001, Park et al. 1999; 2000). These strategies generally attempt to accomplish the following:

- Identify the queue and the queue discharge time.
- Identify the downstream storage available for queue discharge.
- Maximize throughput by avoiding the provision of green time that cannot be used or is inefficiently used because traffic cannot flow during the green periods.

Abu-Lebdeh, et al. [51] developed a dynamic algorithm to obtain an optimal or near optimal-signal control trajectory (i.e. changes to offsets, splits, and cycle length) so system throughput is maximized subject to constraints on state and control variables (green times, and offsets) designed to prevent occurrence of de-facto red. Abu-Lebdeh, et al. [10] extended the previous research with a dynamic traffic control algorithm that can be customized for different priorities to arterial and cross street traffic in order to attain a desired traffic management strategy. The proposed procedure consists of two components: (1) a dynamic signal control algorithm that utilizes queue information to set different signal parameters to maximize the system throughput, and (2) a disutility function that evaluates the algorithm response based on the selected system performance goals. A “real-time queue information feedback” mechanism is needed for practical application of the strategy in the field. With

queue estimation of field conditions, the algorithm could then compare the information with its projected values, and re-solve the optimization problem.

Park, et al. [52] developed an extension to TRANSYT that finds signal timing parameters (cycle, split, and offset) in oversaturation conditions using a genetic-algorithm as the search routine. Kovvali, et al. [35] developed mesoscopic simulations software (Arterial Signal Coordination Software, ASCS) that can optimize diamond interchanges in oversaturated conditions. The strategy also uses a genetic algorithm approach to obtain near optimal solutions that encompass cycle length, phase sequences, and ring lag/ internal offset. Both efforts were focused on the search algorithm and modeling of oversaturation more so than any generalizations or findings about the differences in the resulting timings versus the timings that is produced by naïve TRANSYT or PASSER III. In both cases, demand volumes over time are necessary to run the models and obtain the "optimal" timings and phase sequence.

2.5 TRANSYT-7F

TRANSYT (Traffic Network Study Tool) is a macroscopic, deterministic, simulation and optimization model developed by Dennis Robertson in 1967 [53]. Version 7 of TRNSYAT was modified by Federal Highway Administration in 1981 to accommodate driving on the right. In TRANASYT 7F, signal timing parameters are individually optimized in a serious of hill-climbing searches. A macroscopic simulation is used to evaluate signal timing plans. TRANSYT estimates arrival flow for each approach in the network. In case of link without upstream intersection, uniform arrival rate is assumed. Otherwise, TRNASYT models arrival traffic as cyclic flow profile (CFP) and automatically update links' input and output flows. TRANSYT uses a platoon dispersion model to laterally disperse vehicles in a platoon that is proportion to the distance to the downstream intersection [53]. Queues in TRANSYT's early versions were modeled vertically. Nonetheless, in recent versions after V8.1 queues were estimated horizontally to improve congestion modeling. Therefore, queues size, delay, stops, link's spillback, and storage bay overflow are accurately estimated. Conceptually, this model is better suited for the analysis and optimization of congested (oversaturated) facilities.

TRANSYT uses a hill-climbing technique to search for optimal signal settings that optimize a chosen performance index. TRANSYT heuristic consists of series of optimization stages. First, an optimal cycle length is determined from a pre-specified interval. Then phases' splits can be specified based on degree of saturation of conflicting movements. For the initial optimize cycle, offsets are adjusted by performing a *line search* to improve the global objective function, while splits are kept constant. The search algorithm continues incrementally adjusting the offset as long as objective function value is improving. As soon as the adjustment degrades the objective function, the direction

of the adjustment is reversed with same step size. The search is alternated with different step sizes to escape from local minimum region. Thus, TRANSYT climbing-hill procedure does not guarantee convergence to optimal solution [54].

TRANSYT-7F incorporates genetic algorithm search technique with climbing-hill to improve and accelerate the convergence to optimal solution. This arrangement allows optimizing cycle length simultaneously with splits and offsets. TRANASYT 7F seeks to optimize wide variety of objective functions; including disutility minimization, progression opportunity maximization, throughput maximization, and other hybrid objectives. User can specify only one performance measure as objective function for the entire analysis period. The following is a list of TRANSYT-7F optimization objectives:

1. DI
2. PROS
3. PROS then DI
4. PROS/DI ratio
5. QR * DI
6. THRU/ DI ratio
7. THRU then DI
8. THRU V/C

Where,

DI: Disutility index, measures the traditional delay and stop minimization

PROS: Progress opportunity, measure of the number of successive green signals that vehicle may expect to progress through without stooping at design speed

QR: Queuing ratio, it measure the average back of the queue on a link divided by the maximum number of vehicles that may be accommodated on the link.

THRU: measure of throughput

V/c: volume-to-capacity

The standard delay and sips DI, is the primary objective function within TRANSYT-7F, is defined as follows:

$$DI = \sum_{i=1}^{n}\left[\left(w_{d_i}d_i + K.w_{s_i}s_i\right) + U_i\left(w_{d_{i-1}}d_{i-1} + K.w_{s_{i-1}}s_{i-1}\right) + QP\right] \qquad [2\text{-}10]$$

Where:

d_i: Delay on link (i)

s_i : Stops on link (i) per second

U_i : Binary variable that is '1' if link-to-link weighting has been established, zero otherwise

w_{x_i}: Link-specific weighting factors for delay (d) and stops (s) on link (i)

K : User-specified stop penalty factor to express the importance of stops relative to delay

QP : Queuing penalty

TRANSYT-7F under congested conditions

In the recent releases of TRANSYT-7F, release 8 and up-to-date, a new objective functions were added to handle oversaturation conditions (i.e., minimize queue and maximize throughput). The followings are review of TRANSYT-7F new features for oversaturated network.

1) **Multi-stage optimization**

TRANSYT-7F performs multi-period optimization using the genetic algorithm. The program can optimize the splits and offsets of selected period by the user, or all analysis periods simultaneously.

2) **Congestion matrix**

TRANSYT-7F incorporates congestion matrix. This concept, congestion matrix, represents the physical space available in the network. Thus, links' overflow can be identified and penalized in objective function calculation.

3) **Grouped nodes**

Grouped nodes capability in TRANSYT-7F allows users to operate closed-space intersections as one intersection. Grouped nodes timing can be set initially by user. During optimization process, the grouped node offsets will be held constant as splits and offsets of the surrounding nodes will be optimized normally.

4) **Uncoordinated optimization:**

TRANSYT-7F allows the user to specify "Allow Split Optimization" for each individual intersection, but there is no option to "Allow Cycle Length Optimization" for a subset of intersections. If cycle length optimization is requested, the network cycle length must be optimized at all intersections on the optimization node list, regardless of whether or not they are coordinated.

5) **Phase sequences optimization**

The purpose of phasing sequence optimization generally is to enhance progression, yet delay can sometimes be reduced as well. TRANSYT-7F performs phasing sequence optimization on both major and minor streets. During optimization, TRANSYT-7F examines practically all feasible phasing sequences including leading and lagging left-turns, lead-lag phasing, and split phasing. However, user can specify whether certain approach, intersection, or the entire network to be restricted from being phase-sequencing optimized. In addition, some of TRANSYT-7F features can obstruct phases sequencing function such as: grouped nodes, double cycled nodes, un-signalized intersections, and intersections absent from the optimization list (i.e., "run free intersection") [54].

2.5.1 *SYNCHRO*

SYNCHRO is a macroscopic traffic signal timing optimization tool. SYNCHRO is a delay-based program with an objective function that includes also stops minimization. SYNCHRO generates

coordinated signal timing plans for isolated intersection, arterials, and network. SYNCHRO can analyze fully actuated coordinated signal systems by mimicking the operation of a NEMA controller, including permissive periods and force-off points [55]. The program also optimizes multiple cycle lengths and performs coordination analysis. It optimize cycle length by analyzing all cycles in the predefined range, SYNCHRO optimize offset using a multi-stage process. At each stage, it uses a different step sizes depending on the optimization level selected by the user. Unlike TRNASYT-7F, SYNCHRO's traffic model dose not considered platoon dispersion. As an alternative, SYNCHRO determines which intersections should be coordinated and those should run free. The decision process is based on an analysis of each pair of adjacent intersections to determine the "coordinability factor" using travel time, link distance, and traffic volume.

SYNCHRO calculates intersection and approach delays based on the HCM method. The major difference between the HCM and SYNCHRO is the treatment of actuated controllers, as SYNCHRO incorporates a method to model phase gapping and skipping features for actuated and actuated-coordinated signals. SYNCHRO, generate optimal timing plan for each intersection by averaging the analysis results of five volumes (i.e., 10th percentile, 30th percentile, mean, 70th percentile, 90th percentile), the volume entered by user is considered as the mean and variance of the real volume (Poisson distribution). Thus, delay is calculated based on this averaging method (percentile delay method).

In oversaturated conditions, SYNCHRO measures additional delay incurred by the capacity reduction due queues interactions. Queue delay is part of the objective function used in the optimizations. The queue interactions impact is considered in the SYNCHRO's delay calculation by the following procedures:

1) Determine if the volume per cycle to distance ratio is critical
2) Determine the capacity reduction:
    a. Starvation capacity reduction
    b. Spillback capacity reduction
    c. Storage Blocking capacity reduction
3) Determine the additional delay incurred by this capacity reduction.

However, SYNCHRO optimization procedure fails to protect links storage from spillback. Therefore, intersection's blockage cannot be avoided in advance or even accounted in delay calculations later.

2.6 UTC systems

Urban traffic control (UTC) system is a traffic-responsive system provides signal timing that responds to changing in traffic conditions as measured by detectors. The UTC systems are

composed of two major components: hardware and software components. The hardware component includes signal heads, controllers, central computers, detecting devices, and communication lines. Whereas, the software component includes: the control algorithm, arrival model, departure model, and queue and stops estimation models. Chronologically, UTC systems have evolved over the last 40 years, responding to the needs of different cities around the world, the advances in detection, communication, and control technologies. Based on fix-time and real-time control, UTC system have been classified into five control generations (i.e., 1st generation control (1-GC), 1.5 GC, 2-GC, 3-GC, and 4-GC). The first generation control implements pre-defined signal timing plans, while 1.5 GC implements a combination of first and second generation controls features. The 1.5 GC systems are capable of selecting time plans offline or generate time plans online. The 2nd, 3rd, and 4th generations calculate the optimal timing plans dynamically, yet they differ greatly in their optimization interval and response frequencies [56].

Although most UTC systems have similar objectives such as minimizing delay, UTC systems optimization processes differ significantly based on:

- Whether the system is centralized or decentralized, TRNASYT and SCOOT are examples of centralized UTC systems where each signal continuously communicates with the central computer. In contrast, PRODYN and UTOPIA are example of decentralized UTC systems.
- Level of optimization (i.e., single or bi-level)
- Optimization approaches:
    - Mixed integer linear programming
    - Dynamic rolling horizon
    - Forward dynamic programming
    - Hill-climbing technique
- Detectors coverage and deployment in the network (i.e., upstream or stop-line detections)

The following is a list of the main UTC systems that are currently implemented and have been studied and evaluated:

- TR2
- UTCS-1
- UTCS-2
- UTCS-3
- GERTRUDE
- TRUST
- RTOP
- CALIFE
- SCOOT
- SCATS
- RHODES
- PRODYN
- UTOPIA
- CRONOS
- OPAC
- SPPORT
- ALLONS-D
- OTSCS
- MOTION

Gartner (1982) summarized the reasons of the ineffectiveness of many of the UTC systems by:

- Failure to response to short term traffic fluctuation (i.e., robust prediction models)

- Transitions between optimal plans may offset the benefit achieved
- Detectors errors and frequent malfunction

2.6.1 *UTC Operational Strategies*

The UTC strategy has been developed to address traffic-responsive network-wide signal control, particularly under saturated traffic conditions. The aim of the UTC strategy was to provide, at each cycle, traffic-responsive signal settings, taking into account the overall traffic conditions within an urban network [45].

A list of area strategies was compiled from a literature review and from a pilot study interview at Leeds UTC (Gray and Ibbetson, 1991). Operators at each of the ten cities were asked whether they used any of the following strategies:

- Forced and hold green
- Gating and metering
- Maximum capacity flow
- Negative offset- reverse green waves
- Green waves with cross streets
- Flared green with cross street (increases the green time windows at downstream intersections)
- Diversion away from congestion
- Shorter cycle length
- Longer cycle length

Not all strategies were commonly used by the operators that were interviewed. Force and hold greens were only used in special occasions or events (e.g., emergency vehicles). Metering and extended green were identified as the most effective techniques. Metering was predominantly used to reduce the effect of congestion on the network, while extended green was used to recover from congestion. Negative and simultaneous offset were used within predetermined plans for long queue formations. Quinn suggested that the two techniques could be used for improving operations on arterials

2.7 Adaptive Control Algorithms

Most of the techniques reviewed so far were originally designed to be established “off-line” and operated when oversaturated conditions were anticipated during a pre-set time. The challenge of most “off-line” methods is that the traffic volumes that are used to establish the control parameters of a given strategy are seldom realized in the real-world exactly as they were planned. Particularly for oversaturated conditions, if the traffic demand increases to the saturation level before the capacity-maximizing strategies are scheduled to be applied, queues can quickly build up, leading to a serious

delays and ineffectiveness of the algorithms. To avoid such scenarios, some adaptive traffic control systems can detect changes in traffic patterns before the peak period and begin capacity maximization algorithms or strategies earlier than a pre-scheduled approach might be able to. The main impediment to adaptive traffic control systems, however, is the significant cost in deploying detection systems that can supply the necessary traffic data for on-line decision-making. Some of the strategies can be applied to recurrent congestion conditions as is discussed below.

2.7.1 *SCOOT*

SCOOT (**S**plit, **C**ycle, and **O**ffset **O**ptimization **T**echnique) is essentially a TRANSYT traffic model incorporating an optimizer modified for online application [53]. SCOOT determines cycle length, splits and offsets on cycle-by-cycle basis through three independent optimizers. SCOOT seeks to minimize a performance index (PI), which is a linear combination of vehicular delay and stops.

As a response to the limitations of the second and the third generation of UTC systems, SCOOT was developed with features that provide less disruptive transition process between timings plans. The main features of SCOOT control are: no rapid changes in signal timings, there is no need to *predict* arrival profiles, and less sensitivity to detectors failure and malfunction [56].

The SCOOT control system includes algorithms for dynamic control of individual intersections, arterials, and grids/networks [57]. The core algorithms of SCOOT use link flow profile (a composite representation of volume and occupancy) to tune cycle length time, splits, and offset values of each intersection. These algorithms have been proven to reduce delay in light to medium traffic conditions. However, if queuing occurred right up to the exit detector, SCOOT is not capable to model this condition, and would could not detect the stationary vehicles (i.e. no demand) and reduce the green time, in turn increasing the congestion. Since SCOOT Version 2.4 [57] many features have been added to SCOOT to tackle the problem of address severe congestion :

- Normal/typical cycle time tuning
- “Trend saturation” to schedule rapid increases to cycle time
- Gating and action at a distance

Initially, SCOOT has an algorithm that allows trading of split time from one phase to another based on the degree of saturation of each phase. Conventional actuated traffic control handles this condition with fixed or floating force offs. When force-offs “float” the extra time not used by a phase that gaps-out is provided to the coordinated phase on that ring. With fixed force-offs, the next phase in the sequence can use the extra time if it is oversaturated. In SCOOT, if the %SAT is >1 for a phase, if there are any skipped phases in that ring, the extra time can be used by the oversaturated phase if it follows in the sequence or can be rotated in the sequence.

SCOOT makes adjustment to phase's splits cycle-by-cycle by changing green times within a range of few seconds. SCOOT objective herein is to minimize the maximum degree of saturation of intersection's approaches. This mechanism seeks to maintain a balance in queues grow on all oversaturated approaches. Splits' changes are constrained by their deviations from previous values. Each intersection is handled independently by the optimizers. The optimizer is influenced by *user-specified* links tolerable degree of saturation. Thus splits optimizer seeks to relocate the sparse capacities, if exist, into links with lesser tolerances. Regarding transit priority, SCOOT allows phases to be extended or early terminated to accommodate a passage of transit vehicle through an intersection with certain range of degree of saturation (DOS) (i.e., early termination is allowed if DOS < 95% and extension is allowed, if DOS < 110).

Cycle optimizer

SCOOT operates all intersections located in same sub-area with a common cycle or multiples of the cycle length. SCOOT updates sub-areas cycle's time with a few second increments (i.e., 4, 8, 16 sec) at interval not less than 150 seconds. SCOOT's cycle optimizer seeks to achieve a maximum degree of saturation of 90%at the most saturated intersection in the sub-area intersections, subjected to minimum and maximum cycle length constraints. Then, SCOOT's optimizer allows intersections to operate with double cycling, if they can maintain degree of saturation less than 90%. If the degree of saturation of doubled cycled intersections elevate above 90%, the intersections reinstate to single cycle. Once the maximum degree of saturation fails below 90%, SCOOT iteratively decreases cycle time until the degree of saturation of critical intersection in the sub-area returns back to 90%. SCOOT's cycle optimizer may encounter delay surge as cycle time determination is based solely on intersection's degree of saturation.

Table 1: SCCOT's cycle length increments

Cycle length (C)	Increments
C < 64 sec	4 sec
64 sec < C < 120 sec	8 sec
C > 120 sec	16 sec

SCOOT also includes the concept of "trend saturation" to provide more rapid increases to cycle times during congestion. Apparently, a lower threshold value (say, 80% saturation) can be determined to start rising the cycle time earlier at certain times of day when it is known that the traffic will rise rapidly, such as right before the peak periods.

Offset optimizer

Offset adjustment is determined independently for each signal in the network per cycle. The collected cyclic flow profiles are used to evaluate whether to keep the current offsets or to adjust. The adjustment process considers both upstream and downstream links performance measures for each

intersection. Thus only links where both their upstream and downstream signals exist in same SCOOT sub-area are considered in the offset calculation. This would cause a potential performance deficiency in the periphery links. SCOOT also allows each link to have different weight factor in the offset optimization process. Finally, SCOOT evaluates the total performance measures for all adjusted links for the following actions: (a) alters offset point to start earlier (b) keep offset points unchanged, or (c) postpone the offset point few seconds. If the modification is determined to be beneficial (total delay is reduced), SCOOT will adjust the offset a small amount (e.g. +/- 2 seconds) in the next cycle. When congestion rises further, it becomes more difficult to determine which direction of change will be beneficial, so SCOOT allows the user to provide biasing weights to directional flows by time of day. This pre-sets SCOOT to favor certain directions of travel that the user knows to have heavier flow. While this can be effective, it could make situations worse if there are incident conditions or a special event and the flow patterns change dramatically. In addition to the ability to bias the performance of offsets on certain links, the user can also fix or set an offset value in SCOOT. This is an important concept for intersections that are spaced very closely to each other. If the offset is fixed, then SCOOT will not attempt to tune its value to accommodate changes in the traffic flow. SCOOT also includes a number of features to improve progression performance during heavy flows. Initially, SCOOT (and other adaptive approaches including ACSLITE, RHODES/OPAC, and SCATS) tunes offsets to accommodate changes to directional flows in their base algorithms. SCOOT tunes offsets to improve the total delay and stops on each approach to the intersection based on how the link flow profiles will change when small modifications to the offsets are made [53].

SCOOT during oversaturation:

SCOOT handles congestion with several features, such as: congestion importance factors, congestion offset, gating, and variable node-based target saturation

1. **Congestion importance factor:**

Congestion importance factor (CIF) is specified for each link in the network. CIF is used to influence split calculations in favor of a specific link, when congestion is detected. Congestion weighting factor (CWF) allows the user to specify the importance of conducting the *congestion offset*. Congestion offset is a fixed-value offset, specified by the user, to be used in congested conditions to avoid links spillback. Congestion in this case is defined as occupancy over the upstream detector on the link (at the exit point) continuously for several seconds (e.g. 15 seconds). If this condition is satisfied, then SCOOT will adjust the offset to the pre-set “congestion offset” value immediately without going through a series of small adjustments as would be experienced with its normal offset tuning algorithm. This concept is similar to one that is used in many of the situations reported by

practitioners that they have implemented using traditional actuated controllers with detector logic capabilities. The practitioners typically did not modify offsets in this situation but rather increase the green time of the phase that the queue is served by. The combination of biasing, congestion offsets, fixed offsets, and the normal tuning operation allows significant flexibility in the operation of SCOOT for offsets, although these features must be used carefully together.

2. Gating an action at distance

SCOOT controls excessive congestion in network through gating logic. *"The gating logic allows one or more links to be identified as critical, or "bottleneck", links. A critical link can affect the green time on the gated links" [SCOOT user-manual]*.

Similar to the concept of a congestion offset, SCOOT also includes the capability to use the congestion conditions on one link to affect the splits and offsets at other links. This provides the general ability by using logical conditions (if…then rules) to implement gating (holding more traffic on links that have higher storage capacity) or "flared" green, where downstream link green times are increased to move traffic away from a congested area more rapidly. Martin et al., describes capabilities in SCOOT to define both "congestion links" and "trigger links" and it appears that a "congestion link" refers to the continuous occupancy over the upstream detector for a specific number of seconds, and a "trigger link", for the purpose of beginning gating, is defined by the degree of saturation %SAT measure. For the purpose of "congestion" response, it appears that a single if…then condition is used to start the special operation [58]. For gating response, it appears that multiple trigger links are used in a "cluster" to begin a gating operation in a region where multiple gated links have their green times reduced to restrict in-flow and others may have their green times increased to improve out-flow from the area.

3. Variable node-based target saturation

SCOOT Cycle optimizer normally uses 90% as its target saturation level. However, user can set 80% as "Trend Flag" to give more rapid response. If the saturation of the heaviest movements at the intersections exceeds 90%, the cycle optimizer will add 4, 8, or 16 seconds to the cycle depending on the cycle length. In situations where sever congestion is expected, node-based target saturation levels may be set by user (i.e., to be less than 90%). Assigning low *entry* saturation threshold value will result in early increase in cycle length while high *exit* saturation threshold value will allow early drop in cycle length at end of peak period (i.e., recovery period).

As a natural extension of all of these capabilities, SCOOT includes the ability to weight the importance (i.e. the "congestion importance factor") of these if…then conditions, so that triggering special handling (gating, biasing split optimization, or offset modifications) of congestion at one location can be prioritized versus special handling for congestion at other locations. Hysteresis is also

invoked for the gating operation to dampen the switching from gating operation to normal operation. How all of these features are combined together is not well described in available literature, but it seems reasonable that, under certain recurrent congestion scenarios, the combination of tools in SCOOT could be configured to improve throughput substantially over a traffic control strategy where there is no feedback information available or used. No real-world examples of these operations could be found in open literature, but we surmise that these features were implemented over time in reaction to inefficiencies in the operation of the base algorithms of SCOOT in real-world deployments.

Table 2: SCOOT levels of optimization

Level of optimization	Optimizer in operation (0 = no; 1 = yes)			
	Split	Offset	Cycle time	
			Single only	Single or double
0	0	0	0	0
1	1	0	0	0
2	0	1	0	0
3	1	1	0	0
4	1	1	1	0
5	1	1	0	1

Source Hunt et al. (1981)

SCOOT Limitations

- The gradual evolving of timing plans, would keep SCOOT trapped in local-optima in the search space.
- When optimization constraints are relaxed to accommodate fluctuations, SCOOT endure additional computational burden. This renders SCOOT optimizer to be inefficient for more relaxed operations to accommodate rapid fluctuation.
- SCOOT splits optimizer assign splits based solely on degree of saturation may not result in optimal measures of delay and number of stops, particular in under-saturated conditions.
- SCOOT oscillation between single and double cycle is more likely to cause poor transitions between these two regimes.
- Poor performances at boundaries' links of SCOOT's sub-areas may offset the benefits obtained [53].

2.7.2 *SCATS*

Sydney Coordinated Area Traffic System (SCATS) was developed by the road and traffic authority of New South Wales, Australia. SCATS uses traffic responsive techniques to dynamically update vehicle actuated controller settings. SCATS has no specific traffic model to optimize signal timing, thus the aim to maximize throughput or minimize delay and stop may be better define as goals rather than objective [Shelby]. SCATS operates in two modes; the strategic mode and the tactical mode. SCATS divides the network into sub-systems. These sub-systems are considered the nuclei of the

SCATS strategic control structure. Each sub-system includes only one critical intersection that dictates the cycle length, splits and offset calculations. On the other hand, tactical operation mode permits early termination of phases as demand drops or omits phases with no demand [56]. SCATS's detectors are placed at stop bar in each lane of the intersection. Detectors report vehicle counts during green time and occupancy percentage. This information allows SCATS to calculate the degree of saturation and effective used green time.

The SCATS control system operates similarly to SCOOT as it tunes the cycle, splits, and offsets of intersections in a group or section. Open literature (Martin, 2006) identifies that SCATS has far fewer features than SCOOT that are specifically designed for oversaturation management. SCATS fully utilizes the capabilities of vehicle actuated controller such as: phase extension, phase truncation, and phase skipping. These capabilities render SCATS to be the most robust adaptive system to communication failure and perturbations in traffic [53].

SCATS allows re-allocation of split time from one phase to another at an intersection to balance the degree of saturation on all phases. In addition to split re-allocation, if all stages are maxed-out, then the cycle time is allowed to be increased to keep the most saturated phase at or below a threshold value (typically 90%). In each split plan one or more phases can be designed as the "stretch" phase ("stage", in SCATS terminology). As the cycle time is increased up to a specific pre-specified value, any additional increase in the cycle time will only increase the available green time for the stretch phase. This approach uses existing engineering knowledge of field conditions to allow biasing of the additional split time to the phases that are likely to become oversaturated and cause more damaging results than oversaturation at other stages. In addition to incremental changes to the cycle time (similarly to SCOOT), SCATS includes the capability to quickly jump to a higher or lower cycle time using thresholds on the degree of saturation a given key phase.

Offset selection in SCATS is much simpler than in SCOOT. SCATS provides four options for offsets in each traffic section, two-way progression for low traffic, two-way progression for high traffic, and both inbound and outbound offsets for directional flows, such as are typical during AM and PM peak periods on arterials. Pattern matching is used to select which set of offsets is used under any given pattern.

SCATS Limitations

- SCATS does not allow double cycling for less saturated intersection within sub-system
- SCATS is not capable of estimating platoon progression and effectively provides dynamic offsets since detectors are placed at intersections' stop bars. Alternatively, SCATS relies on pre-defined timing plans settings.

- SCATS collects volume data, saturation flow rates, and saturation level on a lane-by-lane basis.
- SCATS's split adjustment process employs greedy method in favor of current phase, if there is more demand in the current phase. SCATS cannot strike a balance between competing phases as they get saturated simultaneously. It relies totally on its calibrated setting. It is arduous task to manage multi-phase split using vehicle-actuated control in oversaturated conditions [52].

2.8 Multi-Objective Control

In under-saturated conditions, delay minimization and bandwidth maximization are the two main strategies that are used to optimize traffic signals in arterial networks. Targeting one or both of these strategies may not fully provide effective control strategies during oversaturated conditions. In oversaturated conditions, queues develop and grow in length, and the total delay increases exponentially as a function of the elapsed time. Congestion can spread out spatially and temporally due to queue spillback and may cause gridlock. Minimizing delay alone is not sufficient to resolve queues and their impacts, most simply because it cannot be directly measured. In order to possibly maintain optimal states of traffic during over-saturated conditions or transitions between unsaturated and over-saturated conditions, some control strategies have been developed in the past and reformulated based on an integrated criterion that combined delay minimization and queue management through a multi-objective analysis framework.

Lan et al. [59] proposed the COMBAND model to simultaneously minimize delay and maximize progression bandwidth on arterial networks to deal with the conflict between delay minimization and provision of bandwidth. The objective function is to optimize a linear, weighted combination of the delay/stop and bandwidth, subject to maximum queue constraints.

Anderson et al. [60] presented a Multi-Objective Genetic Algorithm (MOGA) for the signal control problem. The aim of their study was to optimize the signal controller off-line with respect to a number of diverse criteria, including reductions of specific vehicle emissions, pedestrians' waiting time, vehicle stops, and vehicle delay. Since the users (vehicles and pedestrian) are competing against each other's for a share of the resource (green time), a value referred as urgency is derived from the raw traffic data using fuzzy logic. This measure of urgency is then assigned for each respective user, which reflects its entitlement on the limited resource (green time). While the MOGA studies are intriguing, the researchers do not include any actual implementation results and queue management criteria are not considered.

Chang-Jen, et al. (2003) adapted multi-criteria decision making (MCDM) methods, including deviation minimization and compromise programming, to develop "compromise" signal control

strategies and investigate the system performance of signalized intersections under various criteria. Their method is claimed to be capable of generating effective timing solutions fairly close to Pareto optimality for a given objective function. However, to achieve robustness of the control strategies, the stochastic process in which vehicular arrivals and queues dynamics are generated are not adequately considered.

Abbas, et al. [61] proposed a use of multi-objective evolutionary programming to examine the effectiveness of alternate strategies for timing oversaturated intersections. The approach allows the optimization of several objectives simultaneously. Unlike traditional methods of assigning pre-defined weights to each objective function, multi-objective evolutionary algorithm produces the Pareto Front of all objectives at the same time. This approach allows the analyst to explore the relationship between different objectives (Delay, stops, throughput, queue length, cycle failures, etc.) and identify the time and locations where a shift from delay-minimization to throughput-maximization may be necessary to alleviate.

2.9 Origin-Destination Matrix Estimation Methods

The origin-destination (O-D) trip table represents the number of travel trips conducted between various O-D zone pairs located within a road network. The trip table plays an essential role in urban planning and traffic operations. Conventionally, trip table have been created through various methods such as home interview surveys, road side interviews, license plate surveys, etc... These traditional techniques are expensive and time consuming. In addition, the rapid changes in land-use development cause O-D table to be obsolete in fast paste. The O-D table estimation problem can be broadly classified based on network configuration and route choice assumptions to the following categories:

- Simple network with no route choice (Arterial network)
- Networks with route choice but no congestion
- Network with both route choice and congestion

The need for obtaining O-D matrix through inexpensive and robust methods emerged in the early 1970s as it was identified that trip tables could possibly be obtained through measurement of traffic volumes on network links using point detection systems that provide link counts. In essence, the existing methods for synthesizing O-D tables from link counts fall into the broad category of one of two types of approaches, namely, *parameter calibration techniques* and *matrix estimation methods*. *Parameter calibration* approaches utilize gravity model to construct the O-D tables that match links flows and zonal socioeconomic characteristics. This method is considered impractical in operational level since it requires updated zonal data to calibrate the parameters used in demand [62, 63] [63]. On the other hand, *matrix estimation techniques* rely only on link traffic counts and prior information in the form of a *seed* trip table. Matrix estimation techniques are therefore easier to implement and they

are more applicable in operational level. A detailed discussion of matrix estimation techniques is provided in the next section [64].

- **Matrix Estimation Techniques**

Matrix estimation techniques can be categorized into two main categories:

1. Maximum entropy/minimum-information theory techniques, and
2. Network equilibrium techniques

Maximum entropy/ minimum-information theory-based techniques: This technique aims to find the relationship between traffic counts collected at discrete points of time on different links and the O-D demand that most likely matches the count observations. The maximum entropy/minimum-information technique aims to calculate the probability of a portion of traffic flow on a link being attributed to a specific O-D pair. This technique estimates the O-D matrix that has the maximum likelihood of producing the observed traffic counts. Given the large number of the expecting possible solutions for O-D pairs, prior information of O-D volumes (a "seed" matrix) is indispensable to improve the reliability of the estimation process. An O-D matrix for a network can be obtained from previous studies or from practitioners' opinion and judgment. The main critique of this method is that the seed matrix is merely used as an initial guess/solution with as little as possible information used from this matrix. In an attempt to address this critique, several researchers have employed Bayesian inference techniques such as in (Maher 1983) or using least squares estimation models such as in [65-69].These methods assign weights to the a priori trip matrix data before incorporating them with the link traffic counts in the maximum likelihood formulation. O-D pairs that have more accurate information (e.g., O-D pairs that are monitored by cameras or some number-plate recognition) would be assigned larger weights during the aggregation process. Van Aerde et al. proved that gravity trip distribution model could be considered a subset of the maximum likelihood solution to the synthetic O-D estimation procedures. They present a procedure for solving the static O-D problem that only requires the Stirling's approximation and does not require flow continuity in observed links flows. This proposed procedure was incorporated in the QUEENSOD software [70]. However, with networks with both route choice and congestion, the proportional assignment is no longer valid because route choice proportion and the O-D table are interdependent [71, 72]. In congestion, it is necessary to incorporate route choice model in the O-D estimation problem.

Network equilibrium-based techniques: The network equilibrium approach is based on the Wardrop principles in traffic equilibrium [73]. This approach is more applicable to dynamic traffic assignment when network congestion effects are prominent. Nguyen et al., and LeBlanc et al., emphasized using a non-proportional assignment to determine the effects of congestion [62, 74]. They account for link's capacity and the non-linear relationships between links and O-D flows. LeBlanc et al., proposed to

formulate the O-D estimation problem as a bi-level problem, where the upper level is the minimization of the Euclidean distance between the solution matrix and the target matrix, and the lower level is the Wardrop user-equilibrium assignment. Cascetta [75] used the Generalized Least Squared method to combine target trip matrix, model prediction and traffic counts within a single framework. (Yang et al. 1992) integrated the generalized least squares technique with an equilibrium traffic assignment approach using a Stackelberg leader-follower bi-level optimization model to account for the uncertainties in both the target O-D matrix and in the traffic link counts. They proposed a heuristic solution method to overcome the inherent difficulty in solving moderate to large sized problems of this type. Sherali et al., [64] proposed a model that also employs the non-proportional-assignment assumption in order to find user equilibrium solution that reproduces the observed link flows. The model recognizes the driver's incomplete information in their selection of routes.

One of the major limitations of the network equilibrium approach is the need for continuity in traffic counts. In addition, the traffic dynamics have to be accounted for in the O-D matrix. For example, a 15-minutes static traffic matrix cannot be used if the travel time from the origin to the destination is, say, 20 minutes. In conclusion, the maximum likelihood approach aims to find a solution that is closest to the priori O-D seed matrix, assuming a proportional assignment. The network equilibrium approach might address the dynamic nature of traffic better, especially in large networks, but successful implementations of this approach are still in development and will not be used in this task.

2.10 Wavelet Analysis

Wavelet is a powerful mathematical tool for data de-noising and signal processing. This tool has proven it capacity in many fields of science and engineering applications such as data mining, systems identification, image processing, and pattern recognition [76]. Wavelet is a wavy function carefully constructed so as to have a certain mathematics properties. A mother wavelet provides a building block function that can be used to describe any class of functions. The wavelet transform utilizes its basic function, the mother wavelet, then dilated and translates it to capture signal features that are local in time and local in frequency. This in particular allows wavelet analysis to provide an insight into the dynamics of time series data beyond that of the orthodox methodologies.

Previous applications of wavelet transform in traffic engineering focuses on traffic flow analysis and incidents detection in freeways. Several researchers applied wavelet transform to raw data targeting noise elimination/ or reduction while keeping the underlying patterns such that the aggregated data using shorter intervals are more predictable. Samant et al., [77] proposed a discrete wavelet transforms (DWT) and linear discriminate analysis (LDA) algorithm to detect traffic incident

features. In the proposed algorithm, DWT was applied to raw traffic data to get the finest resolution coefficients representing the random fluctuations of traffic, and then LDA was employed to the filtered signal for further feature extraction and reducing the dimensionality of the problem. The results of LDA were used as input to a neural network model for traffic incident detection. Further researchers added traffic pattern recognition method to this model [78]. In this model, wavelet analysis was used to de-noise, cluster, and enhance the raw traffic data first, and then the results were classified by a radial basis function neural network.

3 Research Methodology and Organization

3.1 Introduction

To complete the research goals and achieve the associated objectives, a series of tasks were performed. A brief description of each of the research tasks is presented in the following section. The research approach consists of the following stages:

- Thoroughly literature review
- Theoretical development for a timing framework
- Simulation modeling
- Evaluation of the proposed control strategies

The following section presents some of the research methodologies and techniques used in conducting the dissertation tasks. The employed methods include wavelet filtering, k-mean clustering, multi-objective analysis, allocation problem, random number generation, and simulation.

3.2 Wavelet Filtering

Wavelet is a powerful mathematical tool for data de-noising and signal processing. Wavelet filter provides an easy vehicle to study multi-resolution properties of a process, by giving up some frequency resolution. This transform decomposes a process into different time horizons as it differentiates seasonality, reveals structure brake, dissolving the correlation, and volatility cluster in non-stationary or transit time series. The proposed method can be considered as alternative to the orthodox O-D estimation methods. During the peak periods certain movements in the network are considered significant based on their location (i.e., entry and exit points) and flow fluctuation. The proposed method considered the observed detectors' counts profiles resulting from the peak demand and signal operation as non-stationary time series. Wavelet analysis was used in the study to detect the resemblance of these times series of critical movements in their higher resolutions. Then, movements were cluster via K-mean clustering process to establish network's critical routes.

3.3 Simulation

Simulation is defined as a "logical-mathematical representation of a concept, system, or operation programmed for solution on a high speed electronic computer" (Martin, [79]). Drew [80] defined simulation as "a working analogy that involves the construction of a working model presenting similarity of properties or relationships to the real problem under study". In this research, the performance of existing and proposed control strategies was evaluated through simulation. Simulation facilitates system evaluation under wide variety of traffic conditions and control operation modes.

Among the various microscopic simulation models available, VISSIM was used to analyze the existing operation condition as well as the proposed control strategies. VISSIM is a microscopic

behavior-based traffic simulation model capable of analyzing complex traffic operations under both under-saturated and over-saturated conditions. Considering the local behavior of drivers car following settings in VISSIM were calibrated until the VISSIM model generated travel times and queue lengths comparable to those observed during data collection. Several car following parameters, including desired speed and vehicle headway were adjusted in the VISSIM model with the goal of calibrating the VISSIM travel times with the observed field travel times. In additional, VISSIM's signal control modular allows the use of several commercial signal controllers either by means of the built-in control with the VISSIG add-on or with the help of an optional external signal state generator (e.g. optional module VAP, external controller DLL). VISSIM can also be controlled externally through a serial interface to the following controllers: NEMA TS/2, Econolite ASC/3, SCATS, and SCOOT.

3.4 Multi-objective evaluation

In most real-life optimization problems, the utility functions are often multidimensional. Consequently, there is no unique optimal solution but rather a set of efficient solutions, also known as Pareto solutions, characterized by the fact that their cost can't be improved further in one dimension without being worsened in another. The set of all Pareto solutions, the Pareto front, represents the problem trade-offs, and being able to sample this set in a representative manner is a very useful aid in decision making. One approach consists of defining an aggregate one-dimensional utility function by assigned a weighted sum of the all utilities. As a result, each set of weight coefficients will lead to an optimal solution for the one-dimensional problem which is also a Pareto solution for the original problem.

In this research, a multi-objectives approaches to examine the effectiveness of proposed control strategies for oversaturation timing. Unlike traditional methods of assigning pre-defined weights to each objective function, the proposed multi-objective produces Pareto front of all objectives. The shape of the produced Pareto front provides invaluable information to the system analyst and decision makers regarding system performance.

3.5 Dissertation Organization

The proposed methodology presented above implicates accomplishing the following tasks:

3.5.1 *Task1: Literature review*

Conduct detailed and comprehensive literature review about the following topics:

1. Traditional signal control objectives
2. Oversaturation definition and diagnosis
3. Origin-destination estimation problem
4. Wavelet filtering analysis
5. Oversaturation control strategies

3.5.2 *Task 2: Critical routes determination and scenarios development*

In this task, the proposed method investigates the potential of advanced signal processing techniques in determining network critical routes for control purpose can be considered as alternative to the orthodox O-D estimation methods. Wavelet domain processing is used to decompose, de-noise, compress, and extract the common pattern in a set of traffic flow time series collected from several detectors in signalized sub-network in Reston Parkway in Northern Virginia. Results show that several matched patterns in movements can be detected. The obtained route's pattern matches the field observation. The proposed method can identify the critical route under the congested conditions without rely on traditional O-D estimation method

3.5.3 *Task 3: Development of oversaturation control strategies*

In this task, a comprehensive timing framework is developed for oversaturated arterial networks. The framework consists of three major components: control strategy development, timing optimization, and timing plans evaluation. Control strategies are developed based on traffic pattern, and congestion attribute, for the potential problematic scenarios in the network. Several control strategies are pointed out from the extensive literature review conducted previously and from expert opinions. The timing plans were generated according to pre-assigned control strategies and oversaturation principles for cycle length, split, and offset. Finally, the obtained timing plans are evaluated via simulation and multi-objective analysis (i.e., Pareto front).

3.5.4 *Task 4: Data collection and Simulation modeling*

Two networks were modeled using VISSIM micro-simulation software; Reston Parkway network in northern Virginia and Post Oak arterial network in Houston, Texas. The detectors logs and current timing plans from Reston Parkway network were obtained from the North Region Operation (NRO) of VDOT. Similarly, City of Houston provided the information needed to model Post Oak grid network. The proposed control strategies were evaluated using VISSIM with respect to the base case scenarios.

4 Operational of Signal System in Oversaturated conditions

4.1 Introduction

In oversaturated conditions, typical traffic control strategies do not work as efficiently as requires, particularly since the control objectives need to be decided differently when movements are constrained. Unlike under-saturated conditions, operation of traffic signal systems during oversaturation conditions needs consideration and analysis of the underlying causes of the congestion before developing mitigation strategies. This research is investigating the following hypotheses:

- For oversaturated network, considering *different routing scenarios* in designing timing plan will result in adopting different control strategies that need to be evaluated using different performance measures.
- During oversaturated conditions, protecting *critical links/routes* from becoming saturated will reduce the temporal and spatial spread-out of congestion. In addition, it would protect the network from gridlock and secondary congestion symptoms
- During oversaturated conditions, achieving optimality in signal operation requires take into account *system dynamics* in strategies selection, control objectives, and performance measures. Three system states must be considered in timing design: *loading, processing, and recovery*.
- In oversaturated condition, *multi-stage optimization*, *volume spillover accommodation*, and *switch between control objectives* are expected to improve overall network performance.

4.2 Congestion definitions

From the literature and experts views, an oversaturated network can be described according to the following attributes:

- Spatial extent
- Temporal extent
- Recurrence
- Cause(s)
- Symptoms

Table 3: Congestion Attributes Taxonomy

Extent	Duration	Causation	Recurrence	Symptoms
Movement	Situational	Signal Timing	Recurrent	Starvation
Approach	Intermittent	Geometrics	non-Recurrent	Spillback
Intersection	Persistent	Other modes		Storage Blocking
Route	Prolonged	Demand		Cross Blocking
One-way arterial		Planned Events		
Two-way arterial		Unplanned Events		
Interchange				
Grid				
Network				

4.2.1 *Problematic Scenarios*

During oversaturated condition, the following problematic scenarios are expected to take place at intersection level due to high demand and inappropriate traffic signal control. These scenarios represent the fundamental building blocks of an oversaturated situation:

- Spillback
- Starvation
- Residual queues
- Storage blocking

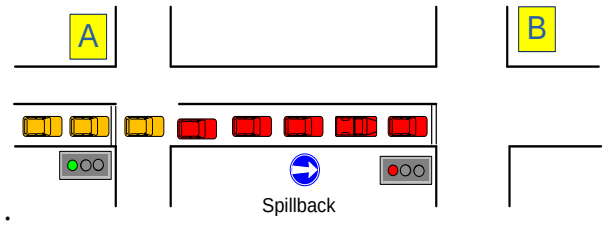


Figure 5: Approach spillback (de facto red)

Spillback occurs when a queue from a downstream intersection uses up all the space on a link and prevents vehicles from entering the upstream link on green. Some literature has defined this condition as causing "de facto red" to the upstream movement since no progression is possible. This is illustrated in Figure 10. Some traffic control strategies can reduce the frequency of spillback occurrence such as negative offsets (i.e., reverse progression), dynamic offset adjustment, or "flared" green times. Operating closely-spaced intersections as one intersection (e.g. like an interchange) can also be an effective strategy to mitigate spillback.

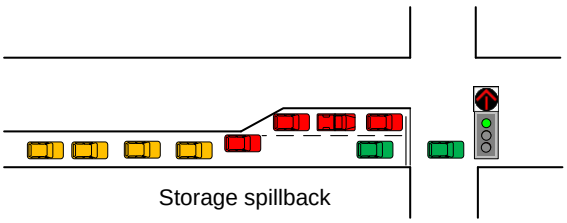


Figure 6: Storage bay spillback

Storage bay spillback, shown in Figure **6**, occurs when turning traffic use up the entire space of the storage lane and blocks the through traffic. The delayed through movement then experiences a starvation. Signalized control treatments include phase truncation, phase re-service, and lead/lag phasing adjustments could be the appropriate strategies at the local level to mitigate storage bay spillback.

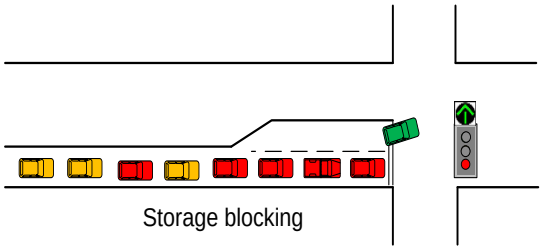


Figure 7: Storage blocking

Turning storage blocking, shown in Figure 7, occurs when queues extend beyond the *opening* of the storage bay. Through traffic thus blocks access to the left or right storage bays. In this situation, the turning movement will experience starvation since the turn signal is green, but the vehicles that intend to turn left are blocked from reaching the turn bay. Phase sequence adjustments and phase truncation might be considered to mitigate storage bay blocking.

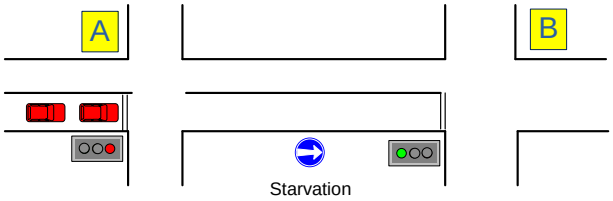


Figure 8: intersection starvation

Starvation occurs when a phase is green, but the phase cannot service at full capacity efficiently due to storage blocking, spillback blocking, or perhaps because the upstream signal is red. These conditions are illustrated in Figure 8. Starvation might be avoided by applying phase truncation strategies, phase sequence changes, and dynamic offset adjustments.

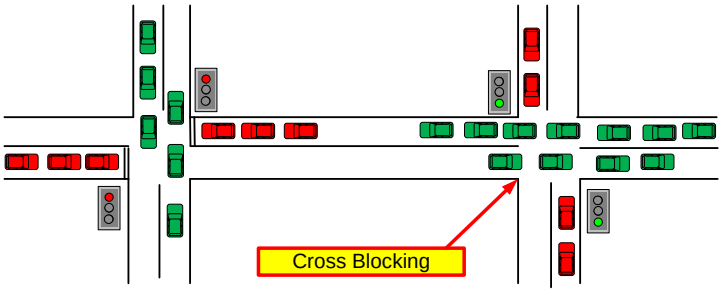


Figure 9: Cross-Blocking effects

Cross intersection blocking, illustrates in Figure **9**, occurs when queues on an incompatible service phase extend into an intersection blocking the progression of vehicles. While many jurisdictions have "*don't block the box*" laws or policies, these types of situations are recurring in grids and networks with short link lengths. Carefully controlled settings of green times and signal offsets are necessary to mitigate these types of situations.

Identification of these *symptoms* of oversaturated conditions is an important component of the identification of appropriate mitigations.

4.3 High-Level System Dynamics

From a high-level perspective, the daily traffic and recurrent events have repeatable patterns. No traffic system is perpetually in oversaturated operation, but rather evolves in three phases of operation:

- Loading phase
- Oversaturated operation (or, "processing Phase")
- Recovery phase

This basic concept of the oversaturation period is illustrated in Figure **10**.

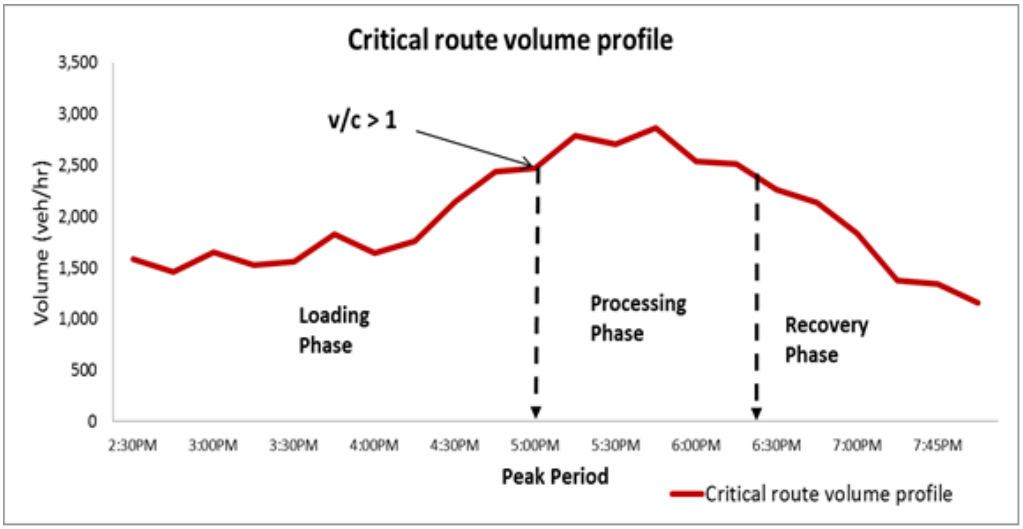


Figure 10: Loading, Oversaturation, and Recovery Phases of Operation

From a high-level perspective, recurrent oversaturation has a repeatable pattern. No traffic system is perpetually in oversaturated operation, but rather evolves in three regimes of operation: loading regime, oversaturated operation (or, "processing Regime"), and recovery regime. During the ***loading*** regime, the network volumes are increasing, and routes volumes' proportions are changing. Inthe case of non-recurrent events, the *triggering event* has occurred. During loading, residual queuing and other symptoms such as storage blocking and starvation begin to emerge. During this regime, switching operational control objective from traditional control of *delay minimizing* to *throughput maximizing* can offer a measurable improvement in performance on the approaches and routes that will shortly become oversaturated. Early application of mitigations is justifiable when the causal effects are recurrent. During the ***Processing*** regime (i.e., oversaturation), both traffic volumes and routes proportions are such that queues and congestion symptoms are not going to be dissipated until either the traffic volumes are reduced, routes proportions are changed (i.e. drivers' avoid the area, adjust their routes, decide to travel later, etc.), or both. Mitigations applied during this regime are expected to protect network from primary congestion symptoms such as intersections blockage, *de facto* red, and gridlock. In the ***recovery*** regime, either of traffic volumes and/or route proportions and/or limited downstream capacity has been adjusted so that the residual queues begin to dissipate. In this regime of operation, certain mitigation strategies can be considered especially effective in returning the system to its steady state faster than continuing to apply the "normal" operational strategies. Considering these regimes of operation in applying or evaluating the performance of mitigating strategies can provide an improvement in performance over time by switching from one control objective to another and therefore, switching their corresponding mitigation strategies.

4.4 A Catalog of Mitigation Strategies

As can be deduced from the literature and expert opinions, multiple strategies are used to address oversaturated problematic scenarios. The oversaturation mitigation strategies are classified into intersection-level, arterial-level, and wide area-level. The can also be grouped based on their objective functionality into: *green utilization* strategies or *queue management* strategies.

The following is a detailed review of different mitigation strategies that are appropriate for oversaturated network.

4.4.1 *Local strategies*

The ***green extension (green flush)*** strategy is used to increase approaches' capacities by forcing the signal to dwell in green while serving the congested approach for much longer than is typically provided in the coordination split. Green flush is often implemented manually, yet an automated mechanism can be implemented by time of day or by a certain threshold of queue formation. Minor objectives would be to keep oversaturated turning bays clear, and to alleviate congestion on some

major side streets. The objective of this strategy is to allow the queue formation on the portions of the roadway with the most favorable storage. Both manual and adaptive flush mechanisms are palliative. Their aim is to reallocate queue development to the least damaging locations. Neither throughput maximization nor residual queues alleviation is anticipated by applying a flush strategy[81].

Phase re-service can be considered when certain movements' volumes exceed its both phase green and storage capacities causing spillback. This strategy allows the heavy movements to be served twice in a single cycle, or to alternate movements to every other cycle, without extending the cycle length and without reducing intersection throughput. The more heavily imbalanced the flows on each movement are at an intersection, the more likely the intersection will benefit from serving the major movement more often in the cycle, for shorter periods of green time during each service.

Dynamic left-turn strategy is applied to prevent an intersection from becoming oversaturated by temporary prohibiting left-turning during high demand periods (i.e., phase reduction). This strategy is implemented extensively in CBD areas during peak periods where pedestrian flow constitutes an additional encumbrance to turning movements.

Lead/lag left turns strategy is implemented to provide more green time for the saturated through movement especially when an intersection has unbalanced turning movement volumes. Lead/lag phase adjustments prevent storage spillback and left-turn phase starvation, respectively. Special consideration to the left-turners' arrival patterns is recommended when designing and applying a lead/lag phase sequence.

Phase truncation is exercised in instances of lack of demand to prevent starvation. A phase can also be terminated if the recipient link is full and/or about to spill back. The truncation of the phase can help to avoid intersection blockage (i.e., the de facto red). The literature pointed out that phase-truncation can also be used as a demand management mechanism to divert left-turners to other intersections by reducing the left-turn capacity [81].

Closely-spaced intersections are preferred to be run together on one controller. Such operation will offer a high level of coordination eliminating spillback and starvation. Therefore, the strategy increases system throughput in addition to managing the queues on the links with limited storage space. Diamond interchanges are the most common examples of locations where this strategy is applied. One case was mentioned by an expert of running three closely-spaced intersections using one controller. One expert offered a rule-of-thumb that if two intersections are less than seven seconds of travel time apart, then the intersections should be run on one controller.

Adaptive and responsive control methods are considered to be robust in delaying the onset of congestion by reducing the development of residual queues in the first place. Adaptive and responsive control approaches also provide faster time-to-response when the demand increases rapidly.

4.4.2 *Arterial Strategies*

The ***offset strategies (i.e., negative progression and simultaneous offsets)*** are used to manage oversaturated arterial networks. Progression strategies maximize the throughput by assuming that vehicles can move down the street unimpeded by residual downstream queues. If the residual queues constitute an obstacle to the progression, negative offsets can be used. The negative offsets strategy has been recognized as being useful for handling long queues, but such a control strategy was considered difficult to implement unless included in a specific pre-determined signal timing plan. Most of the experts noted that they rely on field manipulations rather than using optimization tools to come up with "optimal" timing for negative progression. Most of the expert practitioners advised to use coordination strategies in heavy flow conditions to progress flow not more than "five or six" intersections.

Flaring the green (increasing the green windows at downstream intersections) is considered when residual queues exist in downstream intersections due to excess in turning movements from side streets. Extending the green time for congested movements (***Green flush***) is practiced to increase capacity under the assumption that saturation flow rates remain constant. This assumption is questionable due to the effect of turning traffic on the flow rate. The green time is added to congested movements that create a longer cycle length without adding green time to the other approaches. The extended greens strategy is not frequently used. However, it is useful to recover from severe congestion when handling special events (e.g., protecting a freeway from traffic back-ups on the off-ramp at an interchange). This strategy can have the adverse impact of shifting the bottleneck to the next downstream intersections.

Switching from negative progression to green flush can also be considered as a queue management strategy that requires moving from a short cycle length to a longer one. Queues would be expected to form in the minor approaches.

4.4.3 *Network level strategies*

Metering is an area-wide strategy for queue management. The implementation of this strategy requires the identification of the critical intersections of a network and the links that will be used to store the queues. Metering is applied to impede traffic at appropriate upstream points to prevent the demand from reaching critical levels at downstream intersections. While metering is considered an effective mechanism to alleviate "grid lock" and prevent detrimental effects of queues, experts do not

offer any systematic methodology for allocating the metering points since it must be handled on a case-by-case basis.

Simultaneous offsets and fixed timings are often used in grid network, due to the lack of detection in many CBD areas. Some experts indicated that simultaneous offsets strategy in grids might be implemented in three to five intersections, with different sets of simultaneous offsets applied in alternative fashion (e.g., zero offsets alternating with 65% offsets).

Adaptive and responsive methods can be used to manage grid networks. While adaptive methods can be effective in preventing and responding to congested conditions, the difficulty of setting the system parameters and cost of installation and maintenance remains the main concern with adaptive system deployments.

Cycle length adjustment: The common practice of increasing cycle length to increase intersection throughput is based on the very basic assumption that is saturation flow rate remains constant during green time. This assumption needs more investigation, since many experts reported a significant decrease in stop-line flow rate after the first 30-sec of green. These observations were backed by several recent studies[82]. The recent FHWA report (Signal Timing under Saturated Conditions) pointed to two possible causal mechanisms that can explain the reduction in saturation flow rate. One is the turning vehicles that increase vehicles' gaps and therefore, reduce through flow. The other reason is related to drivers' inability to respond to the green light due to their position in the queue. These assumptions need to be further investigated[81].

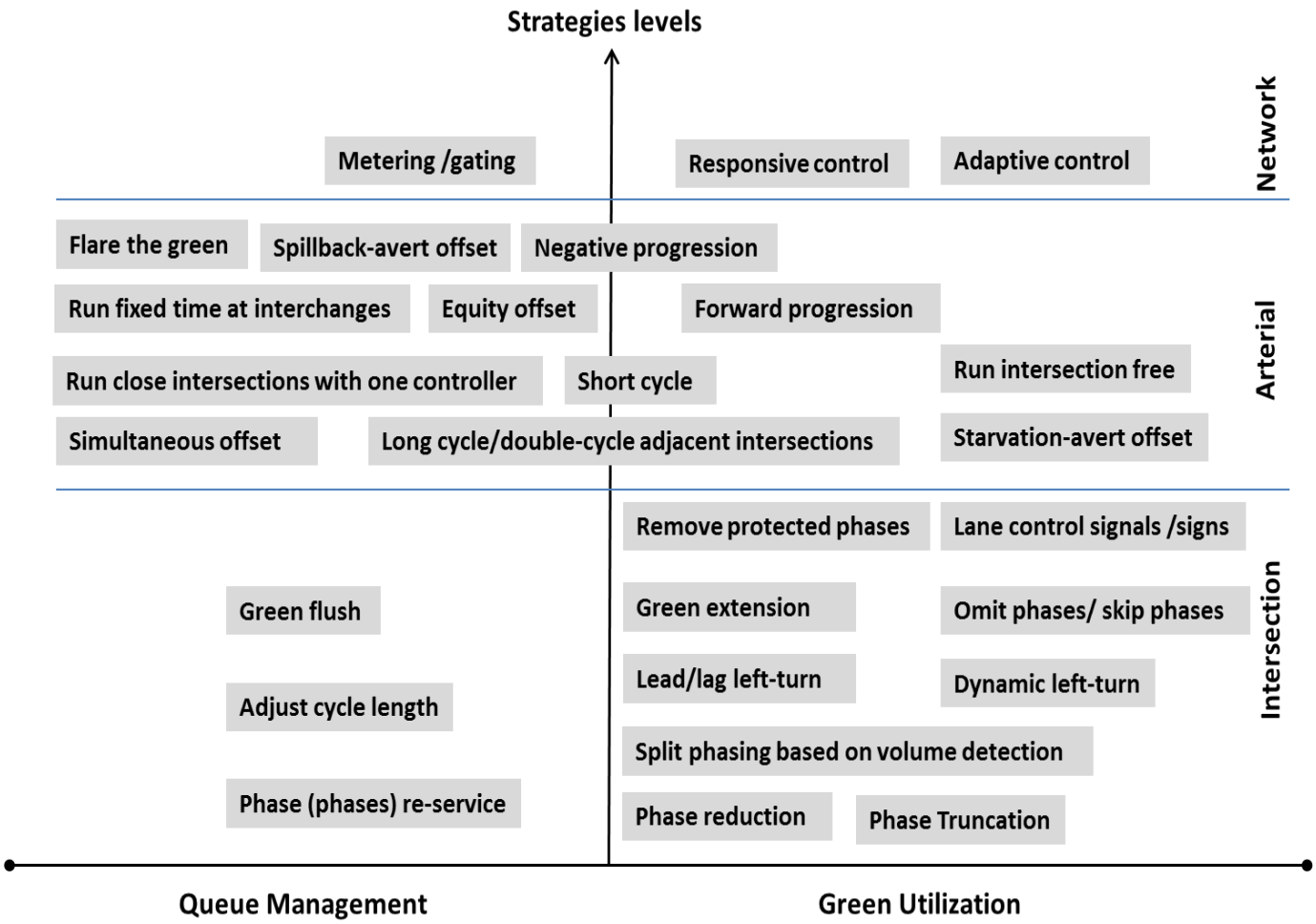


Figure 11: Control strategies operational objectives

4.5 Summary of control strategies

These control strategies can be classified based on the level of its application as follows:

- **Intersection-level strategies**
  - Green extension (green flush)
  - Phase truncation
  - Phase (phases) re-service
  - Dynamic left-turn
  - Lead/lag left-turn
  - Run closely-spaced intersections with one controller
  - Change the barrier structure for certain plans by time of day
    - Omit phases/ skip phases
    - Phase reduction

        - Combine movements, remove protected phases when not needed
    - Time detector extension times appropriately to provide “snappy” operation
    - Use split phasing based on volume detection
    - Allow pedestrian times to exceed cycle time for infrequent pedestrian operation on side streets
    - Use lane control signals or signs to provide flexible capacity
- **Arterial-level strategies**
    - Offset strategies
        - Forward progression
        - Negative progression
        - Simultaneous offset
        - Spillback-avert
        - Starvation-avert
        - Equity offset
    - Flare the green
    - Green flush
    - Switching from negative progression to green flush
    - Adjust cycle length
        - Decrease the cycle time when more than one phase is oversaturated. More opportunities to service the queue because there is reduced efficiency of saturation flow rate in long queues.
        - Increase the cycle time at the oversaturated intersection and double-cycle other adjacent intersections
    - Run closely spaced intersections on one controller, like an interchange to avoid storage of vehicles on short links.
    - Run intersection free
    - Run fixed time at interchanges – early return to green at other locations can actually harm progression
    - Run fixed time, short cycles during construction
- **Network-level strategies**
    - Metering
    - Gating
    - Adaptive and responsive control systems

4.6 Proposed timing framework during oversaturated conditions

The following is a description of a proposed approach for developing and evaluating operational strategies for oversaturated arterial network. The proposed steps materialize from a combination of practitioners' expertise and critical reviews of the available literature. This process can be summarized as follows:

1. **Identify network critical routes:** this step is necessary to understand the traffic patterns on the network during the oversaturated period. Changes in network critical routes will generally result in considering different control objective and therefore, applying different control strategies.
2. **Identify routing scenarios on the network:** different combinations of critical routes will produce different problematic traffic scenarios at network intersections.
3. **Map control strategies to routing scenarios**: not all control strategies are suitable to all problematic routing scenarios. This step narrows down the list of control strategies that can be used for a particular scenario.
4. **Generate optimal timing plans**: in this step, the principals of oversaturation timing will be used to generate optimal timing plans
5. **Evaluate the optimal control plans**: the optimal oversaturated control plans will be evaluated considering multiple objectives performance criteria

In the next chapter, a proposed method investigates the potential of advanced signal processing techniques in determining network critical routes as alternative to the orthodox O-D estimation methods. The principles of the developed oversaturation model are presented first. These principles include cycle length selection, green splits allocation, and offset design during oversaturated condition, followed by detailed description to the calculation procedure and optimization formulation. Finally, the two oversaturated networks were used to investigate the research hypotheses and to evaluate the developed oversaturation timing framework.

5 Critical Routes Estimation Using Wavelet Filtering

5.1 Introduction

Identifying network critical O-Ds is a necessary step to figure out network traffic patterns and therefore identify the potential problematic scenarios. The literature mentions that priority paths or routes should be considered when designing optimal control strategies (Hurdle, 1992). However, it should be noted that a route might become "critical" during the peak periods of the day and not be as "critical" during off-peak periods. This is mainly due to the variation of characteristics of O-D traffic that pass through this route during different times of the day. The orthodox O-D estimation methods are costly, labor-intensive, and time-consuming. In addition, the dynamics of traffic control (i.e., route travel time, delay, etc...) renders the static O-D methods to be ineffective especially during the congestion periods. The proposed method presented in this section utilizes wavelet filtering technique to estimate the network's critical routes. During the peak periods certain movements in the network are considered as to be *significant* based on their locations in the network and their flows fluctuation. The proposed method deemed the count profiles of these *significant* movements as time series and then applied wavelet filtering to capture the similarity between the movements. Basically, wavelet transformation (WT) uses a basic function that is stretched and shifted to capture feature that are local in time and local in frequency. By combining several combinations of shifting and stretching of the basic function, the wavelet transform is able to capture information in a time series and associate with specific time horizons and locations in time. A detailed description about (WT) is provided in the following section [83].

5.2 Background

Wavelet filter provides an easy vehicle to study multi-resolution properties of a process, by giving up some frequency resolution. This transformation decomposes a process into different time horizons as it differentiates seasonality, reveals structure brake, dissolving the correlation, and identifies volatility clusters in non-stationary or transit time series. WT is effectively in representing functions with non-constant (overtime) frequency contents, this due to the Wavelet capacity to localize in time the characteristics of the frequency behavior of a function [84]. Wavelet analysis facilitates using long time intervals where more precise low-frequency information is needed and shorter intervals where high-frequency information is required. This in particular, allows wavelet analysis to reveal many aspects of data that other signal analysis techniques is lacking--aspects like trends and discontinuities in higher derivatives [84]. Signals carry overwhelming amount of data in which relevant information is often difficult to find. However, the time series profile of significant movements can be decomposed by elementary WT into several levels of processing signal.

Wavelet transforms are classified into discrete wavelet transforms (DWTs) and continuous wavelet transforms (CWTs). Both DWT and CWT are continuous-time (analog) transforms. They can be used to represent continuous-time signals. CWTs operate over every possible scale and translation whereas DWTs use a specific subset of scale and translation values or representation grid. It can be considered that both CWT and DWT can be used to analyze the traffic flow series. If the traffic flow *f(t)* is considered to be a *square-integrable* function of a continuous time variable *t*, its CWT is defined as[84]

$$W_f(a,b) = \int_{-\infty}^{\infty} f(t)\psi_{a,b}(t)dt$$

and the two-dimensional wavelet expansion functions (wavelets) $\Psi_{a,b}(t)$ are obtained from the basic function (also known as "mother" or generating wavelet) $\Psi(t)$ by simple scaling and translation

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}}\psi\left(\frac{t-b}{a}\right), a, b \in R, \Psi \in L^2(R)$$

where the parameters $a \neq 0$ and b denote the frequency (or scale) and the time (or space) location, respectively, and R is the set of real numbers. The notation $L^2(R)$ represents the square-summable time-series space of the traffic flow $f(t)$, where the superscript 2 denotes the square of the modulus of the function.

Several families of wavelets have been devolved in the recent years. The first wavelet filter is developed by Haar (1910). Albeit, Haar wavelet filter is easy to visualize and implement, it is inadequate for most real-world applications. Stromberg (1982) developed a family of wavelet. Meyer (1985) introduced the system known as the Meyer basis. Battle (1987) and Lemarie (1988) each independently proposed the same new family of orthogonal wavelets. Ingrid Daubachies introduced a family of wavelet that are not only orthogonal, but also have compact support. Wavelets can be categorized on the following basis: support of function (i.e. compactly or infinite supported), smoothness (i.e. Differentiable or non-differentiable), shape of function (i.e. symmetric or asymmetric), function spaces (i.e. orthogonal, orthonormal, and bi-orthogonal), and based on their input and output (i.e. continues or discrete) [85].

5.2.1 *WAVELET'S APPROXIMATIONS AND DETAILS*

In wavelet analysis, filtering process is the most basic level of WT. herein; the signal is broken into two entities: approximations and details. The approximations are the high-scale, low-frequency components of the signal. On the other hand, the details are the low-scale, high-frequency components [86].. The signal passes through two complementary filters and emerges as two signals as it is shown in Figure 12.

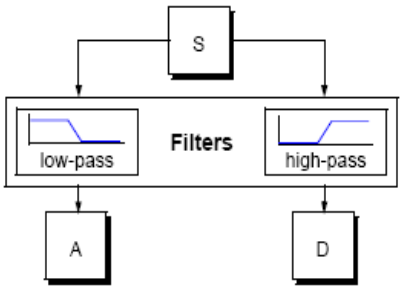


Figure 12: Signal filtering process (Source: MATLAB Wavelet Toolbox™ 4 User's Guide), used under fair use, 2012

The signal decomposition process can be iterated with successive approximations (A) being decomposed in turn, so that one signal is broken down into many lower resolution components. This is called the *wavelet decomposition tree* as Figure 13 illustrates

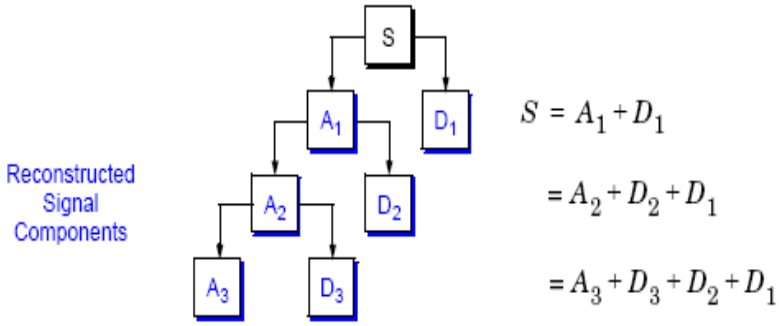


Figure 13: Wavelet Multi level decompositions (Source: MATLAB Wavelet Toolbox™ 4 User's Guide), used under fair use, 2012

5.2.2 *WAVELET COEFFICIENTS*

The signal decomposition process using WT produces coefficients at each decomposition level as detail coefficients (cD) and approximation coefficients (cA). The cDs are small and consist mainly of a high-frequency noise, while the cAs contain much less noise than the original signal does. These coefficients sets describe the signal when the process is reversed. When the separate Wavelet levels are added together the original signal is regained. The shapes of the components of the reconstructed signal depend on the shape of the decomposition analysis.

5.2.3 *DAUBECHIES WAVELET*

Ingrid Daubechies invented what are called compactly supported, asymmetric, orthonormal wavelets — thus making discrete wavelet analysis practical. The names of the Daubechies family wavelets are written dbN, where N is the order, and db the "surname" of the wavelet [87].

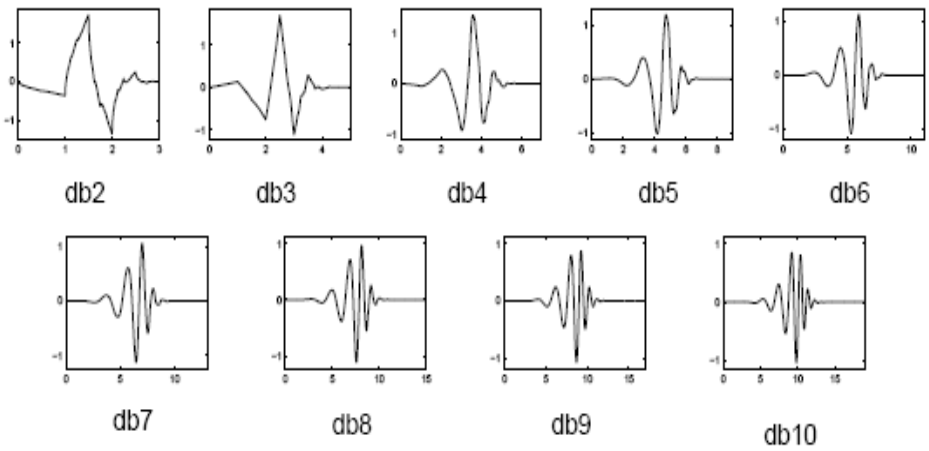


Figure 14: Daubechies' compactly support family scaling function wavelet set (Source: MATLAB Wavelet Toolbox 4 User's Guide), used under fair use, 2012

5.3 METHODOLOGY (WAVELET ANALYSIS AND TIME PLAN)

5.3.1 *Proposed Wavelet Filtering Framework*

The proposed approach is based on three analysis steps; each one of them affects the others significantly. This section describes the sequence of these three steps as well as the main purpose of each one of them.

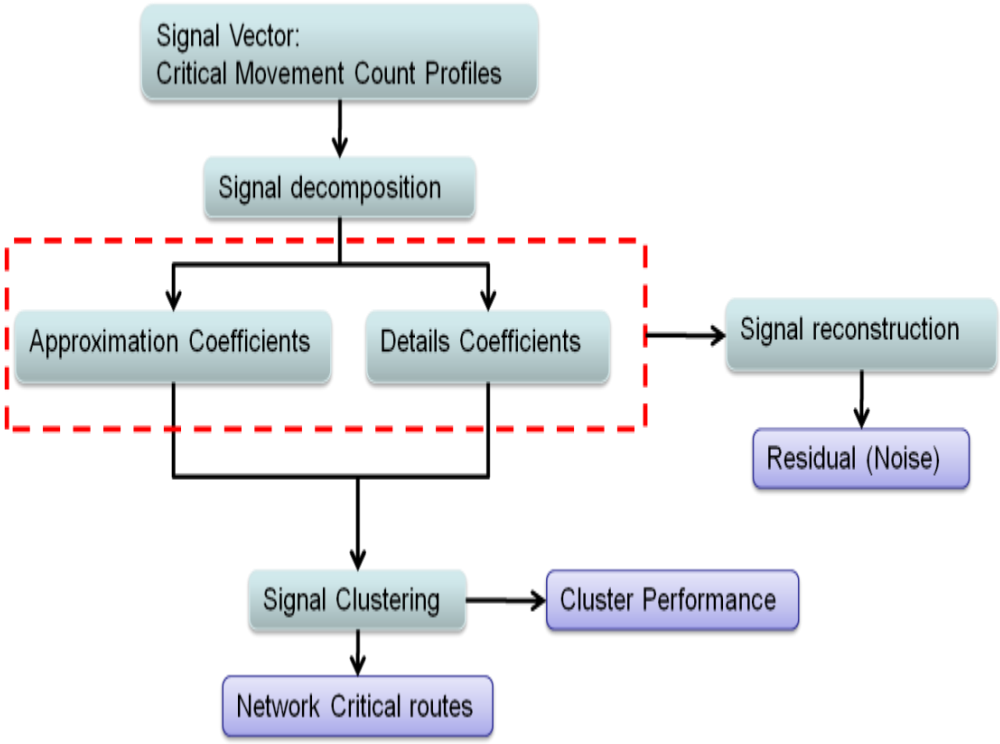


Figure 15: The critical routes identification framework

Step 1: Count Profiles Decomposition

In this step, the count profiles of the significant movements are decomposed using *Daubechies* wavelet filter. Daubechies is compactly support family scaling function wavelet set with nine members as it was described previously. The level of approximation is determined in advance based on the signal magnitude and level of fluctuations.

Step 2: Signal Reconstruction

In this step, the signal is reconstructed back using the original approximation coefficients of the level of previous decomposition for a certain percentage of recovery. The reconstruction process yields a de-noised version of the signal and its corresponding residual. Additionally, in this step the corresponding density of nonzero elements is computed. As for each signal, the percentage of required coefficients to recover a certain percentage of the energy is determined. This illustrates the spectacular capacity of wavelets to concentrate signal energy in few coefficients.

Step 3: Signal Clustering

Signals are clustered based on their approximation and details coefficients generated from multi-level decomposition. K-means uses an iterative algorithm that minimizes the sum distance from each object to its cluster centroid. The algorithm keeps moving object between clusters until the total distance is minimized. The number of clusters is provided to the cluster function as initial input, and then the cluttering process is repeated for a new number of clusters until the optimal silhouette value is achieved. The silhouette value determines how good the clustering of the profiles given the number of cluster is.

5.4 Critical Route Volume Determination

The objective of the proposed heuristic described below in Figure 16 is to assist practitioners to develop feasible set of critical route's volume profiles. These profiles are necessary to determine the appropriate signal control strategies later.

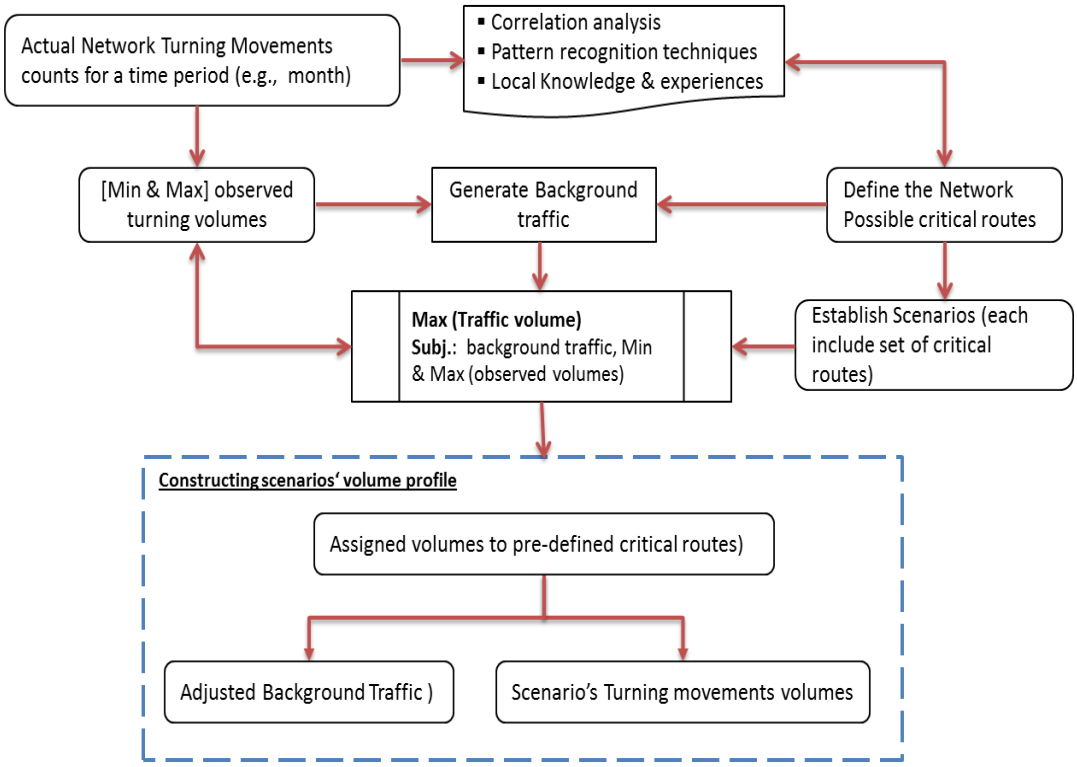


Figure 16: Critical Routes Volume Estimation Framework

a. Data collection

For a given network, movement's counts need to be collected for a long period (e.g., weeks, or months) to reflect traffic variability. The variability in traffic volumes is characterized by the maximum, minimum and average observed volumes in the analysis period. The objective of the step is to establish a range of volumes that movement's counts can vary within.

b. Background volume determination

Network background traffic is composed of non-critical routes' traffic plus the stable fraction of the critical routes' traffic. Basically, background traffic is the least traffic volume observed during the analysis period. The rationale behind extract background traffic is to find the least volume for the signal operation without considering critical routes interactions.

c. Building routing scenarios

In this step, a set of critical routes are combined together to establish a routing scenario. The routing scenarios selection process should consider the possible traffic patterns that reflect the field observations, as incorrect traffic pattern might lead to inaccurate selection of control strategies. However, in case of large networks, considering all combinations of traffic routes and their volumes' levels is impractical. Nevertheless, routes scenarios used to design signal control strategies can be representative not comprehensive.

d. Determining scenarios' volume profiles

After determine the network background volumes, routes volume are set to the maximum levels (i.e. max observed counts) with consideration to other routes that shared same link/links. An optimization procedure is applied to generate feasible solution (i.e., volume profiles of critical routes during the peak period) for each routing scenario.

e. Background traffic adjustment

The background traffic needs to be adjusted again after the critical routes volumes are determined. The objective of this adjustment is to account for the temporal variation of routes volume. This adjustment is crucial, since critical route volumes is intent to be the driving factor the control strategies development.

5.4.1 *Experiment and Analysis*

The Reston Parkway network is located in Northern Virginia, USA. The network consists of fourteen intersections. A major freeway (Dulles Toll Road) passes through the middle of the arterial. Consequently, there is a large variation in the O-D traffic from and to the freeway, during different times of the day. This situation might warrant the division of the Reston Parkway arterial network into two sub-networks north and south of Dulles Toll Road interchange. The orthodox O-D estimation methods, if not used carefully, will result in a biased estimation toward ramps' traffic that defies the field observations. Actual detector data was obtained from Northern Region Operation (NRO) of Virginia Department of Transportation VDOT. The traffic volume was taken for a period of one month starting from May 5th, 2009 to June 6th, 2009. These detectors cover most of the network and have a record every 15 minutes of traffic volume, occupancy, and average speed. According to the field observations during the evening peak-period, the movements are shown in Figure 17 can easily identify as significant movements:

1. Right-turn Movement from Sunset Hill Rd. into Reston Parkway southbound
2. Left-turn Movement from Sunset Hill Rd. into Reston Parkway southbound
3. Southbound through movement at Sunset Hill Rd. intersection
4. Left-turn from Dulles westbound ramp into Reston Parkway southbound

5. Movement exiting westbound at Dulles toll way westbound on-ramp
6. Movement exiting eastbound at Dulles toll way eastbound on-ramp
7. Through movement on Reston Parkway south of the Dulles toll way interchange

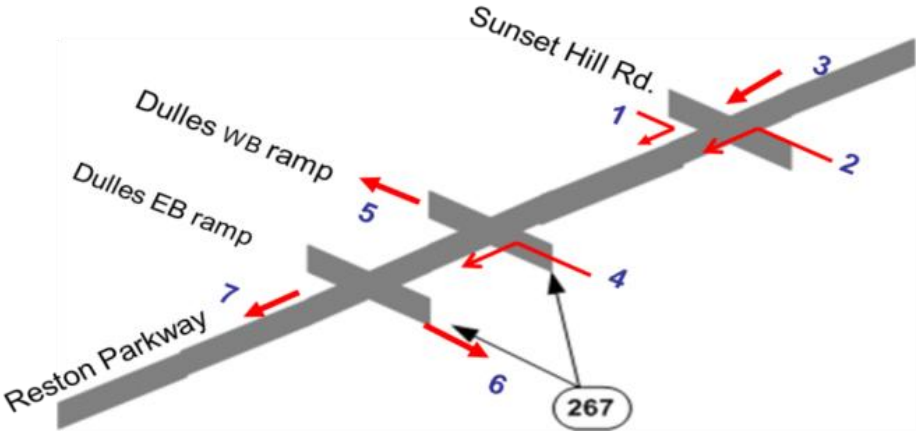


Figure 17: Critical movements on Reston Parkway network in Northern Virginia

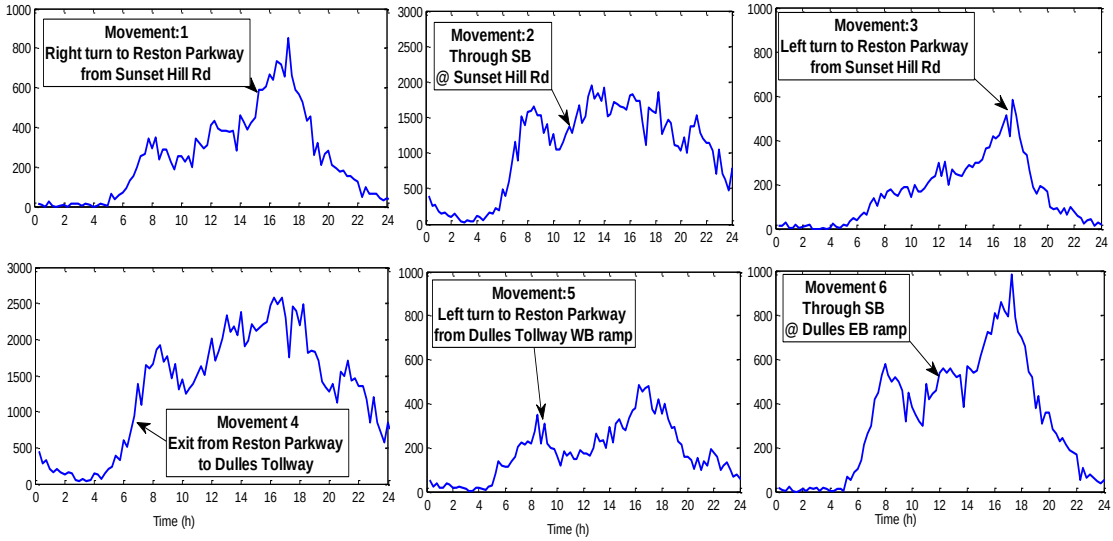


Figure 18: Critical movements count profiles on Reston Parkway network

The 24-hour detectors data of network's significant movements are illustrated in Figure 18. The turning movements count's profiles are considered as signal (time series) in the proposed analysis framework.

5.4.2 *Wavelet Analysis Results*

The length of the line is relative to the magnitude of the coefficients at the same level. The coefficients are spaced at each level to represent their localization in time. Such plot provides good description of where significant change is taking place in the function. The location of the abrupt jumps in the blocky function can be spotted by looking for vertical cluster of relatively large coefficients.

Table 4: Network critical movements level 2 detail coefficients (D2)

Critical movements	Wavelet D2 coefficients										
	1	2	3	4	5	6	7	8	9	10	11
Movement 1	0.09	-0.23	-4.38	2.65	12.16	12.33	5.75	-15.99	-3.53	-6.48	11.57
Movement 2	-1.67	-9.03	-1.94	-2.69	10.35	4.45	12.24	-11.05	13.30	9.03	-2.95
Movement 3	1.76	-1.67	-25.94	10.03	33.87	-9.82	19.94	-17.98	5.31	4.19	7.44
Movement 4	-3.35	-18.24	-4.14	-5.11	1.04	11.54	8.95	1.63	-2.05	4.41	7.67
Movement 5	-5.73	-22.94	-5.22	-22.60	5.10	1.22	-2.87	-4.39	3.49	7.84	6.38
Movement 6	1.73	2.12	-2.73	-1.00	3.29	2.01	12.60	2.73	-0.83	6.39	7.05
Movement 7	-19.61	-75.05	-38.26	17.70	34.98	-23.05	17.94	-21.37	15.70	24.04	10.85

Table 5: Network critical movements level 3 detail coefficients (D3)

Critical movements	Wavelet D2 coefficients										
	1	2	3	4	5	6	7	8	9	10	11
Movement 1	0.47	0.21	-4.38	5.39	-6.17	13.96	-2.35	-3.65	3.30	8.65	2.58
Movement 2	2.80	4.84	-1.96	15.92	5.62	-3.25	10.21	-8.66	3.48	4.16	9.63
Movement 3	-19.79	-64.14	18.24	50.48	-11.86	-26.28	14.51	-0.01	1.40	0.38	6.42
Movement 4	-0.65	-2.04	2.12	0.75	9.06	-6.70	-1.09	-2.29	1.22	4.84	0.56
Movement 5	2.44	9.81	2.82	21.59	-15.74	4.18	-1.61	-29.52	16.14	27.91	-7.05
Movement 6	-0.75	-1.91	-0.29	-1.50	-0.25	-2.54	2.26	-3.25	1.03	4.15	3.06
Movement 7	-19.66	-65.04	22.74	50.44	-7.40	-19.73	10.93	-0.65	1.38	1.94	-0.06

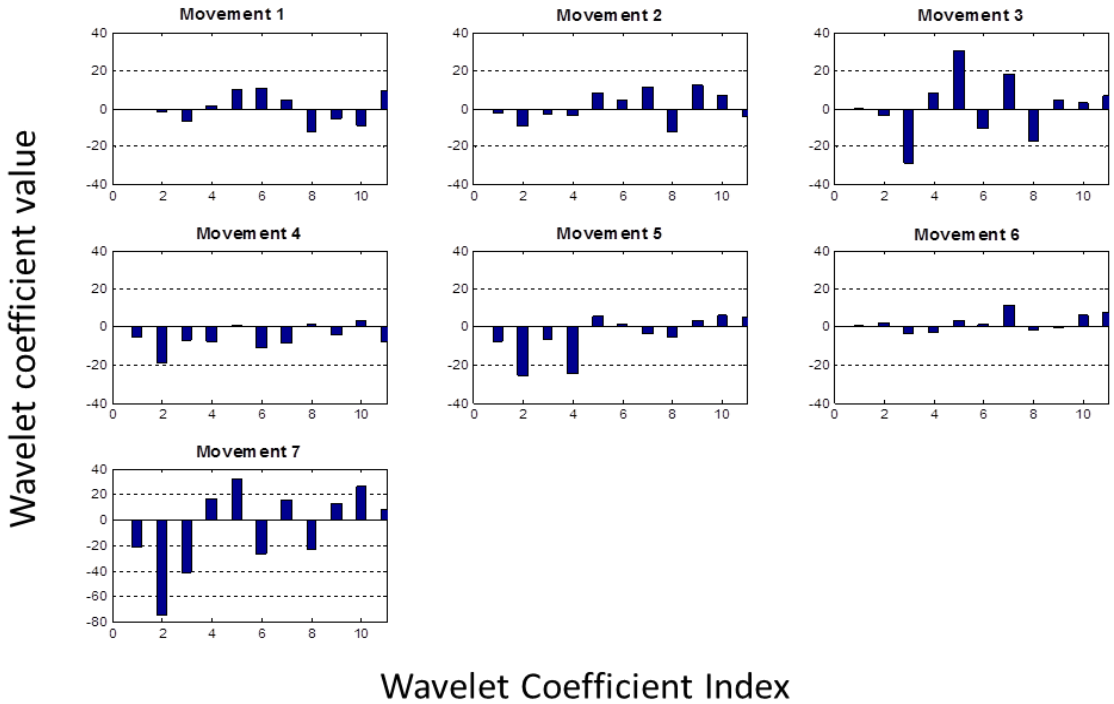


Figure 19: Critical Movements' Detail 2 Coefficients

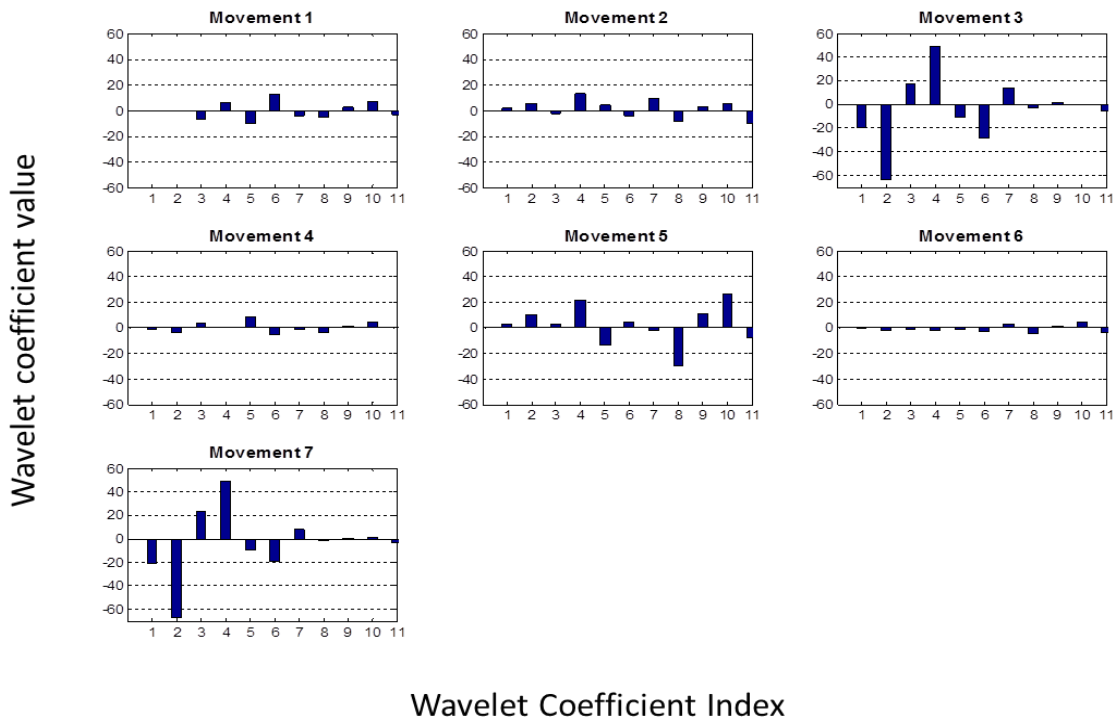


Figure 20: Critical Movements' Detail 3 Coefficients

The preliminary analysis of the decompose signal indicate that detailed level 3 is the most appropriate level considering the fluctuation nature of traffic pattern in this sub-network, unlike detailed level 1 and 2. The noise level produced at each level of decomposition is the factor that determines the appropriate level to be selected. Level 3 is high frequency decomposition of the given volume's that is suitable to capture traffic variation within 15-minute bins. The k-mean clustering algorithm grouped the following movements (3, 7), (1, 5, 2, 6) using both details' and approximations' coefficients. This grouping facilitates in establishing the network's critical routes A, B, and C, respectively as it shown in the following figures. Those findings confirm the field observations (e.g., the heavy turns from Sunset Hill road exit the network at Dulles Toll-way ramps).

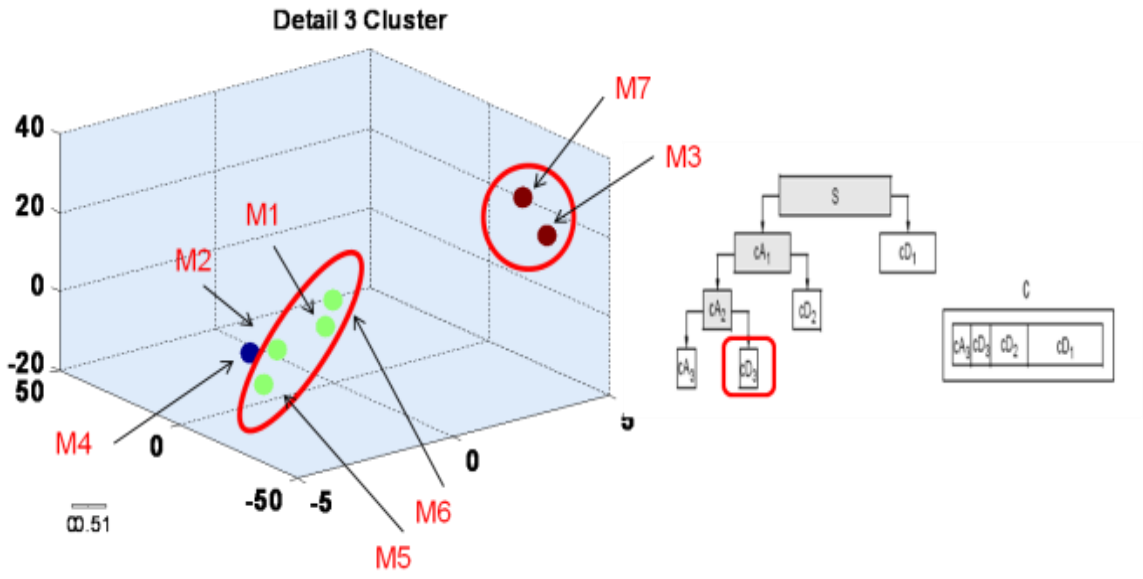


Figure 21: K-means clustering for movements' detail 3 coefficients

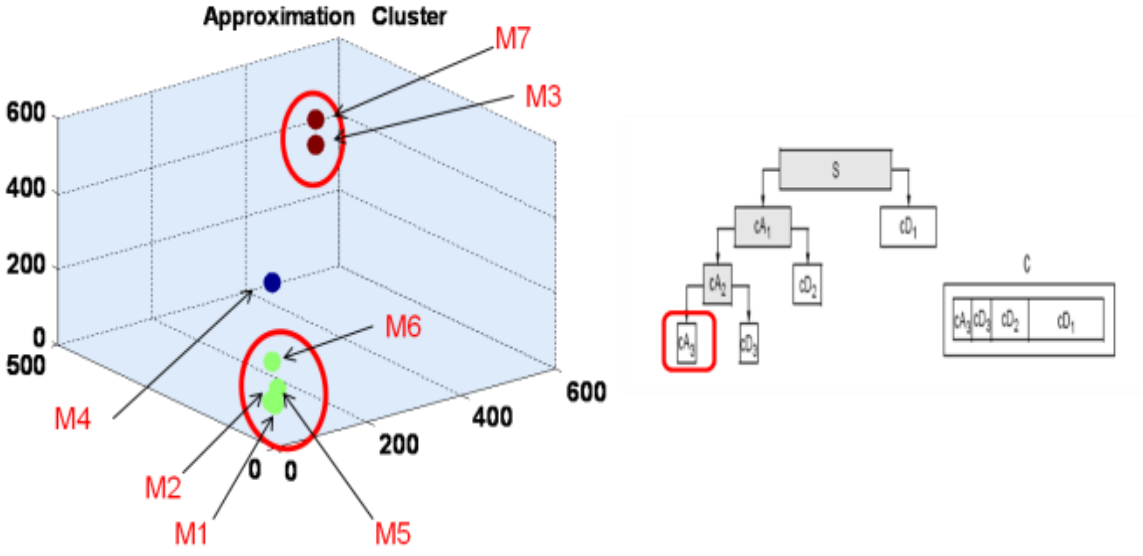
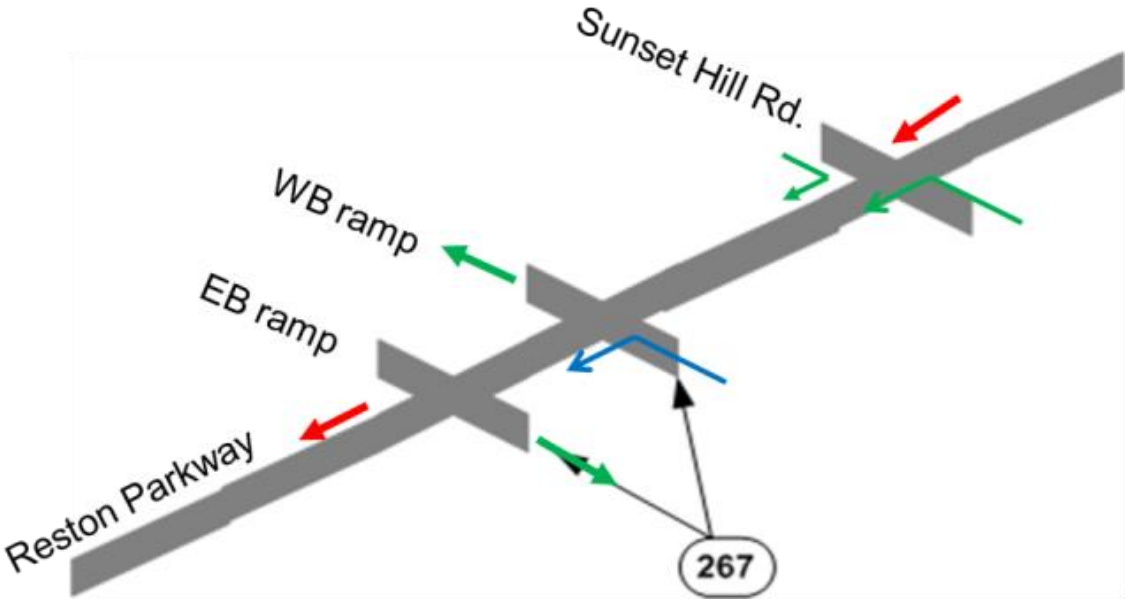


Figure 22: K-means clustering for movements' approximation coefficients



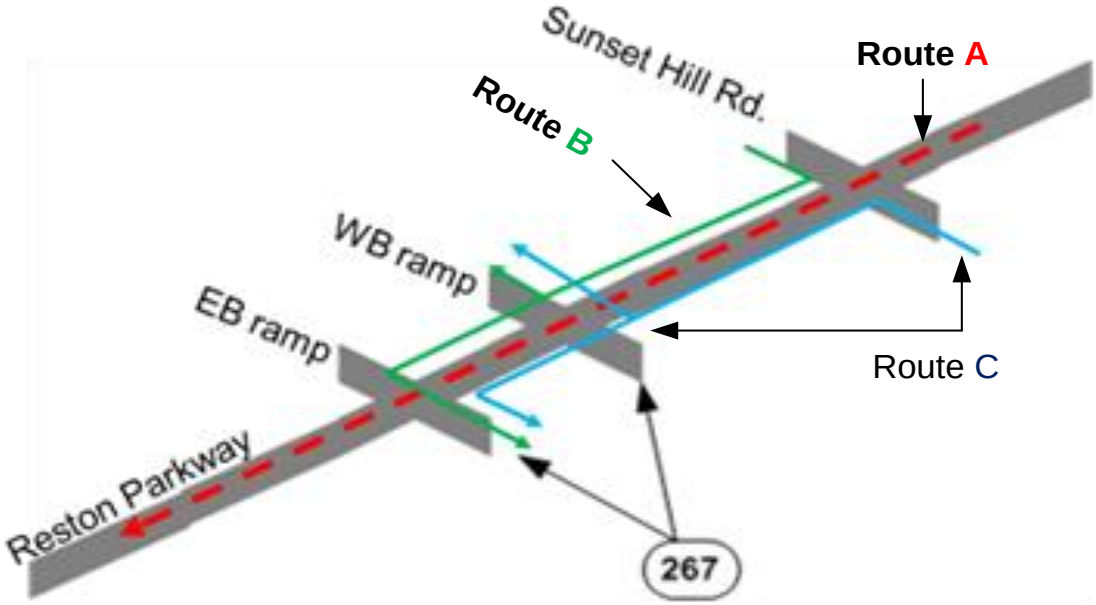


Figure 23: Correlated critical movements and the corresponding routes

5.4.3 *Critical routes volume profile*

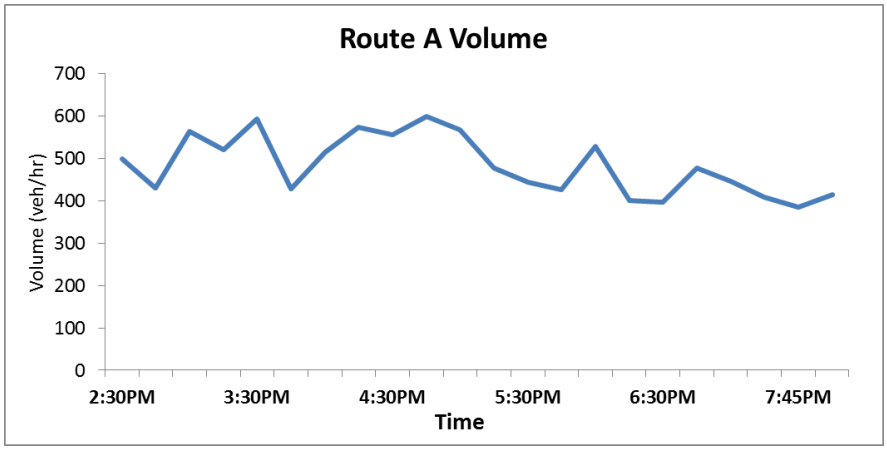


Figure 24 : Route (A) Volume during P.M. Peak Period

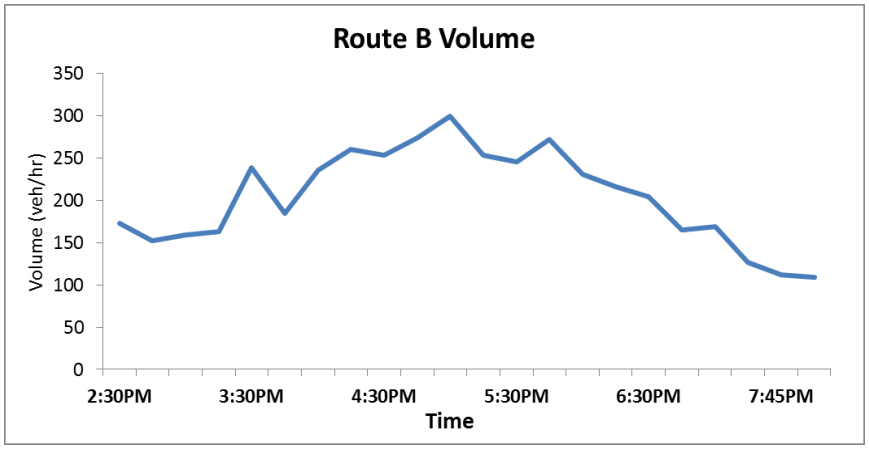


Figure 25: Route (B) Volume during P.M. Peak Period

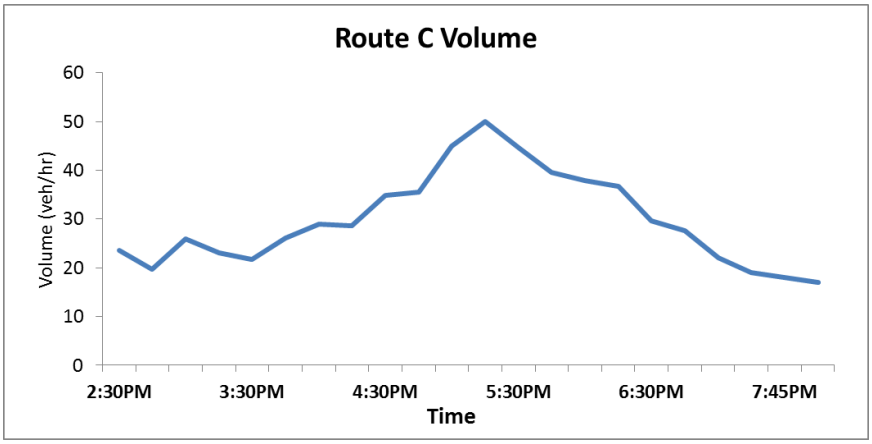


Figure 26: Route (C) Volume during P.M. Peak Period

5.4.3.1 *Summary and Conclusion*

The proposed method investigated the potential of advanced signal processing techniques in determining network critical routes. By performing wavelet decomposition to observed peak hours counts profiles, the individual factors that make up the time series are evolving at different scales and different magnitudes. This process, wavelet filtering, revels the structural breaks and identifies volatility within networks significant movements. The counts profiles were filtered using Daubechies wavelets at detailed level 3 of decompositions that found to be the best representation of these time series. The obtained routes pattern matches the field observation during peak period. Thus proposed method can identify the critical route under the congested conditions without rely on traditional O-D estimation methods.

6 Timing Framework for Oversaturated Arterial Network

6.1 Introduction

This section describes in details the principles of signal timing during oversaturated conditions. The proposed approach steps were developed from a combination of practitioners' expertise and a critical review of the available literature. First, it is critical to be aware of the traffic patterns in the network before designing "optimal" control strategies. This approach would narrow down the possible scenarios that need to be considered in developing the optimal timing plans. The timing plan components (i.e., cycle length, split, and offset) are expected to perform efficiently during oversaturated conditions since they are generated based on oversaturation principles and they account for critical routes flows. The process of the proposed approach can be summarized as follows:

- Identify network critical routes/movements
- Map control strategies to the identified oversaturation problematic scenarios
- Cycle length determination
- Green splits calculation
- Oversaturation offsets design

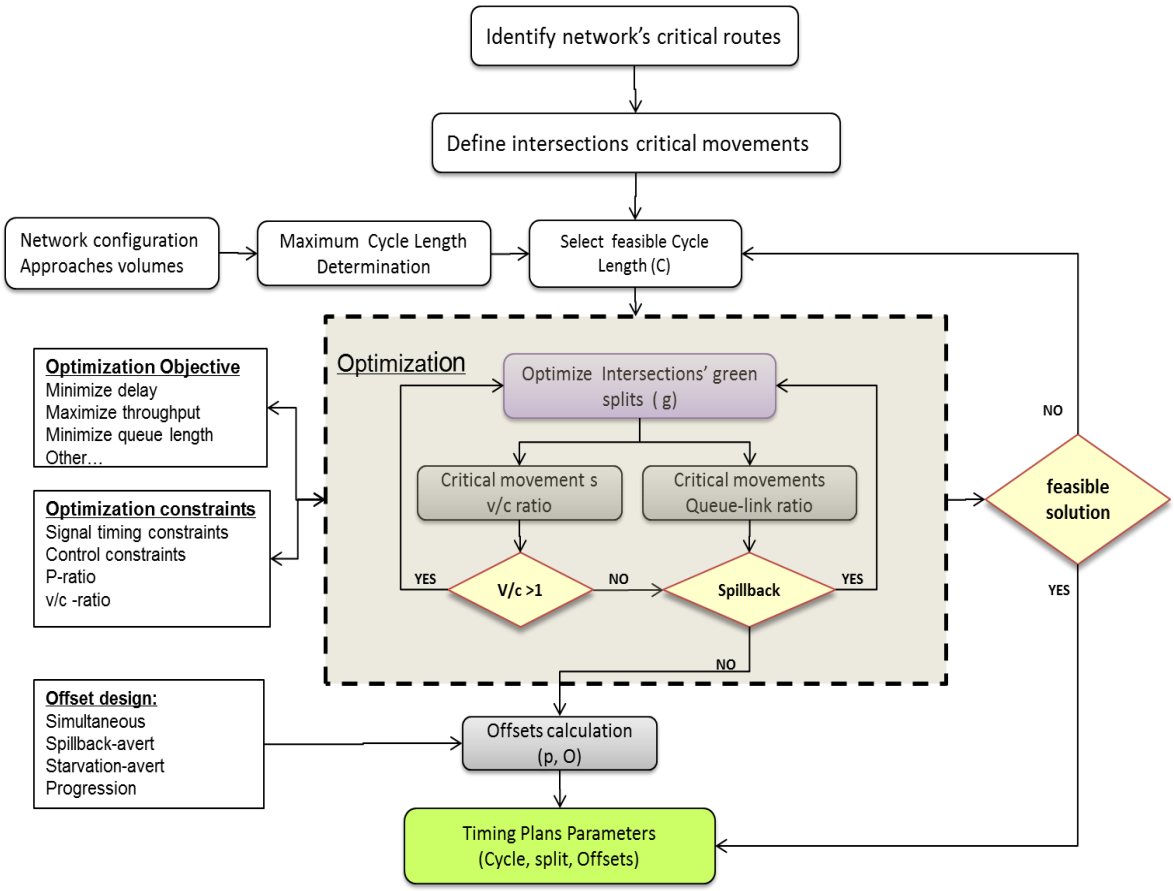


Figure 27: Timing Framework for oversaturated conditions

6.2 Cycle Length Determination

The determination of arterial background cycle length is essential in signal timing design during oversaturated conditions. The failure in determining the appropriate cycle length will increase the possibility of recurrent spillback and cause intersection blockage. In typical under-saturated conditions, the optimal cycle length is determined based on conflicting approaches degree of saturation (Y). Webster equation (i.e., minimum-delay cycle length) calculates the minimum cycle based on total loss time (L) and intersection degree of saturation (Y) as shown in the following equation.

$$C_m = \frac{1.5L+5}{1-Y} \qquad [6\text{-}1]$$

Where:

C_m: minimum delay cycle length

L : Total loss time length

Y : sum of flow (volumes/ saturation flow) ratios for all critical phases

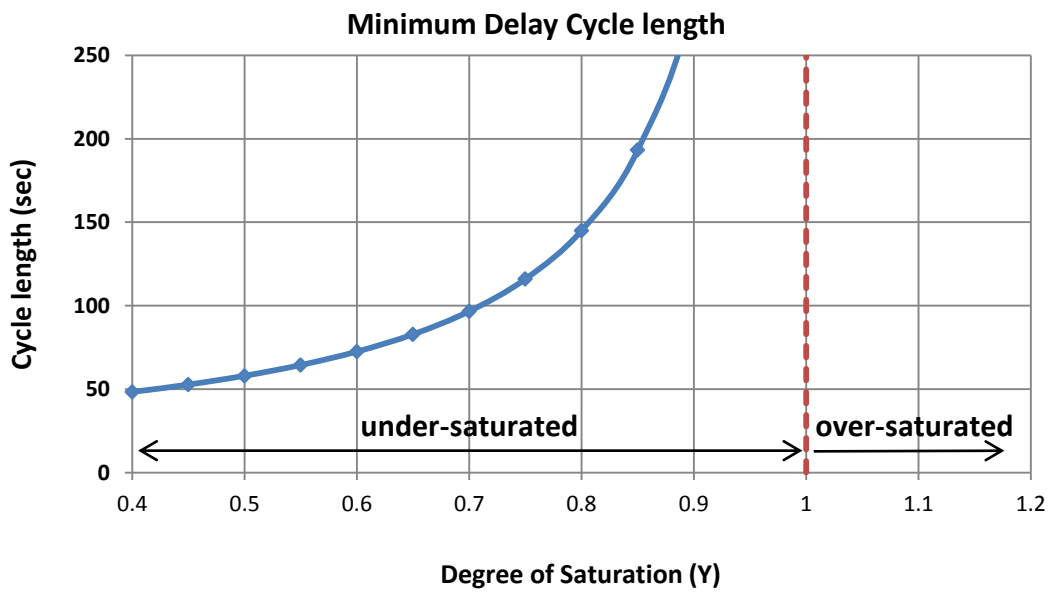


Figure 28: Minimum delay cycle length as function of intersection degree of saturation

In oversaturation conditions, where saturation level exceeds 1.0, Webster's formula is invalid. With "1-Y" in the denominator and Y is greater than 1.0, the formula gives negative value for cycle length. Thus, the minimum-delay cycle length is ***undefined*** when conditions are oversaturated. Therefore, in higher volumes, increasing the cycle length only would have very little effect on intersection delay, yet changing the percentage of green splits that allocated to certain movements are more effective than increasing the cycle itself.

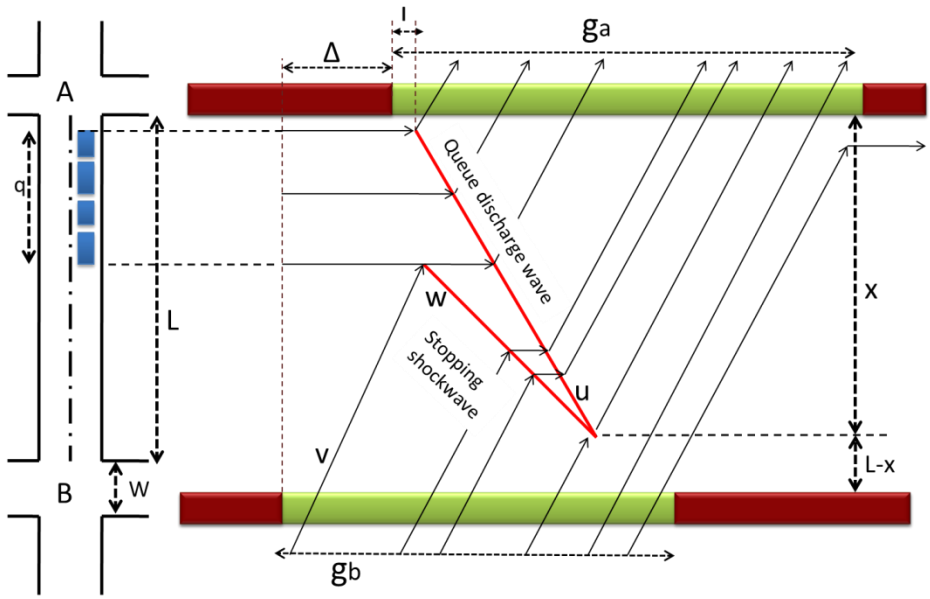


Figure 29: Shockwave at Signalized Intersection

Where:

L: Link length
W: Upstream intersection's width
ga: Effective green of downstream intersection
gb: Effective green of upstream intersection
h: Discharge headway
l: Loss time
Lv: Vehicle effective length
X: Location of where the stopping shockwave is canceled
C: Cycle length
SF: Safety factor (vehicle clearance)
u: Speed of queue discharge wave
v: Speed of leading vehicle in the incoming platoon

$$\mathbf{C \leq \frac{L \times h}{L_v} \times \frac{\left(1 - \frac{W}{L} - \frac{SF \times L_v}{L}\right)}{\left(\frac{ga - l}{C}\right)}} \qquad \textbf{[6-2]}$$

C is solved by iteration

The Internal Metering Policy (IMP) developed by Lieberman et al., [41] provides an upper bound for arterial background cycle length to avoid links' spillback at upstream intersection. The upper bound cycle is the maximum cycle that insures queue forming shockwave dissipates before reaching the upstream intersection. Figure 29 illustrates the maximum cycle length calculation that prevents spillback based on shockwave analysis.

Commonly, higher cycle lengths are considered effective in handling oversaturation (i.e., all queues served during one cycle), or in serving advancing platoons in a progressive system. However, queues formations during the corresponding "long" red time during high cycle would exceed links' storage capacities. Ross et al., [88] provided another formulation for maximum cycle length in oversaturated conditions. Ross formulation determines the maximum cycle for a link without overflow during the red phase. The maximum cycle length (C) is a function of downstream critical approach discharge rate (υ) and link length (L) as the following equation shows

$$C \leq \frac{L}{L_v} \times \frac{3600}{v_i} \qquad \textbf{[6-3]}$$

The minimum of equations [6-2] and [6-3] define the upper-bound of the network cycle length that can be implemented during oversaturation considering spillback avoidance. The lower bound is the minimum cycle length that can meet the safety and operational constraints of signal timing (e.g., pedestrians walking time, minimum green, yellow and all-red time, number of conflicting phases). The upper and lower bound define the feasible range of cycle length during oversaturation conditions. Therefore, in case of networks with short link or links with a heavy arrival rate, additional measures should be considered to avoid using very short cycles. Short links can be protected from spillback by either shortening the cycle length or by reducing arrival rates during red by adjusting offset for the critical routes that pass through this link (i.e., spillback-avert offset). Another approach to avoid link spillback is by providing extra green at downstream intersections (i.e., flaring the green) creating metering effect at downstream intersection.

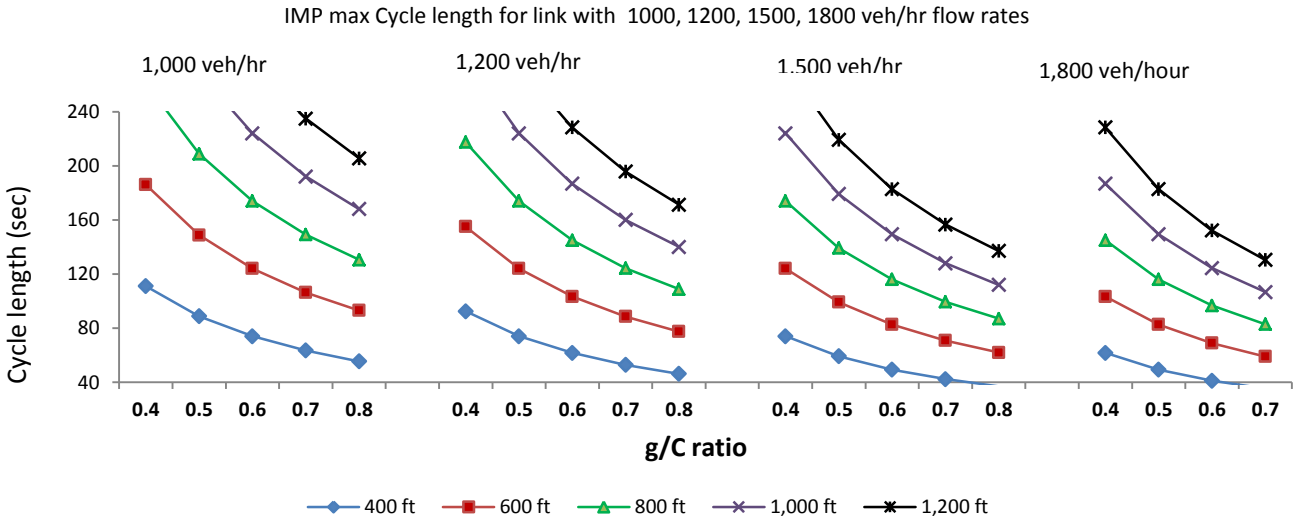


Figure 30: Maximum cycle length that prevents spillback based on approaches g/C ratio and link length

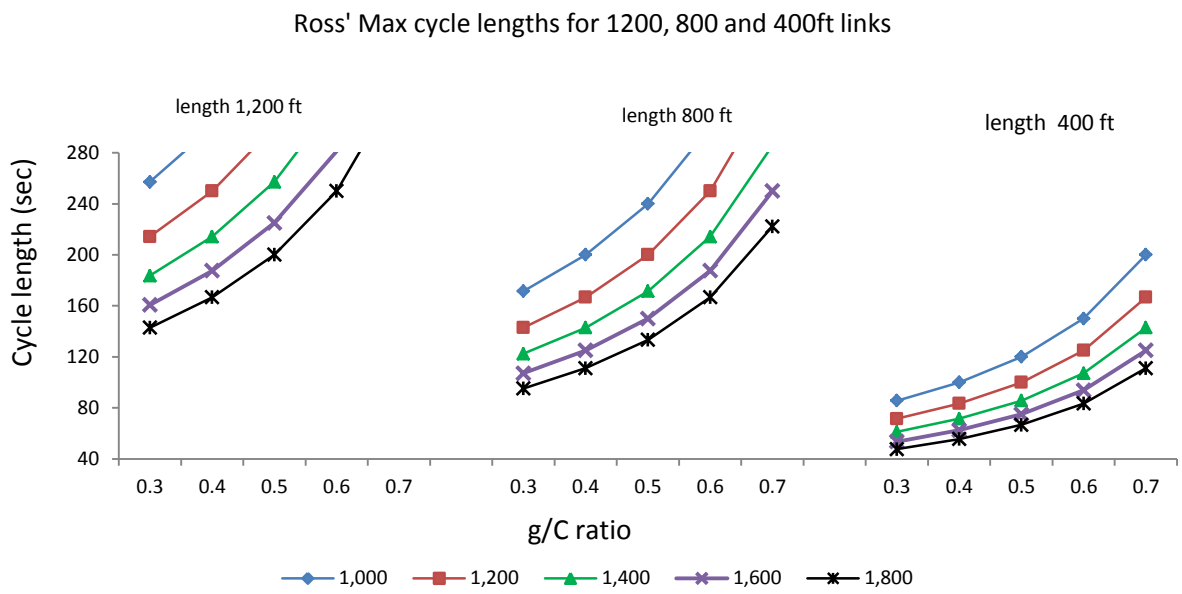


Figure 31: Maximum cycle length that prevents starvation based on R/C approaches R/C ratio and flow rate

6.3 Green Splits Design

Typically, green splits are determined based on conflicting approaches v/c ratios. These splits are constrained by minimum and maximum values assigned for safety and operational purposes (e.g., pedestrian crossing time). In the developed framework, initially, the green splits are determined based on the v/c ratios. Then, in the case of cycle failure (i.e., intersection oversaturated), critical movements would have the priority in green allocation. Green splits will be optimized so that the v/c < 1 for critical movements and the corresponding queues length within the storage capacity (i.e., queue-to-link ratio < 1). If the storage capacity of a predefined critical movement is overflowed, splits will be re-calculated again to avoid spillback occurrence by allowing v/c ratio of non-critical movements to get higher than one.

Several splits–based mitigation strategies are effective in reducing residual queues presence on intersection links. Strategies such as metering critical routes at pre-specified gating links (i.e., links with considerable storage capacities), or by double cycle the subject intersection creating phase re-service effect (in case of heavy arrival during red). The green splits are calculated based on the following assumptions:

- Queued vehicles discharge at saturation flow rate after *loss time*
- No intersection blocking
- No right turn on red
- Volume arrives uniformly

6.4 Oversaturation offset design

Maintaining progression during oversaturated conditions is not an easy task. Queues formation due to insufficient capacity at downstream intersection and heavy turning from upstream side streets disrupt the movement quality of platoons through the arterial, and thereby reduce the overall system performance. There are two competing objectives when it comes to design offsets in oversaturated conditions; prevents queues spillback at upstream intersection and maximizes green usage at downstream intersection. The first objective is achieved by implementing an offset that prevent the formation of a stopping shockwave. This shockwave progresses toward upstream intersection during the green phase reducing the speed of discharge wave of the through vehicles from upstream intersection and even block upstream intersection. The second objective, avoiding downstream starvation, is achieved when first released vehicle joined the discharged queue before crossing downstream intersection. Extra delay in releasing vehicle from upstream intersection will cause downstream starvation as it is shown in Figure 33. Fundamentally, the offset between two signalized intersections relies on links length, vehicle travel time, residual queue, and queue discharge rate. The following is a detailed description of oversaturation offsets calculations.

6.4.1 *Spillback-avert offset*

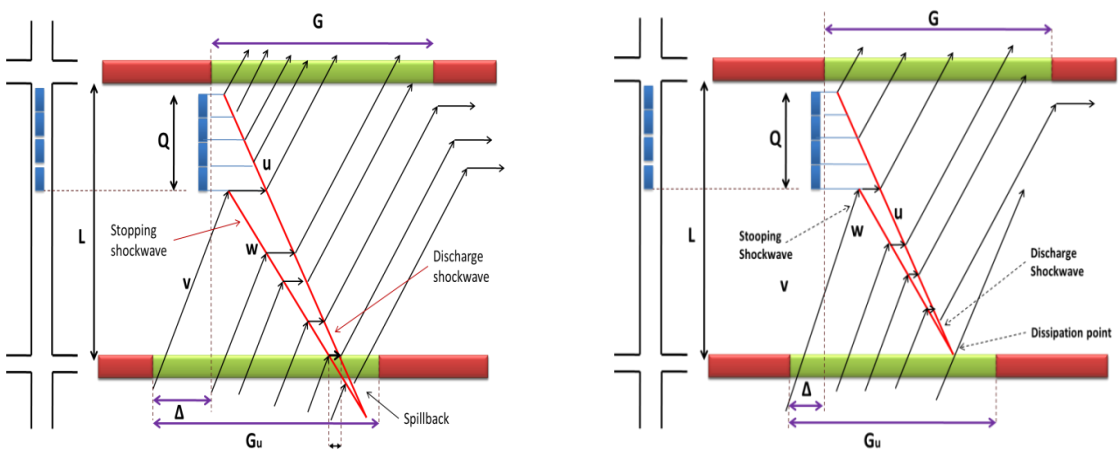


Figure 32: (a) link spillback and upstream intersection blockage (b) Spillback Avert Offset

This offset prevents spillback at upstream intersection by insuring the stopping shockwave dissipate before reaching upstream intersection. The ideal offset can be calculated via several methods that consider vehicle dynamics, queue perception impact that influence approaching vehicle's speed, and platoon dispersion effect. The following equation by Rathi [29]was derived based on the Law of motion to exemplify the concept.

$$\Delta \geq \frac{L}{U_s} - \frac{L\times(1-\rho)}{L_v} \times h \qquad [6\text{-}4]$$

6.4.2 *Starvation-avert offset*

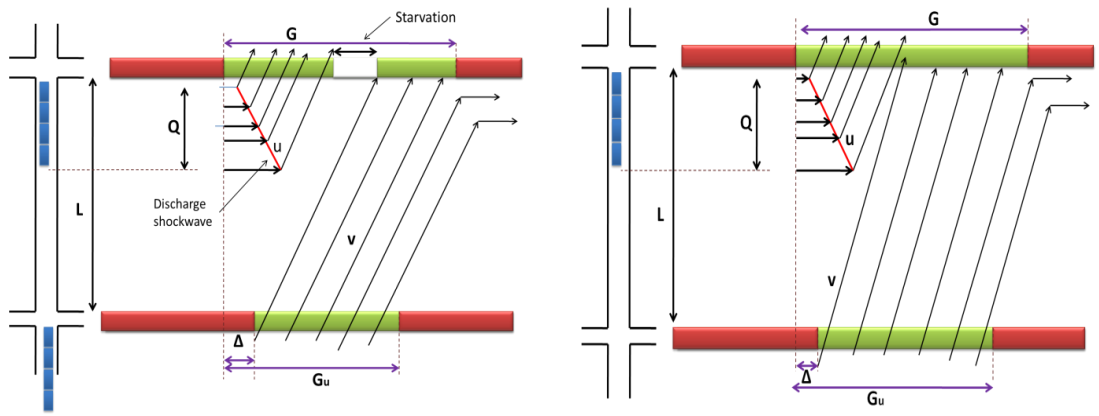


Figure 33: (a) starvation at downstream intersection (b) Starvation-avert offset

Starvation occurs at downstream intersection when the discharge of vehicles at upstream intersection is delayed beyond the ideal offset. Starvation causes capacity loss due to the wasting of limited green time at downstream intersection. The starvation avert offset is the offset that insures first released vehicle joins the discharge queue at downstream intersection. The simple form of starvation avert offset can be derived also as follow:

$$\Delta \leq \frac{L \times \rho \times h}{L_v} - \frac{L}{v}$$

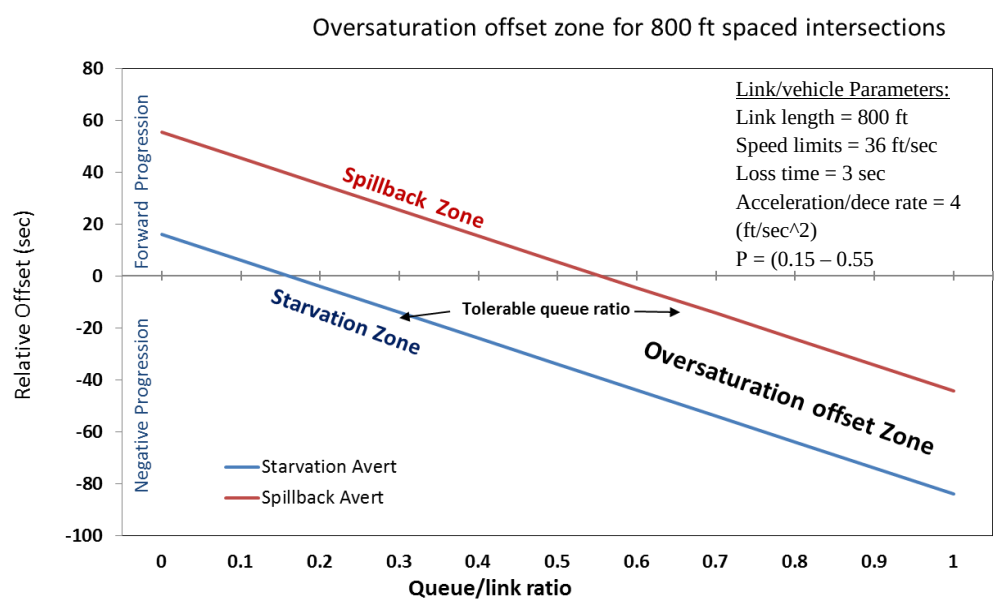


Figure 34: Offset Values Feasible Region

By obtaining the two offsets values (spillback-avert & starvation-avert), a feasible region of oversaturation offset values can be established. This region gives an idea of the optimal offset values that can achieve both objectives (upstream and downstream intersections' green utilization). For a

given length of a link, the key variable in determining the offset value is queue-link ratio (ρ-ratio). The value of (*p*) is used for computing oversaturation offsets values. The value of (*p*) is selected based on expected residual queue, link length, and travel speed. However, if (*p*) differs significantly from its design value, the offset performance will degrade, and could result in a condition of *de facto* red and spillback upstream, or waste of green downstream. Therefore, it is essential to control the value of (*p*) within a certain range. This can be achieved by reduce side street green time, and by prohibit side street turnings during red.

6.5 Splits-offsets calculation

The following is detailed description of the splits-offsets calculation procedure proposed to produce timing plans for oversaturated network. First, splits were determined through an optimization process for particular objective (throughput maximization, delay minimization, or queue length minimization, etc…) satisfying control constraints. The degree of saturation of critical routes approaches are constrained by pre-specified threshold to inhibit ***queues buildup***. The expected queue lengths are estimated for each approach according to their v/c ratio. Next, the resulting queue-to-link ratio (ρ-ratio) are used to determine the upper and lower bounds of oversaturation offsets (i.e., spillback-avert and starvation-avert). The following equations are used in the calculations.

1-Volume per cycle (VPC): approach average arrival during a cycle length:

$$VPC = \frac{V \times C}{3600}$$

2- The Pth percentile approach volume are calculated based on expected volume and peak hour factor (PHF).

$$V_{\alpha} = V \times PHF \times (1 + \sigma_{Pth} \times \frac{\sqrt{VPC}}{VPC})$$

The following percentiles can be used to estimate the arrival volumes: (70th, 85th, 90th, and 95th)

3- v/c ratio for intersection approaches$(v/c)_{ij}$ calculation

$$(v/c)_{ij} = \frac{V_{ij}}{s \times {g_{ij}}/{C}}$$

4- The expected queue length of each approach (Q_{ij}) is determined for the two cases of (v/c)

a) v/c < 1

$$Q_{ij} = \frac{V_{ij}}{3600} \times (C - g_{ij}) \times (1 + \frac{1}{{s}/{V_{ij}}}) \times \frac{L}{n_{ij}}$$

b) v/c > 1

$$Q_{ij} = \left[(V_{ij} - s \times \frac{g_{ij}}{C}) \times \frac{C}{3600} + V_{ij} \times \frac{C}{3600}\right] \times \frac{L}{n_{ij}}$$

5- Queue-link ratio (p_{ij}) is determined based on the expected queue length (Q_{ij}) and the available storage capacity for each approach (L_{ij})

$$p_{ij} = \frac{Q_{ij}}{L_{ij}}$$

6- Oversaturation control offsets are determined according to the (p_{ij}) ratio.

For the main approach

$$\Delta > \mathrm{p}_{ij} \times \frac{L_{ij}}{u_s} - \sqrt{(L_{ij}(1-p_{ij}))}$$

$$\Delta_{min} > \left(\frac{L_{ij}}{u_s}\right) - \frac{L_{ij}(1-p_{ij})}{L_v}.h$$

$$\Delta_{max} < \left(\frac{L_{ij \times p_{ij} \times h}}{L_v}\right) - \frac{L_{ij}}{v}$$

Where:

Let (ij) denotes an approach (j) at intersection (i), $\{j \in J,\ J = \{1, 2, \ldots, 12\}$

Let (v_{ij}) represents expected total volume of approach (j) at intersection (i)

Let (q_{ij}) represents queue length of approach (j) at intersection (i)

n_{ij} : Number of lanes of approach (j) at intersection (i)

L_{ij} : Storage capacity of approach (j) at intersection (i)

C: Cycle length

s_{ij} : Adjusted saturation flow rate per lane of approach (j) at intersection (i)

c_{ij} : Capacity of approach (j) at intersection (i), where $c_{ij} = s_{ij} n_{ij} \left(\frac{g_{ij}}{C}\right)$

g_{ij} : Approach (j) green time at intersection (i)

L_v: Queued vehicle length (ft.)

h_{ij}: Discharge headway approach (j) at intersection (i)

u_s: vehicle approaching speed

$(v/c)_{ij}$: Volume-capacity ratio of approach (j) at intersection (i), $(v/c)_{ij} = \frac{v_{ij} \times C}{s_{ij} \times n_{ij} \times g_{ij}}$

p^r_{ij} : Queue-link ratio for approach (j) at intersection (i) for the critical route (r), $p^r_{ij} = \frac{q^r_{ij}}{L_{ij}}$

PHF : Peak Hour Factor

Δ_{min}: Minimum oversaturation offset (starvation avert)

Δ_{max}: Maximum oversaturation offset (spillback avert)

6.6 Summary and Conclusion

This chapter presents in details the proposed principles of signal timing during oversaturated condition. First, the network's background upper-bound cycle length calculation is demonstrated. The calculations are based on IMP that considered network intersections spacing and arrival demand rates. Second, intersection green splits are determined considering network's critical movement's v/c-ratio and queue-to-link ratio. Finally, oversaturation offsets are calculated based on the calculated splits. The oversaturation offsets are intended to protect network critical routes from both spillback and loss of green. Several mitigation strategies can be applied at different stage of the timing process (i.e., cycle length-based strategies, green splits-based strategies, and offsets-based strategies).

7 Multi-objective Timing Tool for Oversaturated Arterial Network

7.1 Introduction

This section describes an analytical tool that generate near optimal timing plans for oversaturated arterial networks. The generated timing plans employ the theoretical backgrounds and concepts of oversaturation control that had been described in the previous sections. Cycle length, splits, and offset are calculated based on oversaturation control principles. The main idea of the developed tool is to manage queues at pre-defined links/approaches/movements and coordinate the progress of the critical movement in the network links by maximizing the use of green time; preventing downstream intersections starvation and upstream intersections de facto red. The control problem is formulated as a multi-objective problem with throughput maximization, queue management, and delay minimization as time-dependent control objective subject to state and control variables.

7.2 OTS features

The developed oversaturation timing spreadsheet (OTS) employs several features that are essential in generating optimal timing plan in oversaturated conditions such as: un-served demand carryover, multi-period optimization, critical routes/approaches/link/movements protection, and control objectives switching. The following is detail description to the OTS features.

7.2.1 *Timing based on movements' volumes profiles*

Traditionally, signal timing plans are generated based on the highest observed volume of conflicting movements during the analysis period (i.e., design volumes: 1,400 & 1,000 with v/c: 0.58 & 0.42, respectively). This practice is reasonable in cases of under-saturated conditions. However, during oversaturated conditions, approaches volumes are changing rapidly while capacities are limited (i.e., flow conditions are not steady-state).Thus, the percentages of v/c are expecting to change accordingly during over-saturation as **Figure 35** illustrate. Furthermore, each volume interval is not only different but each period has significant bearing on its successive period. Therefore, each developed timing plan is valid during a limited period of time (i.e., optimal plan steadiness) as its performance degrades significantly with the accumulated un-served demand.

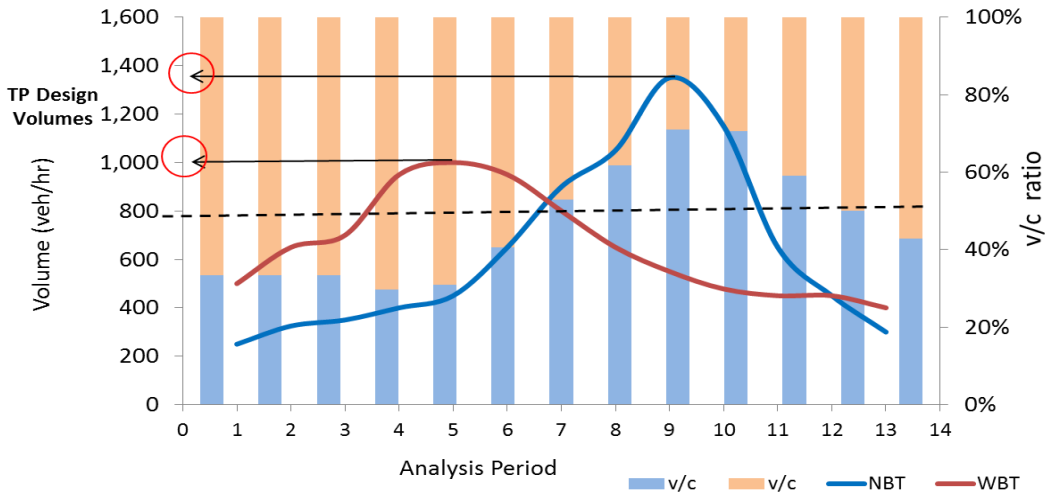


Figure 35: Two conflicting movements' volume profiles and their corresponding v/c ratios

In oversaturated conditions, when a queue does not clear at the end of the green (i.e., volume exceed capacity), these residual vehicles should be included automatically in the following cycle. OTS employs HCM procedure for delay calculations in multiple-time periods. Instead of the most commercial software that optimizes only one period in the volume profile (i.e., the highest volume) and assumes the rest of the control period to be a replication of the optimized period which makes their use in oversaturated flow conditions inappropriate. The OTS procedure analyzes volume intervals sequentially, carryover their final residual queues (if any) to the next interval. Therefore, OTS generates several timing plans using multi-period optimization with demand carryover that accounts for conditions during the current period as well as conditions of the previous periods.

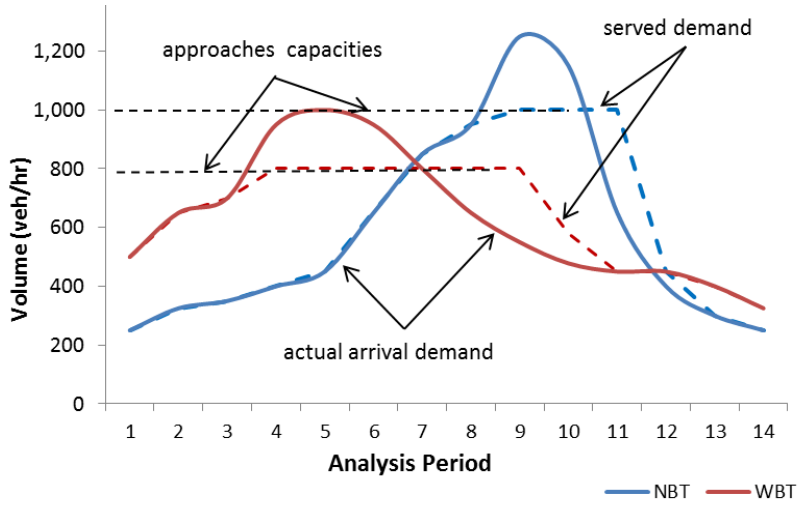
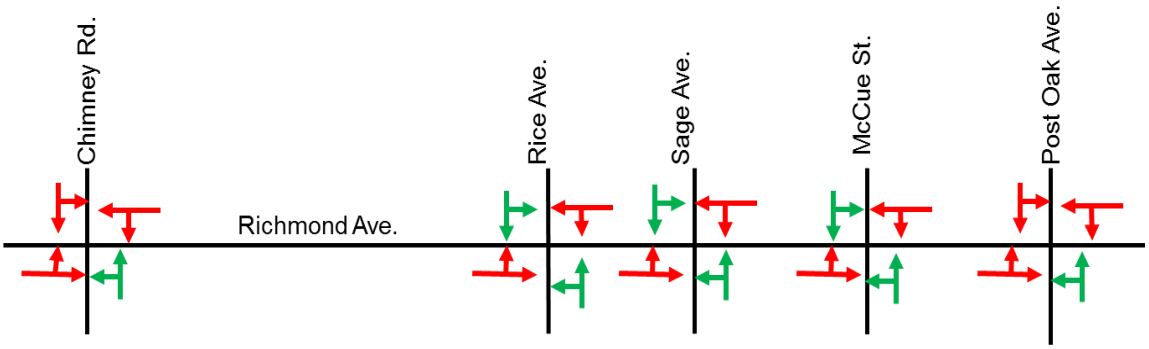


Figure 36: Demand profile during oversaturation condition

7.2.2 *Congestion matrix*



	NBL	NBT	NBR	SBL	SBT	SBR	EBL	EBT	EBR	WBL	WBT	WBR
Chimney Rd	1	1	1	0	0	1	0	0	1	0	0	1
Rice Ave.	1	1	1	1	1	1	0	0	1	0	0	1
Sage Ave.	1	1	1	1	1	1	0	0	1	0	0	1
Mcue St.	1	1	1	1	1	1	0	0	1	0	0	1
Post Oak Ave.	1	1	1	0	0	1	0	0	1	0	0	1

Figure 37: Timing tool congestion matrix, critical movement (1) and non-critical movements (0)

OTS incorporates congestion matrix in the optimization processes, congestion matrix is a binary matrix, as 1 is assigned for non-critical links while 0 is assigned for critical links. Congestion matrix allows the optimizer to distinguish between network critical and non-critical links/movements in order to allocate queues in non-critical approaches. Therefore, the congestion matrix represents the physical space available in the network. Congestion matrix is used to influence split calculations in favor of a specific link, when congestion is detected.

7.2.3 *Multi-objective control*

In under-saturated conditions, delay minimization and bandwidth maximization are the two main strategies that are used to optimize traffic signals in arterial networks. Targeting one or both of these strategies may not fully provide effective control strategies during oversaturated conditions. In oversaturated conditions, queues develop and grow in length, and the total delay increases exponentially as a function of the elapsed time. Therefore, congestion can spread out spatially and temporally due to queue spillback and may cause gridlock. Minimizing delay alone is not sufficient to resolve queues and their impacts, most simply because it cannot be directly measured. In order to possibly maintain optimal states of traffic during over-saturated conditions or transitions between unsaturated and over-saturated conditions, an integrated criterion that combined capacity maximization, delay minimization and queue management must be considered through a multi-objective analysis framework.

There are several reasons to reflect the three phases of oversaturation regimes in control operation objectives and in the performance evaluation. First, the multi-objective analysis performed as part of this research indicates that improved performance can be obtained over time by switching from one mitigation approach to another. Although we have developed limited prescriptive guidelines for application of different strategies during each phase of operation, Abbas, et al; [89], it is clear that many strategies perform differently in traffic conditions that are increasing (loading), constant (oversaturated), or decreasing (recovery). One might switch between strategies using traffic-responsive type of operation or using any other congestion management tool.

1. During *loading phase*, the application of throughput strategies (i.e., green utilization strategies and forward progression) would likely provide a measurable improvement in performance on the approaches, routes, and networks that will shortly become oversaturated. Early application of throughput mitigation strategies is easier to conceptualize when the causal factors are recurrent.
2. Processing phase stage starts once the maximum degree of saturation of critical approaches reaches a pre-specified threshold. In this stage, queue managements control should be applied with an objective of protecting critical routes approaches from spillback.
3. During *recovery phase*, the decision of either to implement delay minimization or throughput maximization strategies is based on arrival demand profiles. If traffic volumes rapidly drop, delay minimization control strategy is recommended. While gradual drop in arrival demand requires a return back to the throughput maximization controls strategy to accelerate system steady-state.

The optimization criterion proposed herein is an application of several choices criteria for different traffic demand phases (i.e., loading, processing, and recovery).

Table 6: Oversaturation phases and the recommended control objectives

Phase	Optimization criteria	Objective
Loading	Capacity maximization or Delay minimization	Utilize network capacity and postpone the start of oversaturated phase
Processing	Queuing Synthetic criteria	**Avoid:** gridlock, capacity reduction, and secondary congestion
Recovery	Capacity maximization or Delay minimization	Quick recovery from oversaturation

7.3 Delay Calculation

OTS calculates delay according to HCM procedure for oversaturated condition [90]. The HCM procedure considered the following delay components; uniform arrival delay, random arrival delay, and residual queue delay. The following equations are used to estimate the three components of delay.

Uniform delay (d1):

$$d_1 = PF \times \frac{0.5 \times C \times \left(1 - \frac{g_{ij}}{C}\right)^2}{1 - \left[\min\left(1, \frac{v_{ij}}{c_{ij}}\right) \times \frac{g_{ij}}{C}\right]}$$

Random arrival delay (d2):

$$d_2 = 900 \times 0.25 \times \left[\left(\frac{v_{ij}}{c_{ij}} - 1\right) \sqrt{\left(\frac{v_{ij}}{c_{ij}} - 1\right)^2 + \frac{8 \times k \times I \times \left(\frac{v_{ij}}{c_{ij}}\right)}{0.25 \times c_{ij}}}\right]$$

Residual queue delay (d3):

$$d_3 = \frac{1{,}800 \times Q_{bij} \times (1 + u) \times t}{C \times T}$$

Where

(i) Denote an intersection in the network, $\{i \in I,\ I = \{1, 2, \dots, N\}$, Where N is total number of intersections

(ij) denotes an approach (j) at intersection (i), $\{j \in J,\ J = \{1, 2, \dots, 12\}$

(v_{ij}) represents expected total volume of approach (j) at intersection (i)

C: Cycle length

c_{ij}: Capacity of approach (j) at intersection (i), where $c_{ij} = s_{ij} n_{ij} \left(\frac{g_{ij}}{C}\right)$

g_{ij}: Approach (j) green time at intersection (i)

$(v/c)_{ij}$: Volume-capacity ratio of approach (j) at intersection (i), $(v/c)_{ij} = \frac{v_{ij} \times C}{s_{ij} \times n_{ij} \times g_{ij}}$

PF: Progression factor

Q_{bij}: Size of initial queue

T: Analysis period for delay calculation, hour

t: Duration of oversaturation within T, hour

u: Delay parameter

7.4 Throughput calculations

OTS assumes a linear increase in the saturation rate from (0 veh/hr) at loss time to maximum saturation flow rate at (t_1) sec after green onset. Several studies concluded that an intersection approach reach its max saturation rate from (7 to 10 sec) after green start. This assumption gives better estimation to the approaches' saturation rates than the previous method. OTS users can insert the value of (l), (t_1), and max saturation rates.

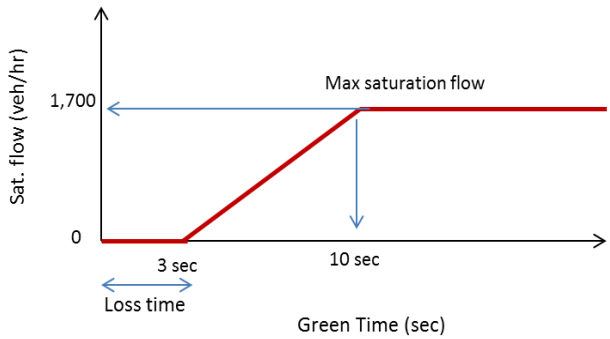


Figure 38: Intersection saturation rate

Throughput flow rate:

$$discharge\ flow = \begin{cases} 0 & if \quad t < l \\ \dfrac{(t-l)}{(t_2-l)} \times (\max\ sat.flow) & if \quad l < t < t_2 \\ \max\ sat.flow & if \quad t > t_2 \end{cases}$$

$$Throughput = min\left\{ \int_l^{t_2} \frac{(t-l)}{(t_2-l)} \times (\max\ sat.flow).dt + \int_{t_2}^{g} (g-t_2) \times (\max\ sat.flow).dt\ , arrival\ volume \right\}$$

<u>Where</u>

l: loss time
g: green time
t_2: Time at max saturation rate

7.5 Control problem formulation

The proposed control procedure explicitly accounts for the principles stated in the previous sections and sets the control parameters to optimize the system selected control objectives (i.e., capacity maximization, delay minimization, or queue management). The procedure achieves this by ensuring that no green time is wasted and has been used to the greatest degree possible. As well, it ensures that queues formations do not block the upstream intersections and ensures that spillback is prevented. As saturation levels change during the peak period, the objectives control needs to respond accordingly. This is one of the major developments in oversaturation timing framework. During loading phase, the control procedure is proactive; it attempts to maximize the traffic entering and leaving the system and ensures to provide control so that undesirable conditions do not develop early. As traffic volumes increase, control objective shift to queue management operation to ensure that queues formation do not grow in the network most critical links. To achieve the objectives of the queue management operation, the control problem constraints need to be slackened. Control constraints such as min/max greens and cycle length can be ignored in the control problem.

In summary, the essence of the control procedure proposed here is to apply different control objective to maximize system performance while (a) ensuring using of all green time efficiently (no wasted green); (b) preventing formation of de facto red at intersections in the system; and (c) prevent

queue formation in network critical routes. The mathematical expressions for these formulation and conditions are presented in the following section.

7.5.1 *Parameters Definitions*

Let (r) denote a critical route consist of a sequences of movements$\{m_1^r, m_2^r, \dots \dots \dots m_n^r\}$ where $m_i \in \{\text{Left, Through, Right}\}, r \in R$

m_{ij}^r Critical movement of approach (j) at intersection (i) for route (r)

Let (i) denote an intersection in the network,$\{i \in I,\ I = \{1, 2, \dots, N\}$, Where N is total number of intersections

Let (ij) denotes an approach (j) at intersection (i), $\{j \in J,\ J = \{1, 2, \dots, 12\}$

Let (k) denote the oversaturated period phase, $k \in K,\ K = \{1\text{: loading phase},\ 2\text{: processing phase},\ 3\text{: reecovery phase}\}$

Let (v_{ij}) represents expected total volume of approach (j) at intersection (i)

Let (q_{ij}) represents queue length of approach (j) at intersection (i)

Let (q_{ij}^r) represents queue length of approach (j) at intersection (i) for the critical route (r)

n_{ij} : Number of lanes of approach (j) at intersection (i)

L_{ij} : Storage capacity of approach (j) at intersection (i)

C: Network cycle length

C_{max} : Maximum cycle length

C_{min} : Minimum cycle length

c_{ij} : Capacity of approach (j) at intersection (i), where $c_{ij} = s_{ij} n_{ij} \left(\frac{g_{ij}}{C}\right)$

g_{ij} : Approach (j) green time at intersection (i)

g_{ijmin} : Minimum green split (sec)

g_{ijmax} : Maximum green split (sec)

Y_{ij}: Approach (j) (yellow + all red) time at intersection (i)

$(v/c)_{ij}$: Volume-capacity ratio of approach (j) at intersection (i),

p_{ij}^r : Queue-link ratio for approach (j) at intersection (i) for the critical route (r), $p_{ij}^r = \frac{q_{ij}^r}{L_{ij}}$

γ_{ij}^r : Maximum allowable value of queue-link ratio for approach (j) at intersection (i) for the critical route (r)

δ_{ij} : Maximum allowable value of v/c ratio for approach (j) at intersection (i)

δ_{ij}^r : Maximum allowable value of v/c ratio for approach (j) at intersection (i) for the critical route (r)

w_0 : Weight of total delay

w_{ij}^r: Weighting coefficients of approach (j) at intersection (i) for the critical route (r)

ϕ : Oversaturation volume-to-capacity ratio threshold, ($\phi \leq 1$)

Definition of oversaturation period

- Loading phase $(0, T_1)$:
- Processing phase(T_1, T_2)
- Recovery phase (T_2, T_3)

  Where:

  $T_1 \coloneqq \min\{t: x(t) \geq \emptyset\}$

  $T_2 \coloneqq \max\{t: x(t) \geq \emptyset\}$

  $horizon \coloneqq [0, T_2]$

$k = 1 \quad if \quad t < T_1$

$k = 2 \quad if \quad t > T_1 \ and \ t < T_2$

$k = 3 \quad if \quad t > T_2$

$X_{i(t)}$: saturation rate at intersection (i) as function of time (t)

$\boldsymbol{Thr}$: Throughput maximization control objective
$\boldsymbol{J_c}$. : Queue management control objective
Delay : Delay minimization control objective

7.5.2 *Objective function*

Let ($\boldsymbol{Z_i}$) be the network performance function. This function represents the utility of the network total control objectives. User can define the sequences of control objectives applied for different saturation regimes (i.e., loading, processing, and recovery). After defining the sequences of the control objectives, ($\boldsymbol{Z_i}$) will be maximized and the corresponding optimal timing plans will be obtained. In under saturated condition, the classic objective is to minimize delay. However, in oversaturated condition, maximizing the network throughput during the loading period may help to postpone the oversaturation phase. During recovery phase, it preferable to minimize delay, if demand drops rapidly otherwise maximize throughput. For the queue management control, the cycle length and min/max green constrains are relaxed. Each intersection in the network can be run with different cycle. The following sequences of control objectives (P_i) during oversaturated regimes are considered.

$$(\boldsymbol{Zi}) = \int_{t=0}^{T_1} P_1 + \int_{t=0}^{T_2} P_2 + \int_{t=0}^{T_3} P_3$$

Thus, the total utility of the control plan ($\boldsymbol{Z_i}$) can be described using the following relationship:

$$\boldsymbol{Z_i} = \sum_{\boldsymbol{t=0}}^{\boldsymbol{T_1}} P_1 + \sum_{\boldsymbol{t=T_1}}^{\boldsymbol{T_2}} P_2 + \sum_{\boldsymbol{t=T_2}}^{\boldsymbol{T_3}} P_3$$

7.5.3 *Problem formulation*

$$Max\ Z_i$$

7.5.4 *Control variables:*

- Cycle length (C) and intersections approaches greens (g_{ij}), or
- Approaches green (g_{ij}) with a fixed cycle length (C)

<u>*Subjected to the following constraints*</u>

1- Cycle length constraints

$$C_{min} < C_k < C_{max}$$

2- Cycle length constraint

$$\sum_j \left(g_{ij}^k + Y_{ij}^k\right) = C_k$$

3- Signal barrier constraints

$$g_{i1}^k + g_{i2}^k = g_{i5}^k + g_{i6}^k$$
$$g_{i3}^k + g_{i4}^k = g_{i7}^k + g_{i8}^k$$

4- Minimum and maximum greens constraints:

$$g_{ij}^k \geq g_{min}^k$$
$$g_{ij}^k \leq g_{max}^k$$

5- Critical approaches P-ratio constraints:

The problem formulation allows user to define an upper limit (γ_{ij}^{r}) for links' p-ratio for each critical link (r). This threshold is set as constraints the optimization process.

$$p_{ij}^{r} \leq \gamma_{ij}^{r}$$

Where

$$p_{ij}^{r} = \frac{q_{ij}^{r}}{L_{ij}^{r}}$$

6- Critical approaches v/c-ratio constraints:

$$\frac{v_{ij}^{r}}{c_{ij}^{r}} \leq \delta_{ij}^{r}$$

7- Non- critical approaches v/c-ratio constraints

$$\frac{v_{ij}}{c_{ij}} \leq \delta_{ij}^{+}$$

$$\forall\, r \in R, \forall\, i \in I,\ \ \forall j \in J$$

$$\forall\, k \in K, \qquad K = \{1, 2, 3\}$$

7.5.5 ***Queue management control***

Queue management mitigation control is expected to protect network from primary congestion symptoms such as intersections blockage and gridlock. This may not be realized, if the network' intersections are constrained by a single cycle operation. For such condition, each intersection need to find its optimal cycle length independently. Therefore, cycle length constraint is relaxed in the optimization formulation along with min/max green constraints.

$$Max\ Z$$

Subjected to the following constraints

1- Cycle length range constraints

$$C_{min} < C_{ki} < C_{max}$$

2- Signal barrier constraints

$$g_{i1}^{k} + g_{i2}^{k} = g_{i5}^{k} + g_{i6}^{k}$$

$$g_{i3}^{k} + g_{i4}^{k} = g_{i7}^{k} + g_{i8}^{k}$$

3- Critical approaches P-ratio constraints:

$$p_{ij}^{r} \leq \gamma_{ij}^{r}$$

$$\forall\, r \in R, \forall\, i \in I,\ \ \forall j \in J$$

$$\forall\, k \in K, \qquad K = \{1, 2, 3\}$$

7.6 Optimal control switching points problem

As the control philosophy during oversaturated conditions assumes system dynamic, it requires changing both of the timing plans and control objectives during the control period. The following section describes a critical element of the OTS, the selection and switching of the optimal timing plans. OTS select optimal timing plans and their switching points for pre-define sequences of control

objectives. The plans that meet the control objectives are selected from a library of earlier optimized timing plans (i.e., multistage optimization).

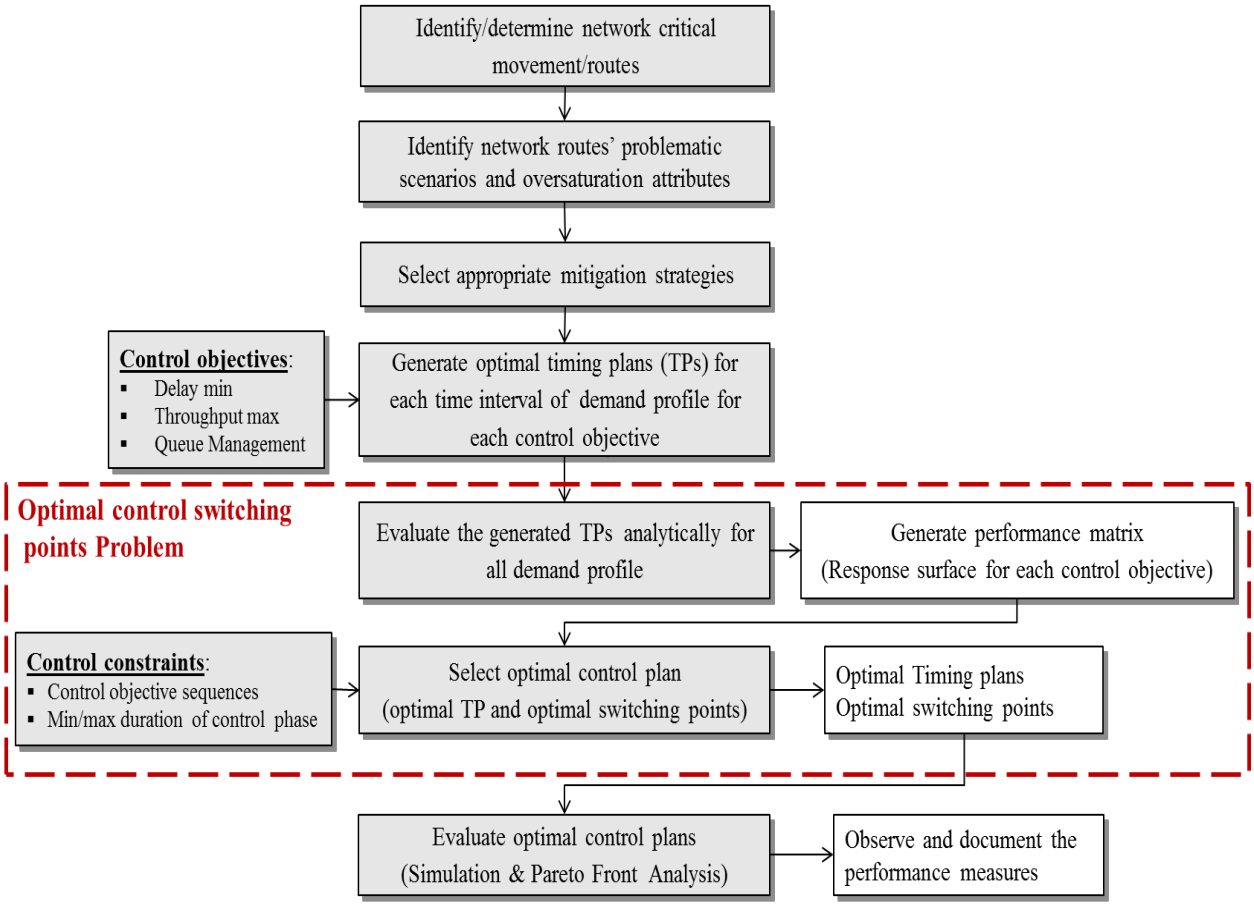


Figure 39: Oversaturation timing Framework

7.6.1 *Optimal switching points Problem formulation*

The selection of "optimal" control plan in oversaturation timing procedure is formulated as a ***linear integer programming problem***. A set of binary variables associated with the generated TPs' assignment is considered as control variables (Y_{jk}) in the formulation. The formulation includes linear constraints for scheduling requirements, an assumed sequence of control objectives, and minimum/maximum duration extents of each control period. The minimum duration of each control strategy is required as an input because when a new timing plan is selected, the traffic controllers enter into transition. During this process, operation in the network is interrupted while each controller adjusts to the new parameters. If the time period of the prescribed new timing plan is too short, the controller may simply end transition and then re-start transition into the next timing plan.

Formulation variables definitions:

Let (k) denote the control objective of the optimal timing plan,

$k \in K,\ K = \{1: Delay\ min, 2: Queue\ management, 3: Throughput\ max\}$

Let (j) denote time period index of the analysis period, $j \in J,\ J = \{1, 2, \dots, N\}$

Let (Y_{jk}) denote the assignment variables of control strategy (k) at time period (j)

Where:

$$Y_{jk} = \begin{cases} 1 & if\ strategy\ (k)\ is\ selected\ for\ time\ period\ (j) \\ 0 & otherwise \end{cases}$$

Let (n_k) denote the number of time periods run by control strategy (k)

Let (τ_k) denote the minimum number of time periods run by control strategy (k)

Let (φ_1) denote the entry oversaturation threshold

Let (φ_2) denote the exit oversaturation threshold

Let (D_{ij}) denote the delay of time period (j) result by applying timing plan (i)

Where $i \in I,\ I = \{i \leq N$ for $k = 1, N < i \leq 2N$ for $k = 2, 2N < i \leq 3N$ for $k = 3\}$

Problem Inputs

- Optimal timing plans result from multi-stage optimization (D_{ij}) and (Thr_{ij}), Where:
  - D_{ij} is the delay resulted from applying optimal TP(i) at time period (j)
  - Thr_{ij} is the throughput resulted from applying optimal TP(i) at time period (j)
- Sequences of control objectives (e.g., delay min - queue management- throughput max)
- Minimum and maximum duration time of control strategies (e.g., min = 30 min, max = 90 min)

Control variable

Assignment variable of applied timing plans (X_i) is a binary variable of (1 if the timing plan was selected, 0 otherwise)

Objective functions

Min (Delay)

Where:

$$Delay = \sum_i \sum_j D_{ij} \times Y_{ij} \qquad \forall i \in I, \quad \forall j \in J$$

$$Thr = \sum_i \sum_j Thr_{ij} \times Y_{ij} \qquad \forall i \in I, \quad \forall j \in J$$

<u>Subjected to:</u>

1- Timing plans assignment constraints

$$X_i \leq 1\ and\ integer \qquad \forall i \in I$$

$$Y_{ij} = X_i \qquad \forall i \in I, \quad \forall j \in J$$

2- Assignment constraints for each time period (j) (only one TP is assigned for each time period)

$$\sum_i Y_{ij} = 1 \qquad \forall i \in I, \quad \forall j \in J$$

3- TP assignment constraint for loading regime

$$\sum_{i=1}^{N} X_i = 1 \qquad \forall i \in I$$

4- TP assignment constraint for processing regime

$$\sum_{i=N+1}^{2N} X_i = 1 \qquad \forall i \in I$$

5- TP assignment constraint for recovery regime

$$\sum_{i=2N+1}^{3N} X_i = 1 \qquad \forall i \in I$$

6- Minimum time periods for each control strategies (optional)

$$\sum_{j} \sum_{i=1}^{N} Y_{ij} \geq \tau_1 \qquad \forall i \in I, \qquad \forall j \in J$$

$$\sum_{j} \sum_{i=N+1}^{2N} Y_{ij} \geq \tau_2 \qquad \forall i \in I, \qquad \forall j \in J$$

$$\sum_{j} \sum_{i=2N+1}^{3N} Y_{ij} \geq \tau_3 \qquad \forall i \in I, \qquad \forall j \in J$$

7- Total sum of time periods for control strategies constraint

$$\sum_{i} \sum_{j} Y_{ij} = N \qquad \forall j \in J, \qquad \forall k \in K$$

8- Consecutive time periods for each control strategies

$$\sum_{m=0}^{n_k} Y_{j+m,k} = n_k \qquad \forall j \in J, \qquad \forall k \in K$$

Where:

$$\sum_{j} Y_{jk} = n_k \qquad \forall j \in J, \qquad \forall k \in K$$

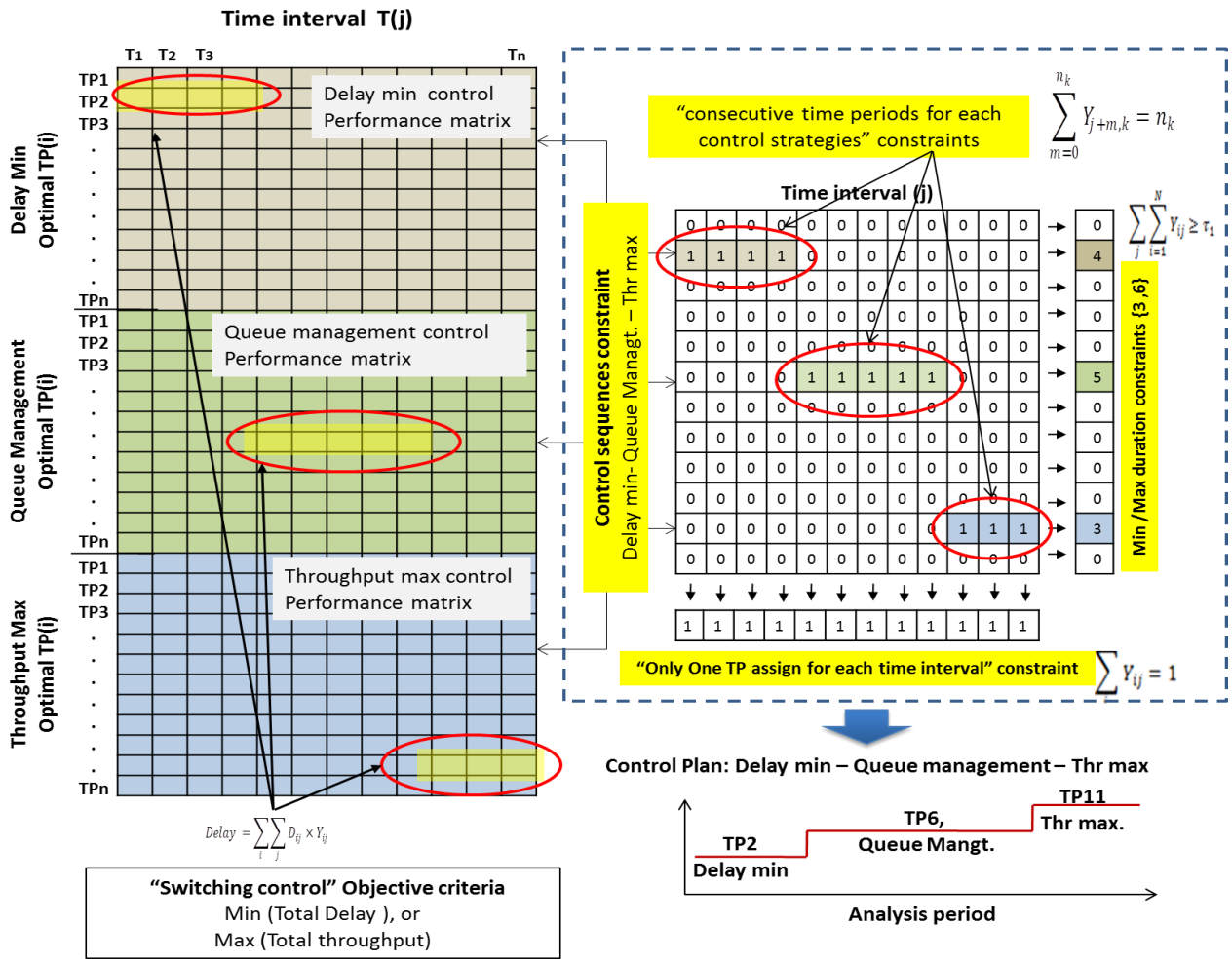


Figure 40: Control Plan: optimal time plans, control objectives, and optimal switch points

7.6.2 *Optimization solver*

A hill-climbing iteration process is used to obtain a set of optimal signal timings in the developed timing tool. Generalized Reduced Gradient (GRG) algorithm, one of the most robust nonlinear programming methods, provided by Excel-solver is used in the timing tool. "The GRG method has been developed and proven to be one of the efficient and effective methods for the Non-linear Programming problem with Non-linear constraints. The basic concept of GRG method entails linearizing the non-linear objective and constraint functions at a local solution with Taylor expansion equation. Then, the concept of reduced gradient method is employed which divides the variable set into two subsets of basic and non-basic variable and the concept of implicit variable elimination to express the basic variable by the non-basic variable. Finally, the constraints are eliminated and the variable space is deduced to only non-basic variables. The proven efficient method for non-

constraints NLP problems is involved to solve the approximated problem and, then, the next optimal solution for the approximated problem should be found."

7.7 OTS optimizer improvement: Initial solution

The random generation of green splits, causes OTS solver, most of the time, to be trapped in unfeasible solutions. The following modifications were added to OTS to improve the solution quality and speed up the solution runs time:

- Generate the initial splits based on triangular distribution with (g_{min}, gv/c, g_{max}) as distribution supports.
- Then the generated green splits are scaled to the ring constraint (i.e., cycle length)

These modifications to the solver insure the search for optima will start from a feasible surface and with initial splits near to optimal solution. The parameters of a triangular distribution herein can be derived directly from minimum green, maximum green, and the green split ($g_{v/c}$) obtained from the approaches saturation levels and v/c ratios.

$$g_{initial} = tringular\ (g_{min},\ g_{max},\ g_{v/c})$$

Where:

$\boldsymbol{g_{v/c}}$: the green splits calculated based on degree of saturation of conflicting phases

$$\boldsymbol{g_{v/c}} = \frac{\dfrac{\boldsymbol{v_i}/\boldsymbol{s_i}}{(\boldsymbol{v}/\boldsymbol{c})_i}}{\sum_i^n \left(\dfrac{\boldsymbol{v_i}/\boldsymbol{s_i}}{(\boldsymbol{v}/\boldsymbol{c})_i}\right)} \times (\boldsymbol{C} - \boldsymbol{L})$$

Then the generated initial splits are scaled down to the ring constraint (cycle length).

$$\widehat{\boldsymbol{g_i}} = \frac{\boldsymbol{g_i}}{\sum_j (\boldsymbol{g_i})_{ring}} \times C$$

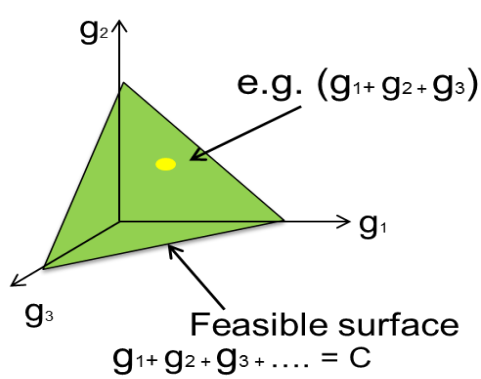


Figure 41: illustration of the initial solution feasible surface

Algorithm variables

r_1 , r_2

Algorithm parameters

m_1 : = maximum number of runs with pre-defined constraints
m_2 : = maximum number of runs with relaxing v/c constraints for non-critical approaches
Δ : = (v/c) increment
SPV: = spillover volume (un-served demand)
δ_{ij}^{nc}: = (v/c) maximum allowable value of network non-critical approaches

Step 1
Define control objective: P_i
Input formulation constraints: (C_k, g_{ijmin}, g_{ijmax}, Signal Barrier, δ_{ij}, δ_{ij}^{r} , γ_{ij}^{r})

Step 2
$r_1, r_2 = 1$
Find Loc Min = 1 {initialized solver}
While (r_2 < m_2 or ***Find Loc Min*** = = 1)
While r_1 < m_1
Initialize g_{ij}^{k} {Initialize g_{ij}^{k} randomly,$U(g_{min}, g_{max})$}
Run ***Timing*** {function that calculates: approach capacity, v/c ratio, expected queue length, p-ratio, and oversaturation offsets}
Run ***Find Loc Min*** {function search for local minima with g_{ij}^{k} as initial guess}
If ***Find Loc Min*** == 0 {local minima has been found}
Return g_{ij}^{k}
Return $P(g_{ij}^{k})$ {Performance measures ofg_{ij}^{k} }
Calculate SPV {determine the demand that exceed the capacity during the analysis period}

Else
$r_1 = r_1 + 1$
End if
End While
Initialize g_{ij}^{k}
Run ***Timing***
$\delta_{ij}^{nc} = \delta_{ij}^{nc} + \Delta$ {Increase the v/c ratio constraints at non-critical approaches}
If ***Find Loc Min*** == 0
Return g_{ij}^{k}
Return $P(g_{ij}^{k})$
Calculate SPV
Else
$r_2 = r_2 + 1$
End if
End While
Return "No feasible solution"
End

Function ***Timing***: this function calculates intersection: capacity, v/c, expected queue, P-ratios, spillover incidents, and oversaturation offsets}

Function inputs

C_k: Cycle length
g_{ij}^k: Green splits
L_{ij}: link length
s_{ij} , : Saturation rates
n_{ij} : approach number of lanes
h_{ij}: Saturation discharge rate
V_{ij}: approach demand

Function variables

---

i, j

Function ***Timing***

For i = 1 to I (for each intersection)
For j = j to J (for each approach)
Calculate approaches capacities $c_{ij}^k = s_{ij}\mathrm{n_{ij}}\left(\frac{g_{ij}^k}{C_k}\right)$
Calculate approached v/c ratio $(v/c)_{ij} = \frac{v_{ij}\times C_k}{s_{ij}\times \mathrm{n_{ij}}\times g_{ij}^k}$

if a $(v/c)_{ij} < 1$ then
$Q_{ij} = \frac{V_{ij}}{3600} \times (C_k - g_{ij}^k) \times (1 + \frac{1}{s/V_{ij}}) \times \frac{L_{ij}}{n_{ij}}$ {Calculate queue length}
else
$Q_{ij} = \left[(V_{ij} - s_{ij} \times \frac{g_{ij}^k}{c_{ij}^k}) \times \frac{C_k}{3600} + V_{ij} \times \frac{C_k}{3600}\right] \times \frac{L_{ij}}{n_{ij}}$ {Calculate queue length}
end if
Calculate p-ratio $p_{ij} = \frac{Q_{ij}}{L_{ij}}$
Calculate oversaturation offsets {spillback-avert, starvation-avert}
Calculate approach delay
Calculate approach throughput
Next j
Next i

Return (delay, throughput, p-ratio, v/c ratio, offsets)

7.8 Illustration example: Initial solution impact on optimal solution

A single oversaturated intersection average delay was optimized using first, solver with random initial solutions set, second with "v/c-based and scaled" initial solutions set. Several runs were conducted and the initial solutions delays were plotted against the optimized delays for both generation methods. The purpose of the example is to evaluate the performance of the improved optimizer with respect to the random initial solution.

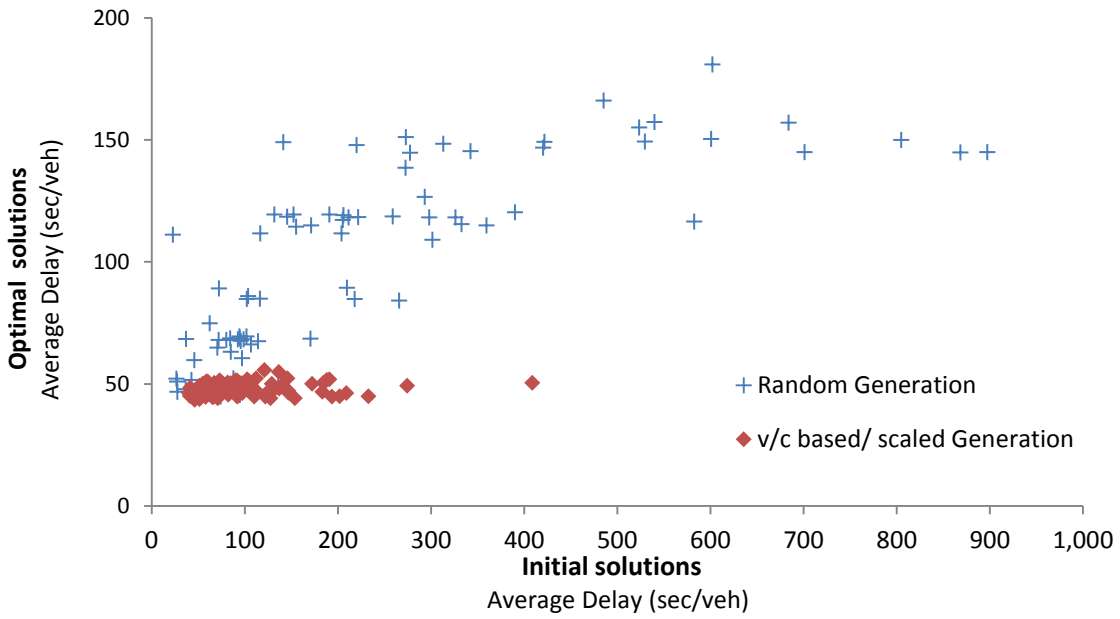


Figure 42: comparison of random & "v/c-based and scaled" generated initial splits solutions

As the above figure illustrates, the solver started searching from a relatively high values when using random generated initial splits compared with the "v/c-scaled & constrained" initials. Due to the nonlinearity in objective function formula, several optimal solutions were resulted when using the random generated initial splits. While the enhanced solver produced, relatively, consistent results when using "v/c-based and scaled" initial splits (i.e., delay = 47 sec/veh)

Number of solution runs

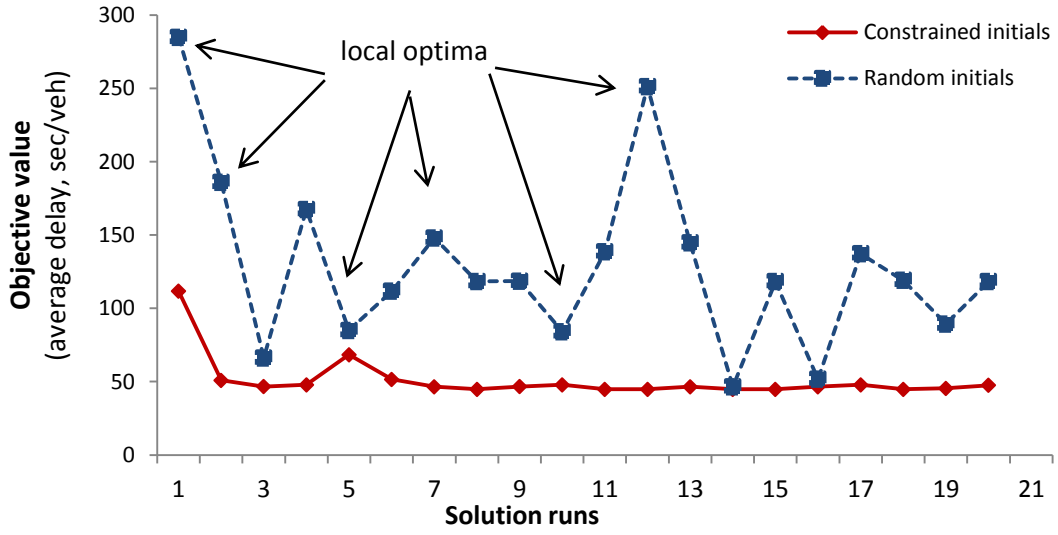


Figure 43: comparison of random & "v/c-based and scaled" generated initial splits solutions

When using "v/c-based & scaled" initial solutions the number of runs required to obtained optimal solution decrease significantly comparing with the random initial. Figure 43 shows the optimal solution progress per solution runs for both type of initial solutions.

7.9 Model verification and validation

This section describes the framework used for the verification and validation of the developed OTS. In order for the developed model to be realized, the model must be valid for the application and must provide results that are credible and reliable. These processes ensure validity, credibility, and reliability of the model outputs.

- ***"Model verification*** is defined to be the process of determining if the rationality that describes the underlying procedure of the model, is accurately captured by the software code". Therefore, the model verification process proves if the corresponding computer program produces the anticipated outputs in terms of accuracy and magnitude. Regardless of the validity of the logic or the theory used in the program.
- ***"Model validation*** is considered to be the process of determining to what extent the model's underlying fundamental rules and relationships are able to adequately capture the targeted emergent behavior, as specified within the relevant theory and as demonstrated by field data." Therefore, delay, throughput, and queues length calculations logic and equations utilized by the developed model should produce the corresponding delay, capacities, throughput, and queues lengths.

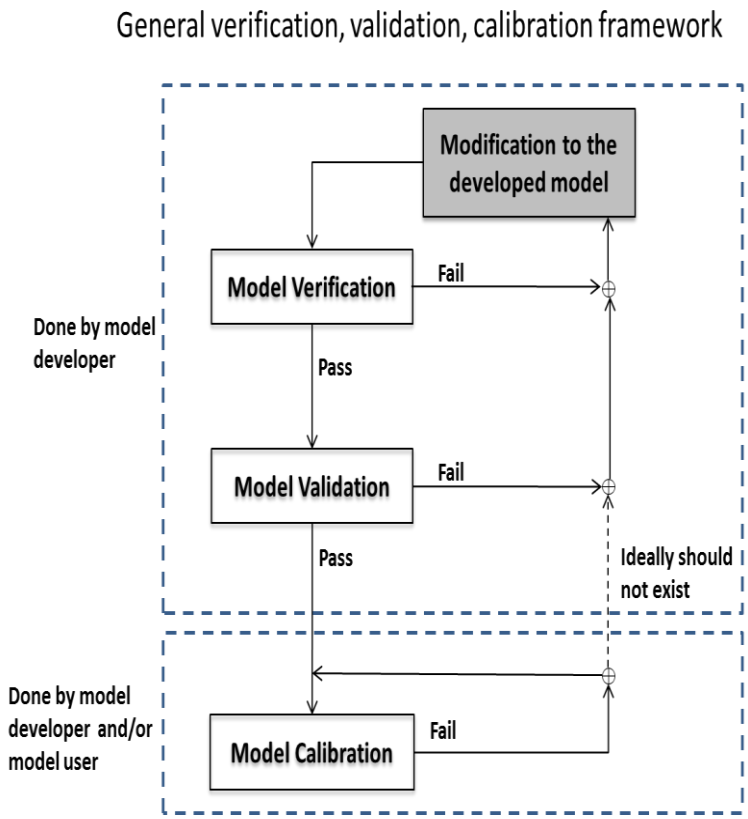


Figure 44: General verification, validation and calibration framework, Rakha, et al., [91]

The processes of verification and validation of a particular model must be carried out sequentially as they interact with each other. First, the verification process where the models codes are tested and examined to be running free of errors is performed. Second, the model validation process is commenced once model verification that been successfully completed. Model validation involves comparing the developed model's outputs with generated analytical solutions (i.e., pre-validated method). The failure in validation requires that some modifications be made to the model logics and equations, and consequently, the verification process must be repeated. This iterative sequences should be is performed until the validation process is successful. This iterative sequence is the responsibility of the model developer. Rakha, et al.,[91]

7.10 OTS Validation: Control delays calculations

The developed OTS was validated using HCM example of "extension of signal delay models to incorporate effect on an initial queue" (HCM 2000, P 16-146) [90]. The HCM example calculates the total delay of an intersection approach that have four time intervals with volume of (800, 1200, 1000, 600 vehicle/hour) with constant approach capacity of 1000 vehicle/hour. The following figure shows the demand profile of the analyzed intersection approach. The outputs of HCM example are depicted in the following figure. The figure shows the estimated delay components of control delay (i.e., uniform delay, random and oversaturation delay, and initial queue delay).

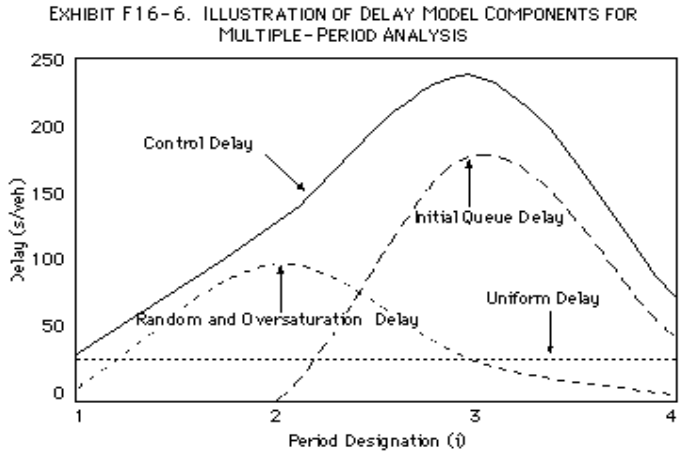


Figure 45: HCM's example; control delay components outputs

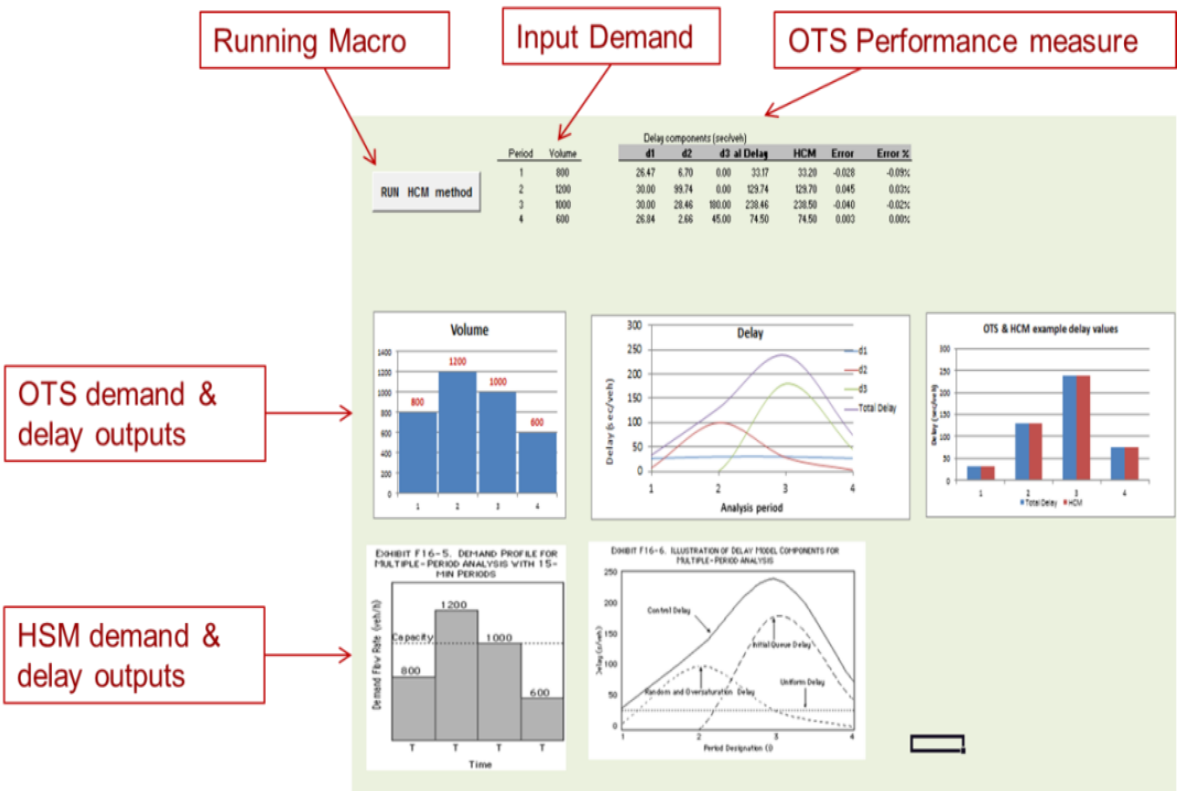


Figure 46: OTS outputs for HCM "multi period delay calculation" example

OTS was executed for the same volume profile and capacity (i.e., green time = 40 sec, cycle = 100 sec, saturation rate = 2500 veh/hr) that used in HCM example. Figure 46 illustrates the components control delay output of OTS as function of demand variation. Figure 47 and Table 7 illustrate the good fit between the OTS estimates and HCM calculations, all estimates were less than 0.03% error. Therefore, the OTS delay estimates follow the HCM delay calculation for multi interval demand.

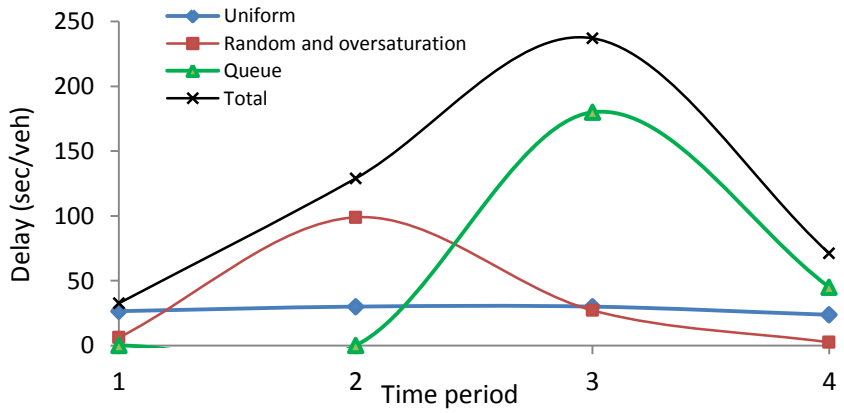


Figure 47: OTS's control delays outputs

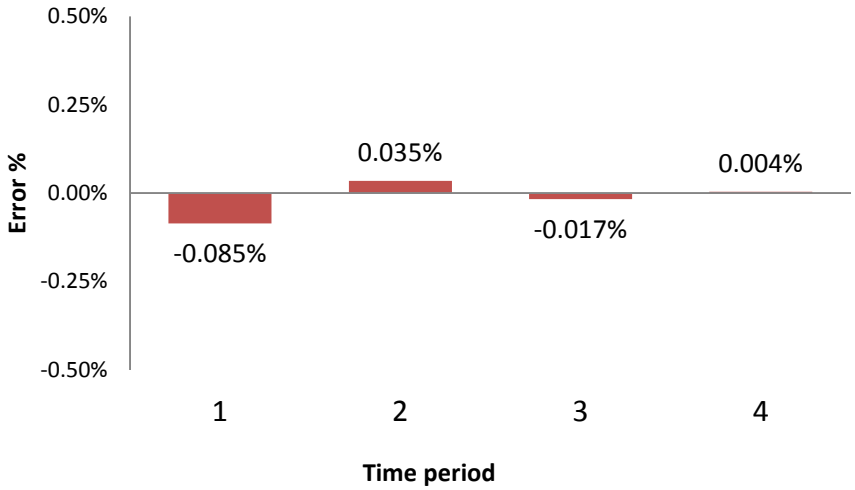


Figure 48: **OTS control delay error percentage**

Table 7: OTS and HCM delay control and the corresponding error

Time Period	Delay (sec/ veh)		Model Error	
	HCM	OTS	error	error %
1	33.17	33.20	-0.028	-0.09%
2	129.74	129.70	0.045	0.03%
3	238.46	238.50	-0.040	-0.02%
4	74.50	74.50	0.003	0.00%

8 Case Study: Arterial network (The Reston Parkway, Herndon, VA)

8.1 Introduction

In Chapter 6 we present the fundamental concepts of the signal timing framework during oversaturated conditions. The principles include cycle length calculation, green splits calculation, and oversaturation offsets design. This chapter describes the implementation of the proposed framework in a congested arterial network in northern Virginia. The case study allows us to investigate two of the previous stated hypotheses:

- "Identifying and protecting network's critical routes from becoming oversaturated would significantly improve system performances"
- "During oversaturated condition, achieving optimality in signal operation requires considering system dynamics in strategies selection, control objectives and performance measures"
- For oversaturated network, considering *different routing scenarios* in designing timing plan will result in adopting different control strategies that need to be evaluated using different performance measures.

First, network critical movements/routes were identified. Based on that, several routing scenarios were established. Then, a set of the oversaturation control strategies were considered to address the problematic scenarios. The applied mitigation strategies include the following:

- Oversaturation offsets
- Phase re-service
- Gating / metering

Then, the proposed oversaturation-timing framework was used to generate optimal \ timing plans, and VISSIM was used to evaluate the timing plans. Next, a multi-objective analysis was performed to eliminate the dominated timing plans and to assist in selecting the appropriate control objectives and their corresponding control strategies.

8.2 Reston Parkway Network

Reston Parkway arterial network is located in Northern Virginia, USA. The network is significantly congested during peak periods as the arterial intersects with the heavily traveled Dulles Toll Road. The network consists of fourteen intersections with a total length of 16,572 ft. (3.1 miles). The spacing between intersections ranges from 524 feet to 3,309 feet. The speed limit for the main arterial is 45 mph. Side streets speed limits range from 15 mph to 45 mph. The volume counts data was provided from NRO, The Northern Region Operation of Virginia Department of Transportation (VDOT), which operates the traffic signals system in Reston Parkway network under VDOT's purview. The current system employs *170* controllers and the central software, Management

Information System for Transportation (MIST). Actual detector data from this network was collected for a period of one month starting from *August 11th, 2009 to September 10th, 2009*. These detectors almost cover the whole network. MIST compiles readings that were aggregated every 15-minute interval. During the evening peak period (2:30 P.M. to 8:00 P.M.) the arterial operates in a coordinated fashion during the day using traditional progression-based offsets. The arterial network is operated with three different timing plans.

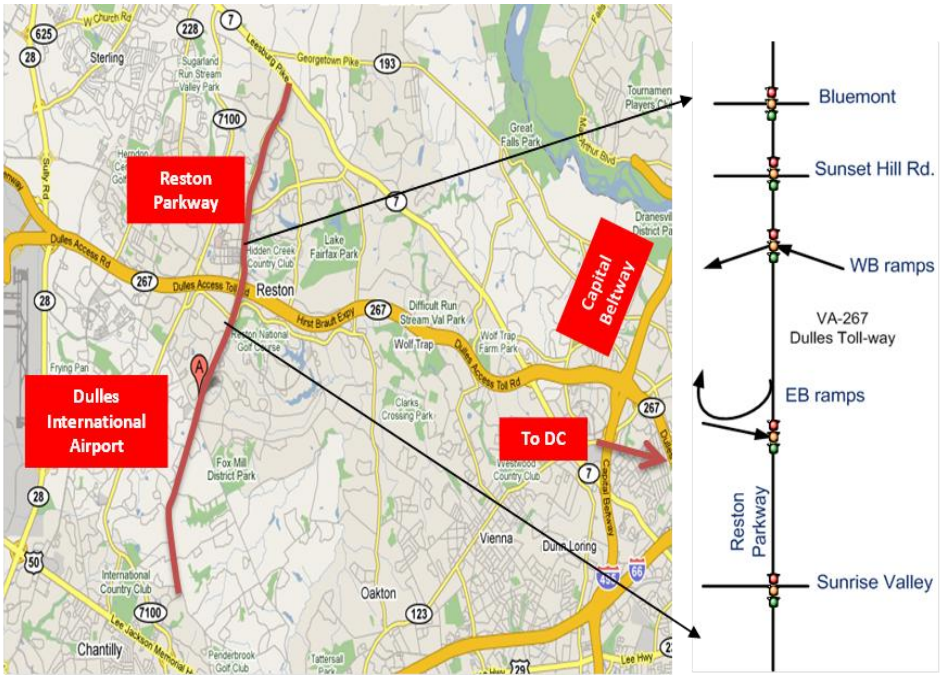


Figure 49: Reston Parkway Network

Table 8: Reston Parkway timing plans during P.M. peak period

Timing Plan	Cycle length (sec)	Applied Time
1	130	2:30 – 3:00
2	180	3:00 – 6:00
3	140	6:00 - 2:30

8.2.1 *Calibration data*

The signal setting information for Reston Parkway network was provided by NRO. The NRO staff gratefully provided us with SYNCHRO files that contain not only the signal timing plans but also the parameters they used to generate timing plans. These timing parameters include peak hour factor, lane utilization, saturation rates, and pedestrian parameters plus the schedule of timing plans through the weekday and weekend. Another set of calibration data was collected from the field during several visits to Reston Parkway network area. The first visit was on *Wednesday March 11th, 2009*. Then we visited the area from *Monday June 1st to Wednesday June 3rd, 2009*. During these visits, a GPS-

instrumented vehicle was driven in the Reston Parkway network from north of Barron Cameron Avenue to the south of Mclean Road. Routes travel time and link traveling speed data were collected using *average-car* technique where the driver was instructed to drive at the approximate average speed of the traffic stream. The collected filed data include the following:

- Travel time
- Queue lengths at critical intersections
- Saturation flow rates at critical intersection
- Traffic patterns

8.2.2 *Traffic pattern in Reston parkway*

During the P.M. peak period, traffic patterns change rapidly in Reston Parkway arterial network. There is a high turning volume into and from the heavily congested Dulles Toll-way ramps. **Figure 49** and **Figure 51** show the changes in arterial's NB/SB traffic volumes at different intersections in the network during a P.M. peak period.

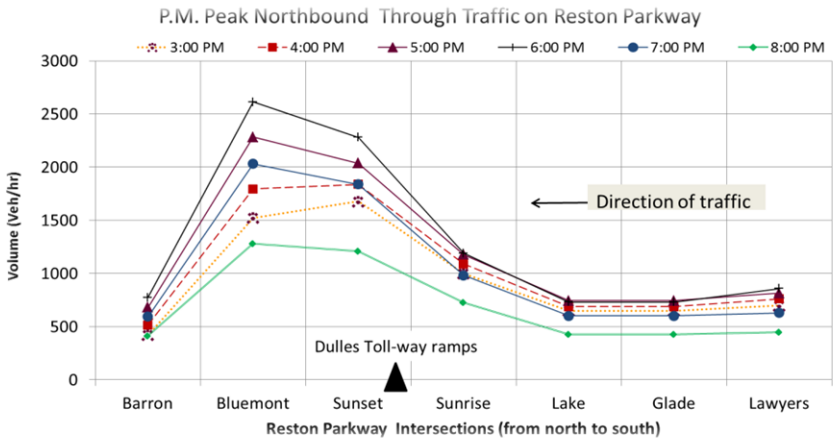


Figure 50: Reston Parkway P.M. Northbound Traffic

Figure 50 shows the northbound through-traffic on Reston Parkway arterial increases rapidly just after Sunrise Valley Rd. intersection, then drastically drops at the Bluemont Rd. intersection. The figure indicates a high turn-on volume between Sunrise Valley Rd, and the Bluemont Rd. intersections. The turn-on volume from Dulles Tollway interchange ramps and Sunset Hill Rd. contribute significantly to the volume spike in northbound traffic during a P.M. peak period. This particular situation cannot be address via traditional arterial timing schemes that prompt arterial through movements' progression. The network's highest volume of 2600 veh/hr was observed at 6 P.M. at the Bluemont way intersection. This volume quickly drops to around 2000 veh/hr at 7 P.M. and then to nearly 1400 veh/hr at 8 P.M. The northbound volumes at the intersections Lake, Glade, Lawyers are between (500-900) veh/hr during the entire P.M. peak period

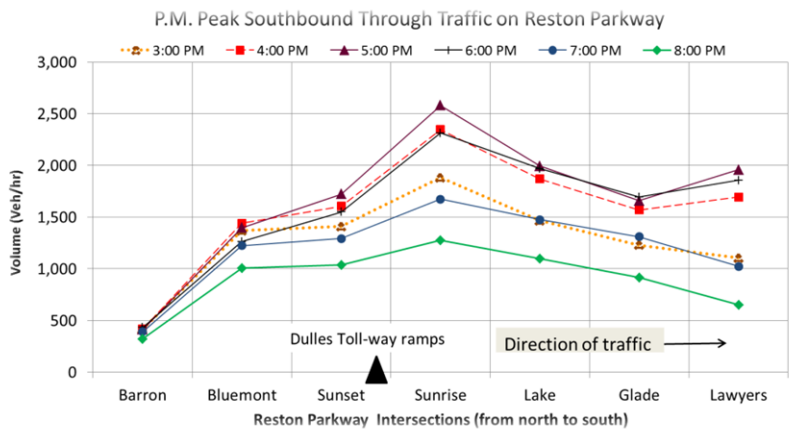


Figure 51: Reston Parkway P.M. Southbound Traffic

Figure 51 shows southbound through traffic on Reston Parkway arterial increases in a gradual fashion unlike the northbound through-traffic. The heaviest volume in the network is observed around 5:00 P.M. at the Sunrise Valley intersection (i.e., the study network exit intersection). Network southbound peak is observed earlier than northbound approaches at 5:00 P.M. During this time, 5:00 P.M, at the Sunset Hill Rd intersection the southbound through-traffic is nearly 1,700 veh/hr whereas northbound through-traffic is almost (2,400 veh/hr).

8.2.3 *Building Volume scenarios*

Network routing scenarios were established based on volume analysis of the network observed volume. The volume analysis includes: correlation analysis, cluster analysis, and pattern reconnection techniques for the movements volumes. For each considered routing scenario, background traffic volume was determined and then critical routes volumes were maximized to match the actual observed volume. Table 9 and Figure 52 show six critical routing scenarios in the Reston Parkway network. The first three scenarios are the basic arterial routing arrangement that consist of northbound, southbound, and north/southbound traffic. The other three scenarios contain more routes as counts data and field observations indicate.

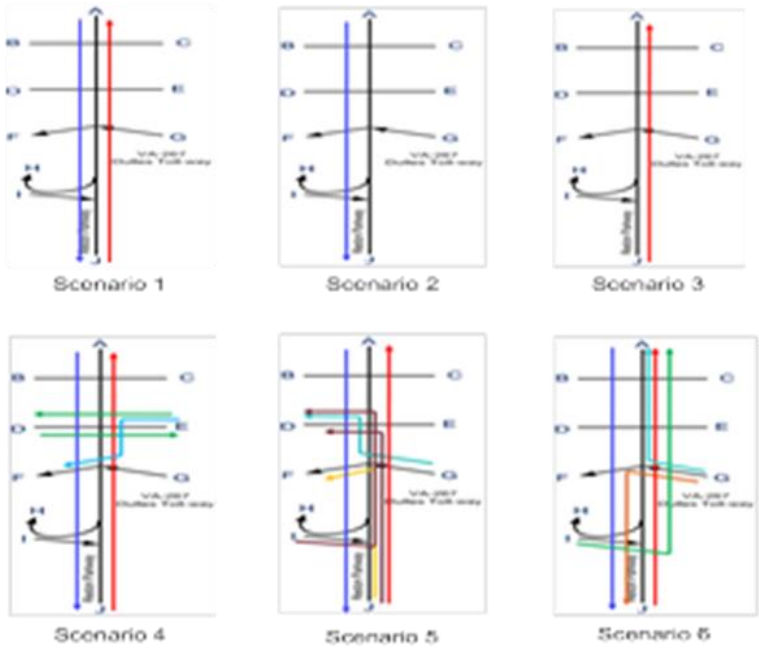


Figure 52: The Reston Parkway Network Critical Routes Scenarios

Table 9: Reston Parkway network critical routes scenarios

Scenarios	Network critical routes description
Scenario 1	1. NB through on Reston Parkway 2. SB through on Reston Parkway
Scenario 2	1. SB through on Reston Parkway
Scenario 3	1. NB through on Reston Parkway
Scenario 4	1. NB through on Reston Parkway 2. SB through on Reston Parkway 3. EB through on Sunset Hill road 4. WB through on Sunset Hill road 5. WB on Sunset Hill road, left into SB on Reston Parkway and right into Dulles WB on-ramp
Scenario 5	1. NB through on Reston Parkway 2. SB through on Reston Parkway 3. NB through on Reston Parkway then left into Dulles WB on-ramp 4. NB through on Reston Parkway then left into WB Sunset Hill road 5. Left from Dulles EB off-ramp into NB Reston Parkway and left into WB Sunset Hill 6. Right from Dulles WB off-ramp into NB Reston Parkway and left into WB Sunset Hill road
Scenario 6	1. NB through on Reston Parkway 2. SB through on Reston Parkway 3. NB through on Reston Parkway then left into WB Sunset Hill road 4. Left from Dulles EB off-ramp into NB through Reston Parkway 5. Right from Dulles WB off-ramp into NB through Reston Parkway 6. Left from Dulles WB off-ramp into SB through Reston Parkway

Problematic Scenario # 5

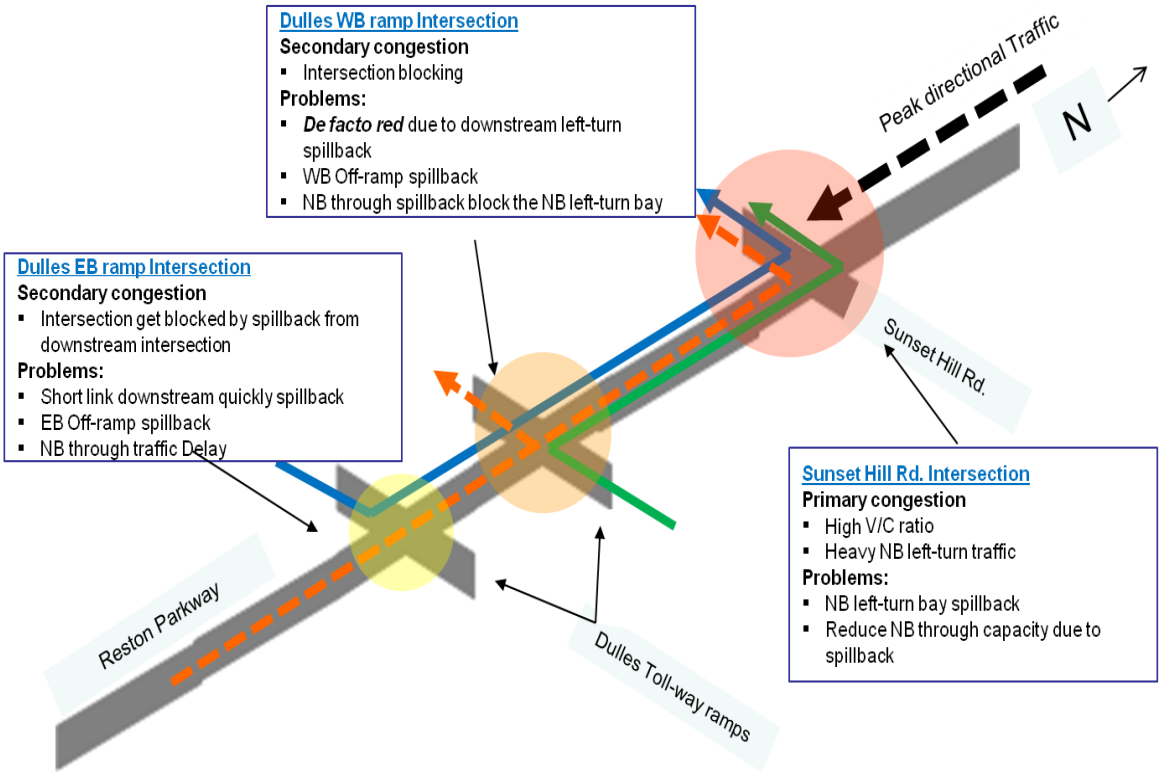


Figure 53: Problematic Scenario 5 critical movements and congestion diagnosis

Figure **53** illustrates one of the network problematic scenarios, scenario number 5. Critical routes, critical movements, congestion attributes, and proposed control strategies were presented in the Figure **53**. Then, each of the proposed strategies was further discussed in detail for the expected improvements and drawback. The intent of this specific example is to demonstrate how mitigation control strategies were developed based on multilevel (i.e., network-wide, tactical, intersection) considering the congestion attributes. Figure 53 illustrates four of the critical routes at the interchange of the Dulles Toll Road with the Parkway. These routes are considered critical during P.M. peak periods. In this scenario, the traffic turning left into Reston Parkway NB from the eastbound off-ramp of the Dulles Toll Road has limited storage between the interchange ramps—in this particular case, the two ramp junctions are operated as independent intersection. At the next intersection (Dulles Toll Road WB ramps) a significant amount of traffic turns left. This traffic should not be stopped at EB ramps intersection, if possible, due to the limited storage capacity between the interchange ramps intersections. The traffic should progress through Dulles Toll Road WB ramp intersection and then left turn into Sunset Hill intersection. Thus, the left turn phase is perhaps a better choice for "coordination" of the movements at the three consecutive intersections rather than the Northbound

through phase at the Northern-most intersection. In addition, there is a similar critical route coming from the Dulles EB off-ramp of the toll road that also turns right into Reston Parkway NB and then turns left at the same intersection (Sunset Hill Rd.) that has a limited left-bay storage. This route is shown in blue in Figure **53**. In particular, these two route flows arrive at the left-turn of the Sunset Hill Rd. intersection at different times in the cycle. This is where determination of the critical routes and dynamic mapping of the growth and dissipation of congestion demonstrates its value. In this scenario, we find that perhaps ***phase re-service*** combined with *gating* traffic upstream of the Dulles Toll Road EB ramp intersection, along with appropriate settings for *offsets*, could result in more efficient processing of the critical route flows and therefore mitigation of the oversaturation that results with normal, under-saturated operational strategies. Mitigation strategies and combinations of specific mitigation strategies will be described in more detail in following section.

Table 10: Reston Parkway network Scenario 5 oversaturation attributes

Extent	Duration	Causation	Recurrence	Symptoms
Movement	Situational	**Signal Timing**	**Recurrent**	**Starvation**
Approach	Intermittent	**Geometrics**	non-Recurrent	**Spillback**
Intersection	**Persistent**	Other modes		Storage Blocking
Route	Prolonged	**Demand**		Cross Blocking
One-way arterial		Planned Events		
Two-way arterial		Unplanned Events		
Interchange				
Grid				
Network				

Control Strategy 1: Increase the critical route capacity downstream (Phase Re-service)

Control Strategy

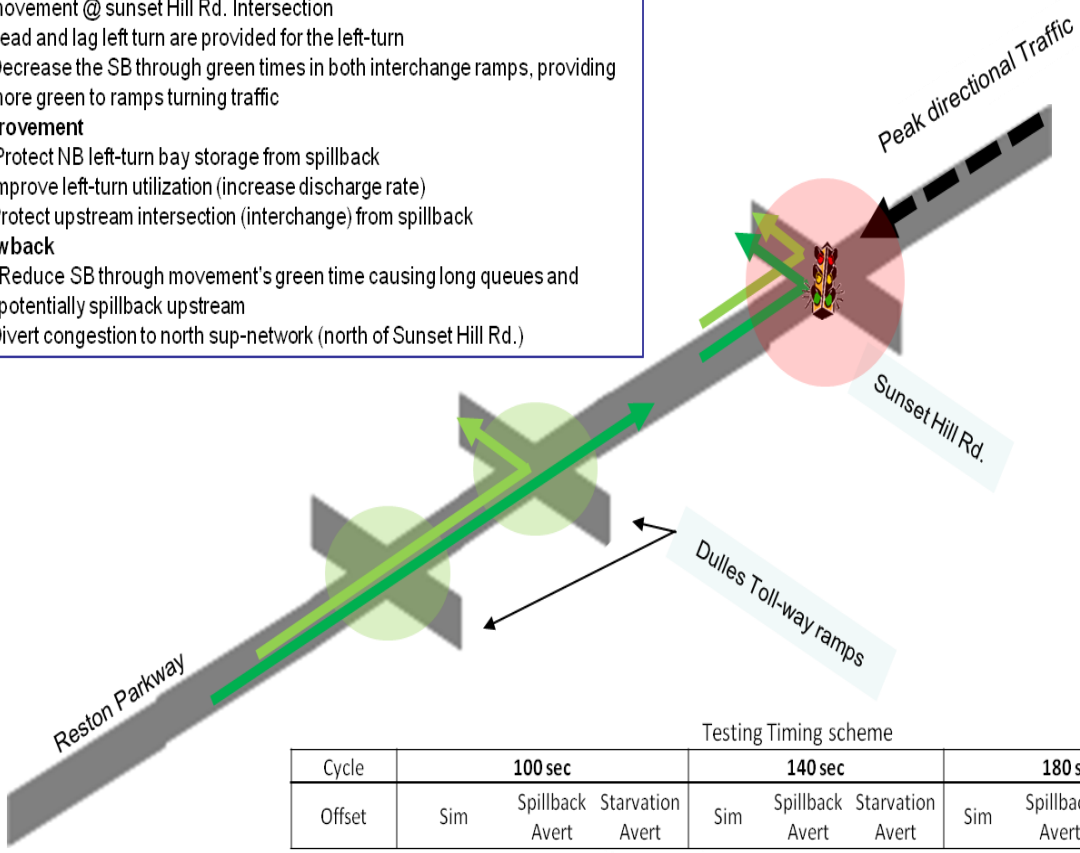
- Increase NB left-turn capacity by provide two phases for the left-turn movement @ sunset Hill Rd. Intersection
- Lead and lag left turn are provided for the left-turn
- Decrease the SB through green times in both interchange ramps, providing more green to ramps turning traffic

Improvement

- Protect NB left-turn bay storage from spillback
- Improve left-turn utilization (increase discharge rate)
- Protect upstream intersection (interchange) from spillback

Drawback

- Reduce SB through movement's green time causing long queues and potentially spillback upstream
- Divert congestion to north sup-network (north of Sunset Hill Rd.)



Testing Timing scheme

Cycle	100 sec			140 sec			180 sec		
Offset	Sim	Spillback Avert	Starvation Avert	Sim	Spillback Avert	Starvation Avert	Sim	Spillback Avert	Starvation Avert

Figure 54: Scenario 5 control strategies 1, Phase re-service at Sunset Hill Rd. intersection

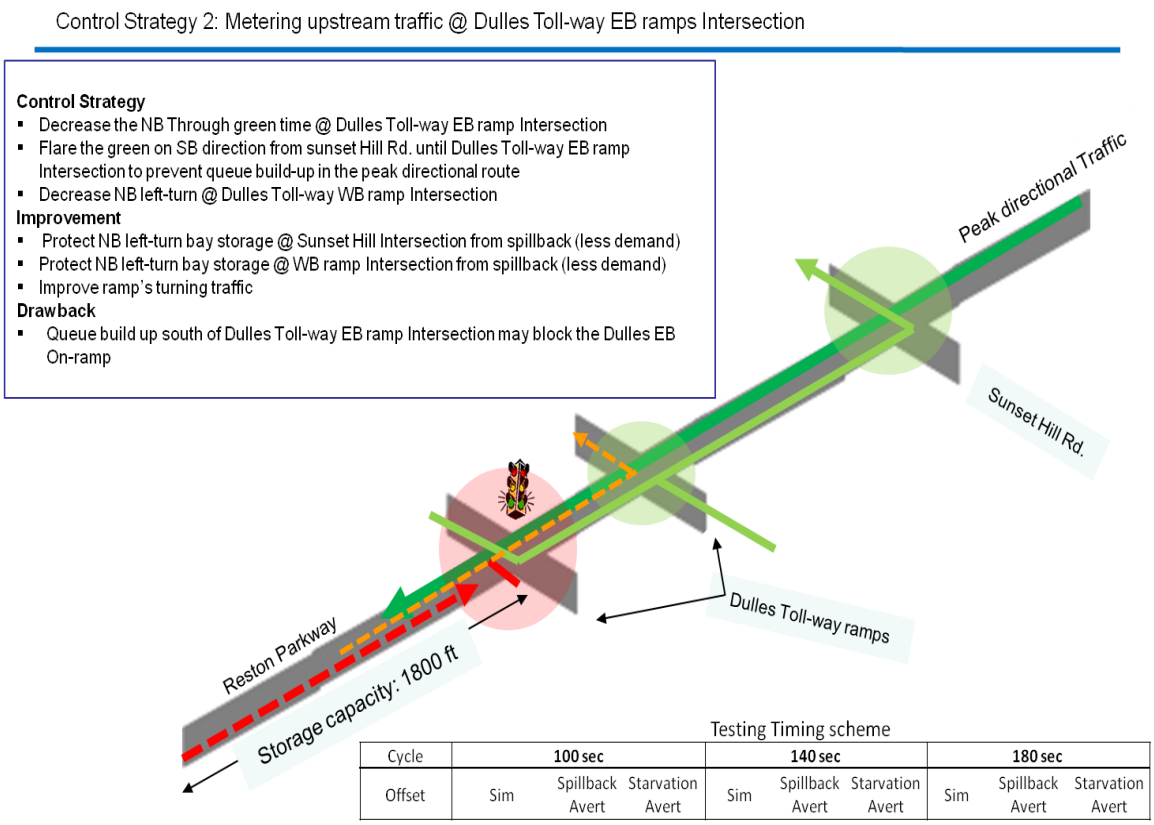


Figure 55: Scenario 5 control strategy 2, Metering at Dulles Toll road EB off- ramp intersection

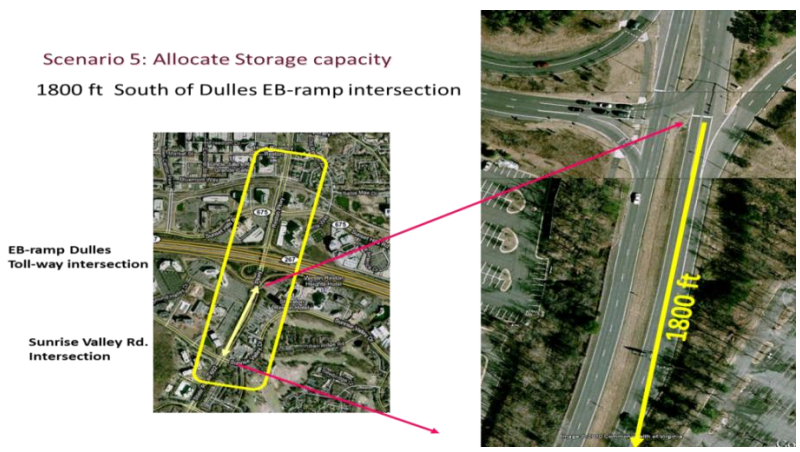


Figure 56: Metering intersection, gated link and the storage capacity in Reston Parkway Network

Figure 56 shows the queuing capacity available south of the Toll Road EB off-ramp intersection (1800 ft.). The link, NB between Sunrise Valley Rd. intersection and Tool Road EB-ramp intersection, is suitable to be assigned as gated-link in the network. The NB traffic was metered at this location to provide extra capacities to the off-ramps (EB and WB ramps).

Table 11: The operational objectives of the applied control strategies

Control Strategy	Expected Improvement	Operational Objective
Phase re-service	Prevent storage bay spillover Improve left-turn green utilization	Queue Management
Simultaneous offset	Clear residual queue Improve green utilization Accommodate bi-directional critical routes	Queue Management
Negative progression	Accommodate turning traffic Clear residual queue Improve green utilization (Prevent starvation)	Queue Management
Forward progression	Improve arterial progression	Stops minimization Delay minimization
Long Cycle	Minimize loss time Improve arterial progression	Stops minimization
Short Cycle	Prevent spillback Improve green utilization Provide equitable service	Throughput maximization Queue management
Gating /metering	Prevent congestion spread-out Offset congestion	Queue Management

8.2.4 *Timing Plans developments*

After determining the proper mitigation strategies, signal timing plan parameters (i.e., cycle length, splits and offset) were generated accordingly. Three cycles length values (i.e., 180, 140 and 100 seconds) were selected to be evaluated in this case study representing long, medium and short cycle length. For each selected cycle length, the initial splits were determined based on v/c ratios, then green splits were adjusted based on the critical approach storage capacity with consideration to the applied control. Then oversaturation offsets (i.e., spillback-avert and starvation-avert splits) were calculated according to expected queue length, link storage capacity, vehicle operation characteristics, and shockwaves parameters. The following acronyms are used to represent oversaturation offsets in the next sections.

- ***Sim***: Simultaneous (zero) offset
- ***Max***: Offset that prevents starvation at downstream intersection
- ***Min***: Offset that prevents spillback at upstream intersection (de facto red)
- ***Med***: medium value between Max and Min

Figure 57 shows an example of oversaturation offsets values (i.e., Min & Max) for both NB and SB traffic. These offsets values were developed from optimized splits for a aprticular cycle length. In Figure 57, the feasible range of oversaturation offsets values of the northbound movement are mainly a forward progression. In contrast, southbound movement offsets values are mostly reverse progression.

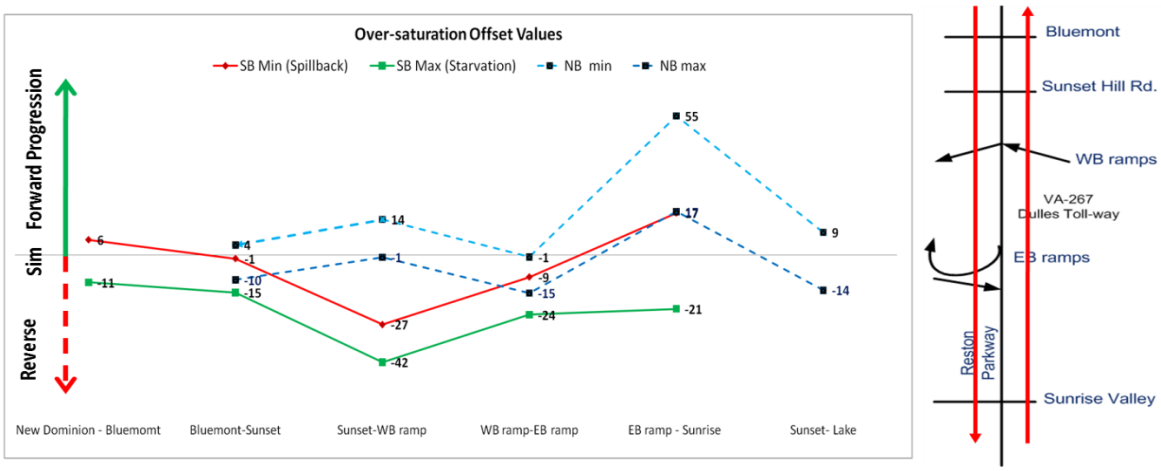


Figure 57: Oversaturation offsets on Reston Parkway

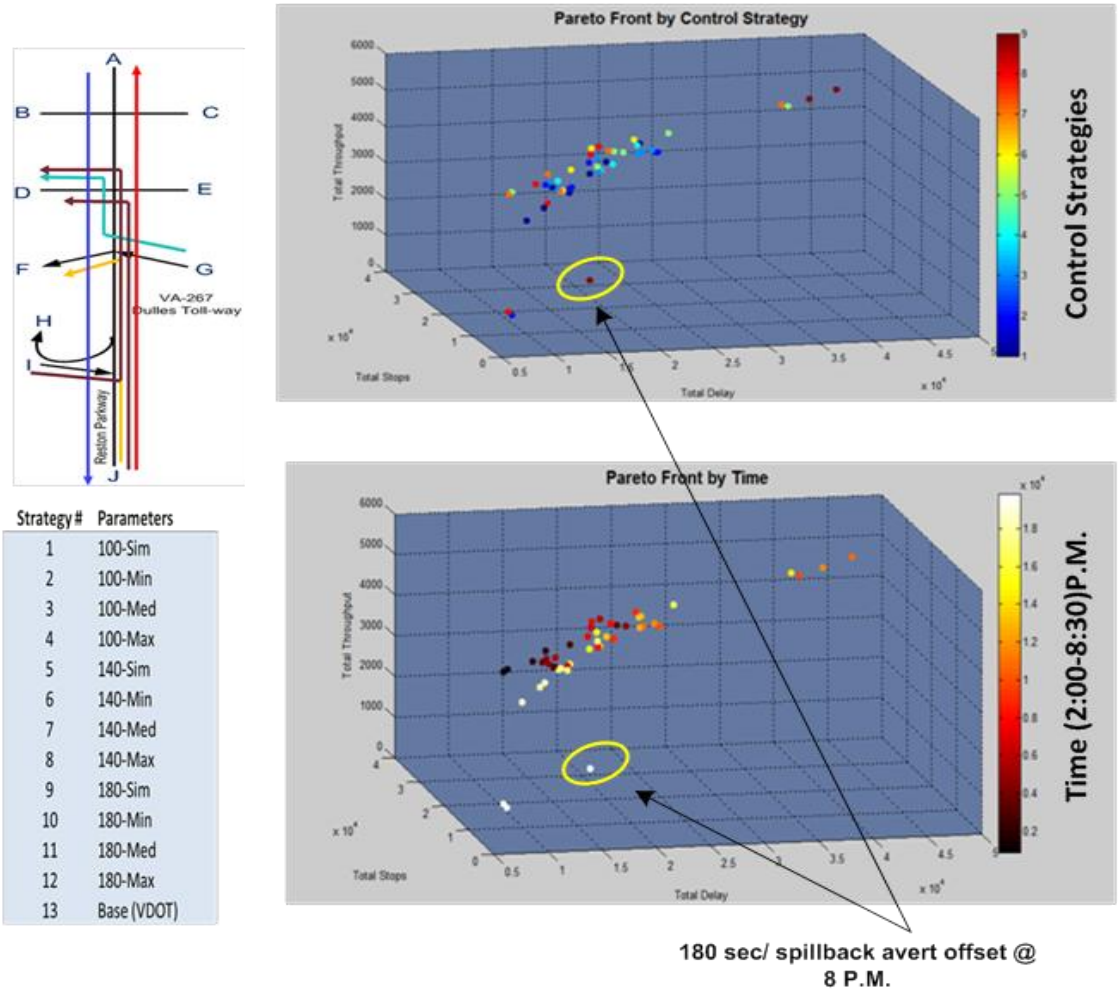
The following table illustrates the control strategies and timing plans for scenario 5. Twelve control strategies were developed for phase re-service strategy and another twelve for metering control strategy.

Table 12: scenario 5 timing plans and applied control strategies

Scenario	Strategy	Control Strategy	Applied Location	Action	Cycle Length (sec)	Offset Design	Split
5	1	Phase Re-service	NB Left-turn Sunset Hill Rd. Intersection	Double the left-turn phase (lead & Lag)	100	Simultaneous	based on V/C ratio with priority to Scenario 5 critical routes
	2					Spillback Avert	
	3					Medium	
	4					Starvation Avert	
	5				140	Simultaneous	
	6					Spillback Avert	
	7					Medium	
	8					Starvation Avert	
	9				180	Simultaneous	
	10					Spillback Avert	
	11					Medium	
	12					Starvation Avert	
	13	Base timing plan (VDOT)			140, 180, 130	Progression offset	v/c ratio
5	1	Metering	NB Through @ EB ramp Intersection	Reduce the NB through green time by 20%	100	Simultaneous	based on V/C ratio with priority to Scenario 5 critical routes
	2					Spillback Avert	
	3					Medium	
	4					Starvation Avert	
	5				140	Simultaneous	
	6					Spillback Avert	
	7					Medium	
	8					Starvation Avert	
	9				180	Simultaneous	
	10					Spillback Avert	
	11					Medium	
	12					Starvation Avert	
	13	Base timing plan (VDOT)			140, 180, 130	Progression offset	v/c ratio

8.3 Simulation Result and Pareto front Analysis

VISSIM simulation output results were collected every 15-min including network total delay, total number of stops, and total throughput. Figure 58 shows the dominating timing plans performance measurements in a 3D temporal diagram. These diagrams directly demonstrate the temporal optimality of each tested strategies. For example, for metering scenarios, longer cycles perform better during the end of peak period (recovery period)). On the other hand, for Phase re-servicing, shorter cycles dominate both build-up and recovery periods. Regarding throughput only, shorter cycles perform poorly at the beginning of peak period, then system throughput start to improve as demand increases in the network, toward the end of peak period the shorter cycles return to their poor performance. These innovative concepts of temporal diagrams convey invaluable information to the decision makers regarding system performance. They also assist analysts to build a combination of an effective control strategy that meet agency control objectives. Then, the selected strategy/strategies can be evaluated and the average performance measurements improvements during the analysis period can be obtained

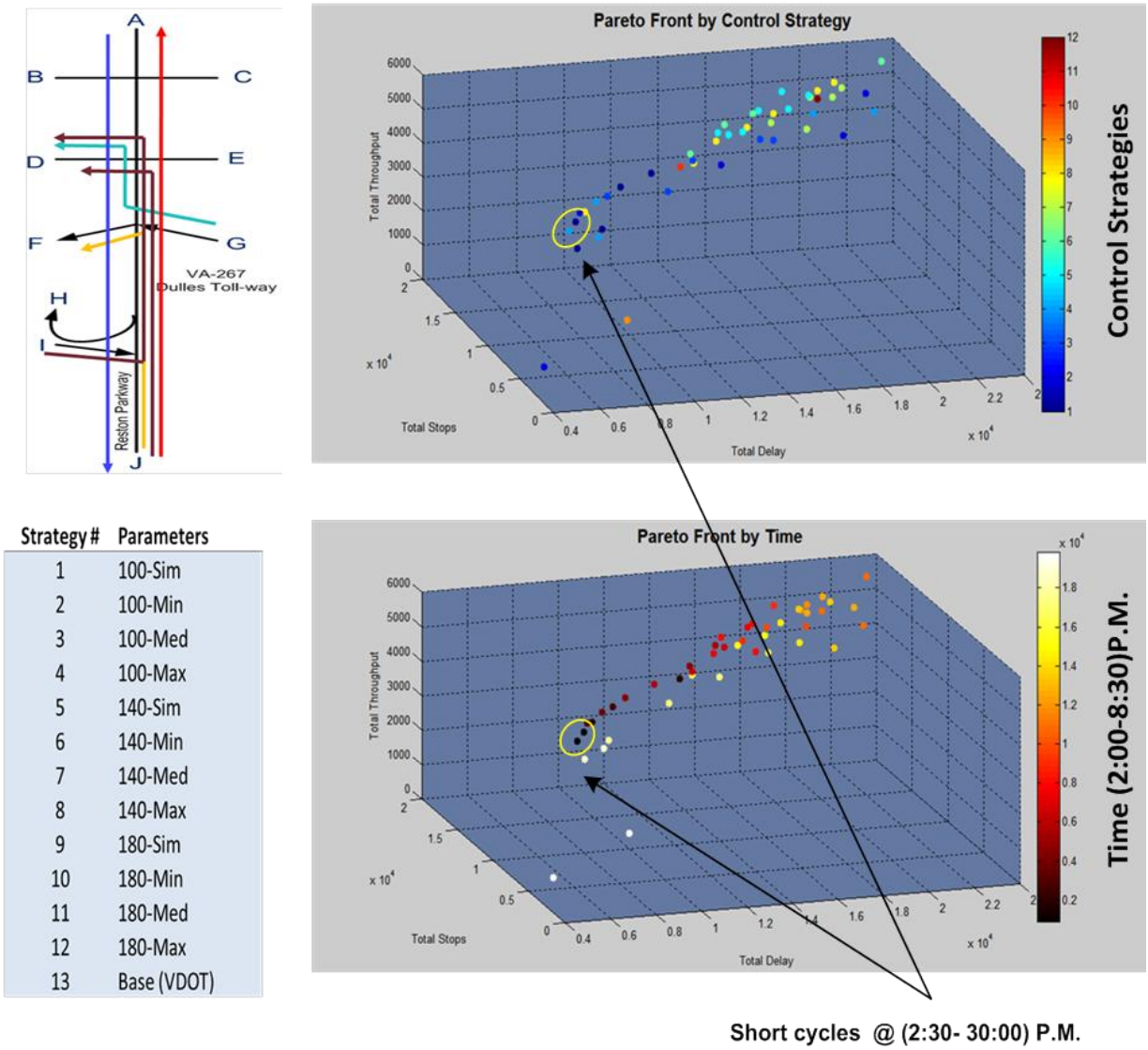


Strategy #	Parameters
1	100-Sim
2	100-Min
3	100-Med
4	100-Max
5	140-Sim
6	140-Min
7	140-Med
8	140-Max
9	180-Sim
10	180-Min
11	180-Med
12	180-Max
13	Base (VDOT)

Figure 58: Scenario 5 with Phase Re-service: Network performance measures of the optimal control strategies

Routing Scenario (5) with phase re-service control findings:

- Short cycle lengths and mid offsets minimize delay
- Mid cycle lengths and mid offsets maximize throughput
- In general, Phase re-service control strategy reduces throughput
- Throughput maximization strategies reduced delay by 29% and reduced throughput by 1%
- Delay minimization strategies reduced delay up to 29% and reduced throughput by 1%



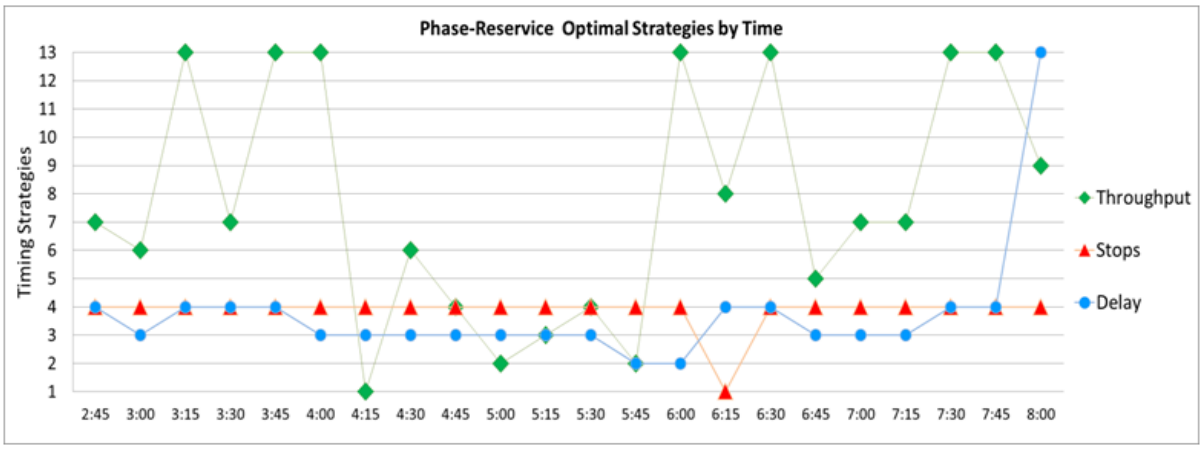
Strategy #	Parameters
1	100-Sim
2	100-Min
3	100-Med
4	100-Max
5	140-Sim
6	140-Min
7	140-Med
8	140-Max
9	180-Sim
10	180-Min
11	180-Med
12	180-Max
13	Base (VDOT)

Figure 59: Scenario (5) with Metering: Network performance measures of the optimal control strategies

Routing scenario (5) with metering control findings:

- Short cycles with min/mid offset minimize delay and stops
- Mid cycle length maximize throughput
- Longer cycle length increases the delay significantly
- Throughput maximization strategies reduced delay by 13% and increased throughput by 11%
- Delay minimization strategies reduced delay by 35% and increased throughput by 7%

The following figures show the optimal timing control plan for each time interval during the peak period. The steadiness of optimal timing plan can be considered in the evaluation process, as shifting between time plans during peak period may induce undesirable transition effects. For phase re-service operation, delay and stops minimization optimal strategies appear to be stable during most of the peak period (Strategies 3 & 4), while throughput maximization optimal strategies shows instability as Figure60(a) illustrates. For metering control strategy, Delay minimization & throughput maximization optimal strategies appear to coincide with each other most of the peak period (shorter cycles). Regarding timing plan steadiness, plan 3 and 4 appear to be the best of the tested timing plans as Figure60 (b) illustrates.



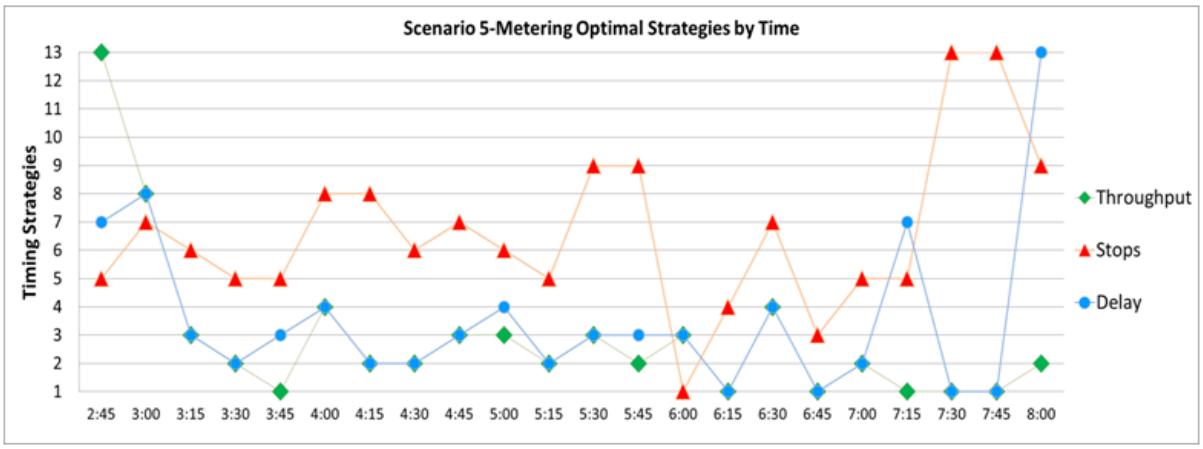


Figure60: Scenario 5 Optimal Control Strategies Time-of-peak period for performance measures (Delay, Stop, and Throughput)

Finally, the sub-optimal timing plans were evaluated for each 15-min interval with respect to the current implement timing plan (i.e., VDOT timing plans). The percentage of performance measures improvements were presented in **Table 13** for individual strategy.

Table 13: Scenario 5 control strategies Total Improvement Percentage during Peak Period

Strategy	Cycle (sec)	Delay	Stops	Throughput
Phase Re-service				
1	**100**	26%	45%	-1%
2		26%	45%	-1%
3		29%	42%	-1%
4*		29%	60%	-1%
5	**140**	-17%	7%	-1%
6		-13%	22%	-1%
7		-14%	15%	-1%
8		-9%	18%	-2%
9	**180**	-117%	-25%	-8%
10		-142%	-19%	-12%
11		-90%	-29%	-7%
12		-122%	-29%	-41%
Metering				
1	**100**	35%	32%	7%
2		35%	29%	7%
3		34%	23%	8%
4		33%	14%	6%
5	**140**	6%	17%	10%
6		17%	32%	10%
7*		13%	25%	11%
8		14%	22%	9%
9	**180**	-18%	-17%	11%
10		-15%	-13%	8%
11		-24%	-20%	6%
12		-24%	-9%	6%

*Optima timing plans

8.4 Statistical analysis

A standard t-test was performed to determine if there was any different in the performance measures between VDOT timing plan and the proposed control strategies. The null hypothesis in this context could be that the mean of VDOT's timing plan performance measurements are equal to that of the proposed timing plan performance measurements. The t- hypothesis tests with 95% confidence level were examined performance measures, total delay and total number of stops, for unequal variances. The hypothesis test was mean total delay difference between VDOT plan and the proposed timing plan is not equals to zero. Each pair would have equal mean if t-statistical < t-critical. The results show significant different between the values of delay and number of stops as measures by the t-test

at a 95% confident level. The statistic summary of t-tests for all locations was described on the following table.

Table 14: t-Statistical test for the phase re-service control strategies' performance measures

Time Plan	Cycle length		Mean		Error		t-statistics (95%)	
	sec	offset	Delay (hr)	Stops	Delay	Stops	Delay	Stops
1	100	sim	2,361.8	116,536	5.2%	8.5%	-18.7	-21.6
2		min	2,361.8	116,536	6.7%	7.8%	-17.3	-21.9
3		med	2,266.0	122,893	6.3%	2.2%	-20.0	-21.9
4		max	2,266.0	84,754	7.5%	5.9%	-18.8	-30.8
5	140	sim	3,734.1	197,052	7.8%	3.4%	8.2	-3.5
6		min	3,606.5	165,269	6.5%	2.5%	7.2	-11.4
7		med	3,638.4	180,101	3.2%	5.7%	10.2	-7.1
8		max	3,478.8	173,745	7.7%	3.2%	4.6	-9.2
9	180	sim	3,734.1	264,855	7.2%	3.1%	8.7	12.3
10		min	4,532.0	252,142	5.5%	2.6%	22.5	9.6
11		med	6,064.0	273,330	7.3%	2.7%	32.3	14.5
12		max	3,893.7	273,330	9.0%	4.6%	9.4	13.3
VDOT (Base)	140-180-130		3,191.6	211,884	6.6%	10.4%	-	-

Table 15: t-Statistical test for the metering control strategies' performance measures

Time Plan	Cycle length		Mean		Error		t-statistics (95%)	
	sec	offset	Delay (hr)	Stops	Delay	Stops	Delay	Stops
1	100	sim	2393.684	144,081	6.6%	6.3%	-16.55	-15.57
2		min	2393.684	150,437	12.3%	7.4%	-12.05	-13.61
3		med	2106.442	163,151	5.3%	6.8%	-24.92	-10.82
4		max	2138.358	182,220	8.6%	8.8%	-20.66	-5.95
5	140	sim	3000.084	175,864	6.6%	7.0%	-3.63	-7.82
6		min	2649.011	144,081	9.9%	8.4%	-8.82	-14.75
7		med	2776.674	158,913	8.0%	6.7%	-7.40	-11.85
8		max	2744.758	165,269	5.5%	12.4%	-9.42	-8.49
9	180	sim	3766.063	247,904	11.7%	7.7%	6.44	6.77
10		min	3670.316	239,429	6.9%	8.3%	7.96	5.09
11		med	3957.558	254,261	7.9%	8.1%	11.17	7.71
12		max	3957.558	230,953	4.2%	8.3%	15.63	3.58
VDOT (Base)	140-180-130		3,191.6	211,884	6.6%	10.4%	-	-

8.6 Lessons learnt and guidance from Reston Parkway Case Study

The following is a list of major findings from the Reston Parkway Study:

- Critical route movement determination is essential for determination of optimal strategies
- Cycle length should be determined such as to accommodate maximum queues
- Offsets should be determined within a range to prevent both spillback and starvation
- Queues are to be considered as constraints in the optimization problem; throughput and delays are objectives
- During the network loading, optimal strategies maximize throughput and during recovery, optimal strategies minimize delay
- High cycle lengths are associated with higher delay without increasing throughput
- More guidance is needed to understand the relationship between different objectives and traffic scenarios
- A third phase of the research should extend the scope to include real-time route volume determination and scenario identification
    - Maximum likelihood determination of scenarios
    - Accounts for system dynamics in optimal control

8.7 Summary and Conclusion

This test case analyzed an oversaturated network in Reston Parkway in Herndon, VA. The volumes analysis indicates that several routing scenarios could be considered in the timing procedure. One routing scenario was selected in this study and two sets of mitigation strategies were developed. A combination of Phase re-service, cycle time adjustments, green time allocation, negative offsets, and metering strategies were tested. The main feature of the first set was metering traffic at Dulls Toll-way eastbound ramps intersection. The second set of strategy focus on increasing downstream capacity of several critical routes by applying phase re-service at the NBL movement at Sunset hill Rd. intersection. The case study concludes several findings such as: short cycle lengths with simultaneous offsets minimized total system delay. Medium-length cycle times were found to be optimal in maximize total throughput. Metering strategies that maximizing network throughput were found to decrease total delay by 35% and increased total throughput by 11%. Strategies that were focused on minimizing total delay could reduce delay by up to 35%, but increased throughput by only 7%. Strategies using phase re-service that were focused on minimizing total delay could reduce delay up to 29%, but decreased throughput by only 1%. A three-dimension temporal Pareto front diagram is developed to illustrate the variation and the duration of optimal strategies throughout peak period. This innovative diagram will assist analysts to build a combination of an effective control strategy that meet agency's control objectives. Results revealed that the proposed framework led to significant

improvement of system performance. Results also revealed that different control parameters within the optimal range lead to different amount of improvement considering different performance measures.

This test case illustrates the importance of considering different objectives when different assumptions or information about critical routes are available. In this test case, the loading, processing, and recovery regimes were not considered. Only one set of strategies were applied during the entire duration of the scenario.

The application of each of the mitigation strategies is dependent on the situation and causation of oversaturation. In many, if not most, cases, oversaturation conditions are a “system-wide” problem and cannot be isolated to only one location. As the spatial and temporal scope of the oversaturated scenario grows, it is unlikely that any single mitigation approach is, by itself, enough to handle a complex and dynamic control situation. It is not possible to specify a general approach to combine mitigation strategies, but many mitigation strategies together can be additive in their curative and palliative effects on any scenario. Based on the sheer number of potential combinations, it was not possible to explore every possible situation in this case study.

9 Case Study: Post Oak Network, Houston, TX

9.1 Introduction

This chapter describes the implementation of OTS in a congested arterial network in Houston, Texas. This case study allows us to investigate two of the previously stated hypotheses:

- **"**During oversaturated conditions, achieving optimality in signal operation requires considering system dynamics in strategy selection, control objectives, and performance measures. At least two of system states (i.e., demand) must be considered in timing plan design: loading and recovery."
- "During oversaturated conditions, multi-stage optimization, volume spillover, and switching between control objectives will improve overall network performance."

This is a real-world test case concerns a heavily traveled arterial that becomes significantly oversaturated on several critical routes. This oversaturation situation happens recurrently during AM and PM peak periods due to surges in traffic simply due to day-to-day variability. In this test case, we have studied the application of various mitigation strategies at the five intersections segment of Richmond Ave. OTS was used to generate control plans that focus on each of the previously defined oversaturation regimes: *loading*, *processing*, and *recovery*. The generated control plans include also the optimal switch points between control objectives (i.e., capacity maximization, delay minimization, and queue management). The following are the sequences of control plan development and evaluation process:

- Define network critical routes/ movements
- Define oversaturation attributes and the corresponding problematic scenarios
- Generate multiple TP for different control objectives
- Select the optimal TP & their optimal switching points
- Evaluate the selected control plans using multi performance criteria

9.2 Study network: Post Oak Network

Figure 61: I-610 loop and US-59 interchange, Houston, TX

Houston, Texas, is the fifth largest metropolitan area in the United Sates. The Post-Oak / Galleria mall area experiences serious oversaturated conditions during peak periods and high traffic volumes throughput the day. The study network is adjacent to the HI-610 loop and US-59 interchange as shown in Figure 61. The I-610 West Loop is Houston's second busiest freeway. It is often referred to as the West Loop parking lot because of the severe congestion that occurs between I-10 and US-59. Westheimer Rd., on the North side of the sub-network shown below, is one of the most heavily traveled arterials in the region. Post-Oak network s is a good example of a combination of arterial and grid operations, and the influence of freeway on- and off-ramp traffic on arterial operations.

Figure 62: Houston Galleria, Uptown District, skyline view

The Uptown/Galleria district of Houston, Texas, is a dynamic, urban community in Houston's West Loop, about five miles west of Downtown. The area is bordered by Interstate I-10 on the north, Interstate HI-610 on the east, US-59 on the south and Chimney Rock on the west. About 42,000 residents live in the Houston Galleria's apartments, modern townhouses and single-family homes. A major feature of Uptown Houston is The Galleria, the largest shopping mall in the state of Texas and the fourth largest in America. The Uptown District measures about 5 million square feet of retail space. Uptown is also host to Houston's largest hotels, which host about 20 million visitors a year. Therefore, the area is considered a major attraction in Houston Metropolitan area. The area commercial activities and its proximity of major highways generate different traffic patterns during A.M. and P.M. peak periods.

9.3 Step 1: Development of Critical Route Scenarios

The first step in the process of developing effective control strategies is to identify the critical routes throughout the network. Once the critical routes, movements, and bottleneck links are identified, the possible problematic traffic scenarios can be established. Then, a wide range of control strategies can be considered to encounter and reduce the detrimental effects of these possible scenarios.

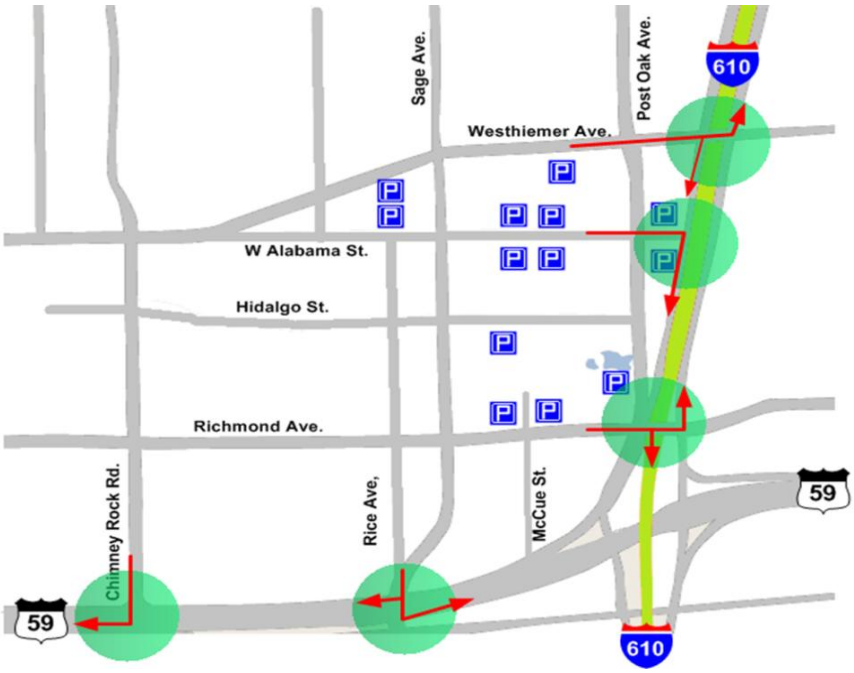


Figure 63: Post Oak parking lot facilities and highways accesses points

Figure 64: Network exits & routes to the highways interchange ramps

During the P.M. peak period, a large number of trips are generated from the network, as employees leave their offices. The large number of parking lots facilities and their substantial capacities are indicators to the significant contribution to the P.M. traffic load. Moreover, another traffic pattern is recognized is the network. That is, the through traffic is passing the network east/west to and from

HI-610 loop, south toward US-59. In the analysis, two traffic patterns are considered to establish effective control strategies. Because of its location relative to major freeway connections to central Houston, there is a pattern of high commuter traffic outbound the network in the PM peak. The major PM peak's routes are illustrated in Figure 64. The volume Figure 64 analysis consists of correlation analysis of data from detector during the peak period. Correlated movements were clustered to construct critical routes. The outcomes of this process (i.e., network critical routes) were verified and complemented with knowledge provided by local practitioners.

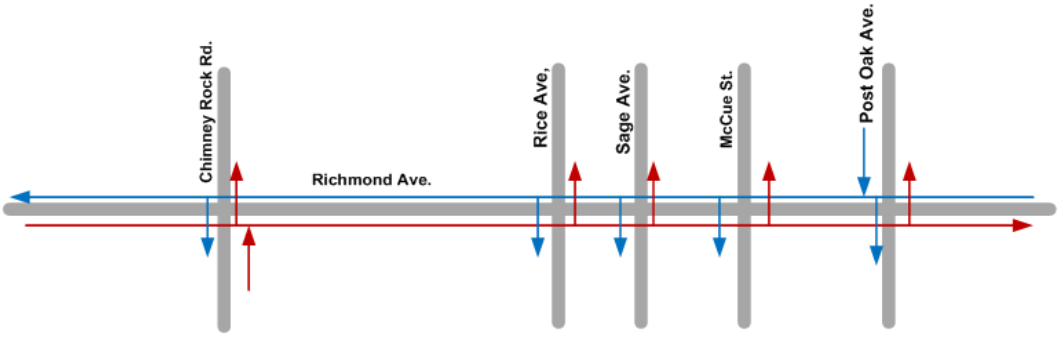


Figure 65: critical routes pass through Richmond Ave.

9.4 Step 2: Control Strategy Development

During the peak hours, Post Oak network faces serious recurrent congestion. The congestion occurs during evening peak periods when the traffic demand increases significantly due to large trips generated from the commercial office parks. The symptoms of the congestion are characterized by slower speed, intersection starvation, storage blocking, and entry drives blocking (i.e., secondary congestion). Initial evaluation using VISSIM simulation software indicated that the dominant factor affecting the traffic operation in the network was recurring spillback, queues formation, and secondary congestion. Left-turn storage bays spillback serve to block and disturb traffic flow on main approaches. The left-turns storage spillbacks are caused by inadequate green times and limited storage capacity. The following table defines network oversaturation attributes.

Table 16: Post Oak network oversaturation attributes

Extent	Duration	Causation	Recurrence	Symptoms
Movement	Situational	Signal Timing	**Recurrent**	**Starvation**
Approach	Intermittent	Geometrics	non-Recurrent	**Spillback**
Intersection	**Persistent**	Other modes		**Storage Blocking**
Route	Prolonged	**Demand**		Cross Blocking
One-way arterial		Planned Events		
Two-way arterial		Unplanned Events		
Interchange				
Grid				
Network				

The process of choosing mitigation strategies began with addressing the most obvious problem(s) first and working outward from there to identify other symptoms and potential mitigations. Since persistent, recurring spillback of left-turn bays was one of the dominating factor influences the operation during the peak period, it was essential to develop a control strategy that explicitly addresses this problem and reduces its impact on through movements. Left-turn phase re-service strategy provides extra capacity to the left-turn movements. To prevent short links spillback, green flaring strategy was implemented. "Flare the green" strategy provided extra capacity to downstream intersection in order to clear the existing queues. Flaring the green is increasing the green windows at downstream intersections. Network primary arterial is characterized by both long-spaced high-capacities and close-spaced intersections as well. These configurations allow the usage of large cycle length. However, the implementation of such large cycles in the interior intersections would cause queues buildup blocking entry drives and storages bays. Therefore, running interior and short-spaced intersections with half cycle would reduce the queues formation and would provide better circulation. The following tables and figures summarize the mitigation strategies used in the developed control plans, their locations, expected improvements, and operational objectives

Table 17: Post Oak applied mitigation strategies

Mitigation Strategy	Description
Oversaturation Offsets	Offsets are determined to accommodate both spillback and starvation. This is considered a queue management strategy.
Simultaneous Offsets	Due to the short link between Sage Ave, and Rice Ave., intersections, simultaneous offsets were implemented for both directions
Phase re-service	Implemented for the WBL and EBL at Sage Ave. and Rice Ave., due to the short left-turn bays' storage
Run intersections free during the process regime	Each intersection in the network is allowed to decide its optimal cycle length independently
Flaring-the-green	In this mitigation strategy, the EBT movements are flared from Chimney Rd. to McCue St.

9.5 Step 3: Control Plans Development

Three control plans were developed based on the above routing scenario using OTS. The control objectives and sequences for the developed plan are illustrated in Table 18: The minimum duration for each control strategy was set to be 30 minutes interval.

Table 18: Oversaturation control plans objectives and schedules

Control Plan	Loading regime	Process regime	Recovery regime
Base*	Delay minimization	Delay minimization	Delay minimization
TP 1	Throughput max	Queue management	Delay minimization
TP 2	Delay minimization	Queue management	Delay minimization
TP 3	Throughput max	Queue management	Throughput max

** Timing plan was generated using the peaks volumes using SYNCHRO*

Table 19: Enumeration of optimal timing plans in the switching problem

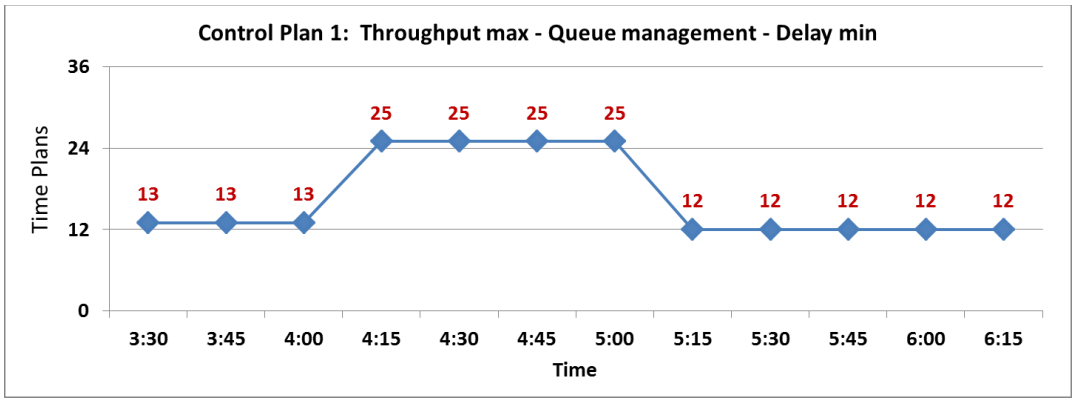
Post Oak Network optimal timing plans			
	Control Objectives		
Optimize for time period	Delay Minimization	Queue Management	Throughput Maximization
3:30 - 3:45 P.M.	1	25	13
3:45 - 4:00 P.M.	2	26	14
4:00 - 4:15 P.M.	3	27	15
4:15 - 4:30 P.M.	4	28	16
4:30- 4:45 P.M.	5	29	17
4:45 - 5:00 P.M.	6	30	18
5:00 - 5:15 P.M.	7	31	19
5:15 - 5:30 P.M.	8	32	20
5:30 - 5:45 P.M.	9	33	21
5:45 - 5:00 P.M.	10	34	22
6:00 - 6:15 P.M.	11	35	23
6:15 - 6:30 P.M.	12	36	24

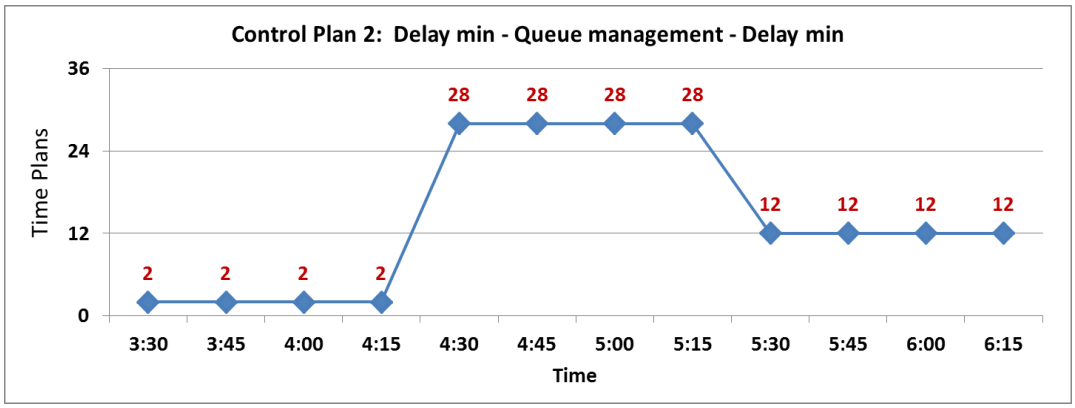
Table 20: Optimal Control plans description

Plan	Loading Phase				Processing Phase				Recovery Phase			
	Operation Objective	Timing Plan #	Cycle Length	Starting time	Operation Objective	Timing Plan #	Cycle Length	Starting time	Operation Objective	Plan #	Cycle Length	Starting time
1	Thr max	13	120	3:30	Queue Mangt.	25	Vary	4:15	Delay min	12	120	5:15
2	Delay min	2	100	3:30	Queue Mangt.	28	Vary	4:30	Delay min	12	100	5:30
3	Thr Max	13	180	3:30	Queue Mangt.	25	Vary	5:00	Thr max	22	100	6:00

Timing Plans (1-12) are optimal delay-min-based plans, (13-24) are throughput max-based plans, and (25-36) are optimal queue management-based plans

*For queue management intersection are not coordinated





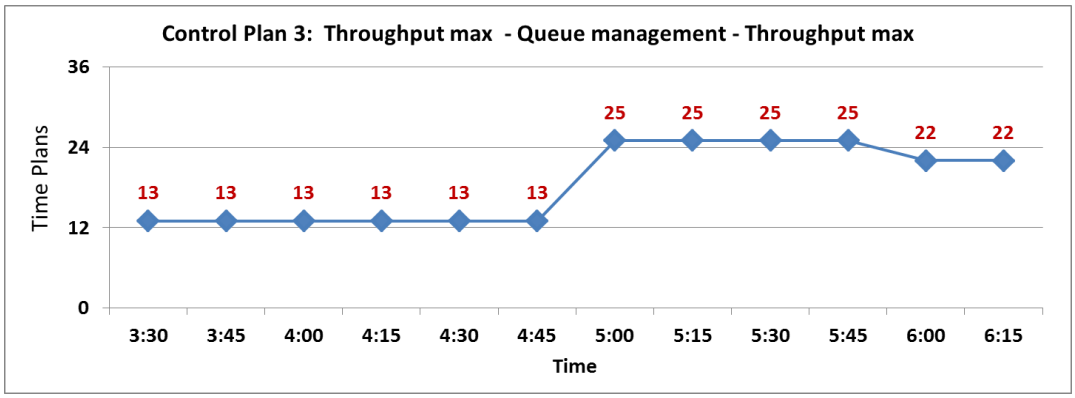


Figure 66: Control Plan 1, 2 and 3's optimal timing plans and their optimal switch points

Table 21: Optimal TP for the developed control plans (1, 2, and 3)

Control plan 1 timing parameters

Throughput max	Green splits (sec)									Offsets
Intersection	S1	S2	S3	S4	S5	S6	S7	S8	Cycle	WBT
Chimney Rd.	20	33	18	49	20	33	18	49	120	36
Rice Ave.	13	14	25	68	13	14	25	68	120	33
Sage Ave.	13	34	28	45	13	34	28	45	120	31
McCue St.	15	24	24	57	15	24	24	57	120	39
Post Oak Ave.	16	29	21	54	16	29	21	54	120	0

Queue management	Green splits (sec)									Offsets
Intersection	S1	S2	S3	S4	S5	S6	S7	S8	Cycle	WBT
Chimney Rd.	23	41	21	98	23	41	21	98	199	29
Rice Ave.	16	17	27	65	16	17	27	65	140	37
Sage Ave.	17	23	22	62	10	23	22	62	140	53
McCue St.	14	29	20	40	9	29	20	42	119	44
Post Oak Ave.	17	29	23	64	17	29	23	64	149	0

Delay min	Green splits (sec)									Offsets
Intersection	S1	S2	S3	S4	S5	S6	S7	S8	Cycle	EBT
Chimney Rd.	19	27	20	54	19	27	20	54	120	65
Rice Ave.	13	24	18	65	13	24	18	65	120	65
Sage Ave.	13	26	22	59	13	26	22	59	120	72
McCue St.	14	34	19	53	14	34	19	53	120	56
Post Oak Ave.	17	41	15	47	17	41	15	47	120	0

Control plan 2 timing parameters

Delay min	Green splits (sec)									Offsets
Intersection	S1	S2	S3	S4	S5	S6	S7	S8	Cycle	WBT
Chimney Rd.	11	22	14	53	11	22	14	53	100	72
Rice Ave	11	20	15	54	11	20	15	54	100	72
Sage Ave.	14	22	19	45	14	22	19	45	100	75
McCue St.	11	26	18	45	11	26	18	45	100	56
Post Oak Ave.	14	23	14	49	14	23	14	49	100	0

Queue management	Green splits (sec)									Offsets
Intersection	S1	S2	S3	S4	S5	S6	S7	S8	Cycle	WBT
Chimney Rd.	9	16	8	42	9	16	8	42	75	68
Rice Ave.	13	14	17	59	13	14	17	59	102	67
Sage Ave.	12	26	24	49	12	26	24	49	111	76
McCue St.	9	29	15	38	9	29	15	38	91	59
Post Oak Ave.	11	29	9	25	11	29	9	25	74	0

Delay min	Green splits (sec)									Offsets
Intersection	S1	S2	S3	S4	S5	S6	S7	S8	Cycle	WBT
Chimney Rd.	11	22	18	49	11	22	18	49	100	67
Rice Ave.	11	22	16	51	11	22	16	51	100	67
Sage Ave.	12	20	21	48	12	20	21	48	100	74
McCue St.	11	21	18	50	11	21	18	50	100	58
Post Oak Ave.	13	24	19	44	13	24	19	44	100	0

Control Plan 3 timing parameters

Throughput max	Green splits (sec)									Offsets
Intersection	S1	S2	S3	S4	S5	S6	S7	S8	Cycle	WBT
Chimney Rd.	24	45	34	77	24	45	34	77	180	58
Rice Ave.	17	38	31	94	17	38	31	94	180	61
Sage Ave.	21	34	34	91	21	34	34	91	180	56
McCue St.	24	33	19	104	24	33	19	104	180	44
Post Oak Ave.	27	56	22	75	27	56	22	75	180	0

Queue management	Green splits (sec)									Offsets
Intersection	S1	S2	S3	S4	S5	S6	S7	S8	Cycle	WBT
Chimney Rd.	32	39	27	71	32	39	27	71	169	19
Rice Ave.	17	23	26	84	17	23	26	84	149	47
Sage Ave.	17	36	17	85	17	36	17	85	155	48
McCue St.	18	24	23	97	18	24	23	97	162	41
Post Oak Ave.	24	64	18	59	24	64	18	59	165	0

Throughput max	Green splits (sec)									Offsets
Intersection	S1	S2	S3	S4	S5	S6	S7	S8	Cycle	EBT
Chimney Rd.	18	32	22	78	18	32	22	78	150	15
Rice Ave.	15	36	27	72	15	36	27	72	150	32
Sage Ave.	18	43	26	64	18	43	26	64	150	54
McCue St.	20	29	24	77	20	29	24	77	150	48
Post Oak Ave.	25	64	15	47	25	64	15	47	150	0

9.6 Simulation Experiment

Post Oak network was coded in VISSIM, a time-step behavioral-based microscopic traffic simulation mode [92]l. Ring barrier controller, RBC, was used to implement the developed control plan (i.e., control strategies, timing plans, and switch points. 30-minutes warm-up period was conducted before data collection process starts as well as 30 minutes to clear the network at the end of the volume loading period in the simulation. Each applied control plan was evaluated via 5 simulation runs with different seed number to account for the randomness of drivers and vehicles characteristics, and to improve the level of confidence in the results by reviewing the average performance across the 5 runs.

9.7 Performance measures evaluation

Analyzing oversaturated network is a complex endeavor due to the conflicting control objectives and performance measures. Each one of the performance measures assessed one aspect of system performance. However, collectively they were able to convey a more complete picture of the state of the system at different points in time. That also allowed a better assessment of the overall merits of each control plan as compared with its intended objective. For instance, the network traffic load can reveals very valuable information. Nevertheless, this is only an incomplete picture if the overall effectiveness of the system is not defined. It does not show the consequence of high traffic load on arterial speed and level of progression. Therefore, other performance measures should be examined

such as holdback traffic and throughput rates before a decision is made on which control plan is more effective. The following performance parameters were used to evaluate the control plans:

- Network overall performance measures
  - Network total delay
  - Average delay per vehicle
  - Network total stopped delay
  - Average stopped delay per vehicle
  - Average number of stops per vehicles
  - Average travel speed
- Queue length analysis
  - Links' P-ratio
  - Reserve capacity percentage
- Travel time analysis
  - Critical and non-critical routes travel time profiles
  - Routes travel time reliability
- Throughput analysis
  - Network traffic loads
  - Intersection throughput rates
  - Holding back traffic

9.7.1 *Network overall performance measures*

The overall network performance for the base timing and developed control plans are illustrated in the following table. The results of the overall performance measures average indicated that each of developed control plan reduces the overall delay, average delays, and improve average speed.

Table 22: Overall network performance for the base timing and the developed control plan

Parameters	Base	TP1	TP2	TP3
Total delay (hour)	2022	1828	1926	1850
Average delay (sec/veh)	231	212	229	213
Average speed (mph)	5.7	6.5	6.1	6.5
Average stopped delay (sec/veh)	150	143	158	148

9.7.2 *Queues analysis*

In oversaturated conditions, the subjects of signal control and links queues management are closely related more so than in under sat-saturated conditions. The following section presents the results for network links queues lengths. All presented results are stated on a per lane basis. The main arterial links P-ratios are presented as they relate to different control plan and temporal system characteristics as shown in the figures below. The figures demonstrate that the base timing resulted in P-ratio greater than one in all arterials links with exception of the link between Chimney Rock Rd, and Rice Ave. (with length of 1700 ft.). The OTS control plans kept the arterial links' P-ratio under one during the analysis period most of the time.

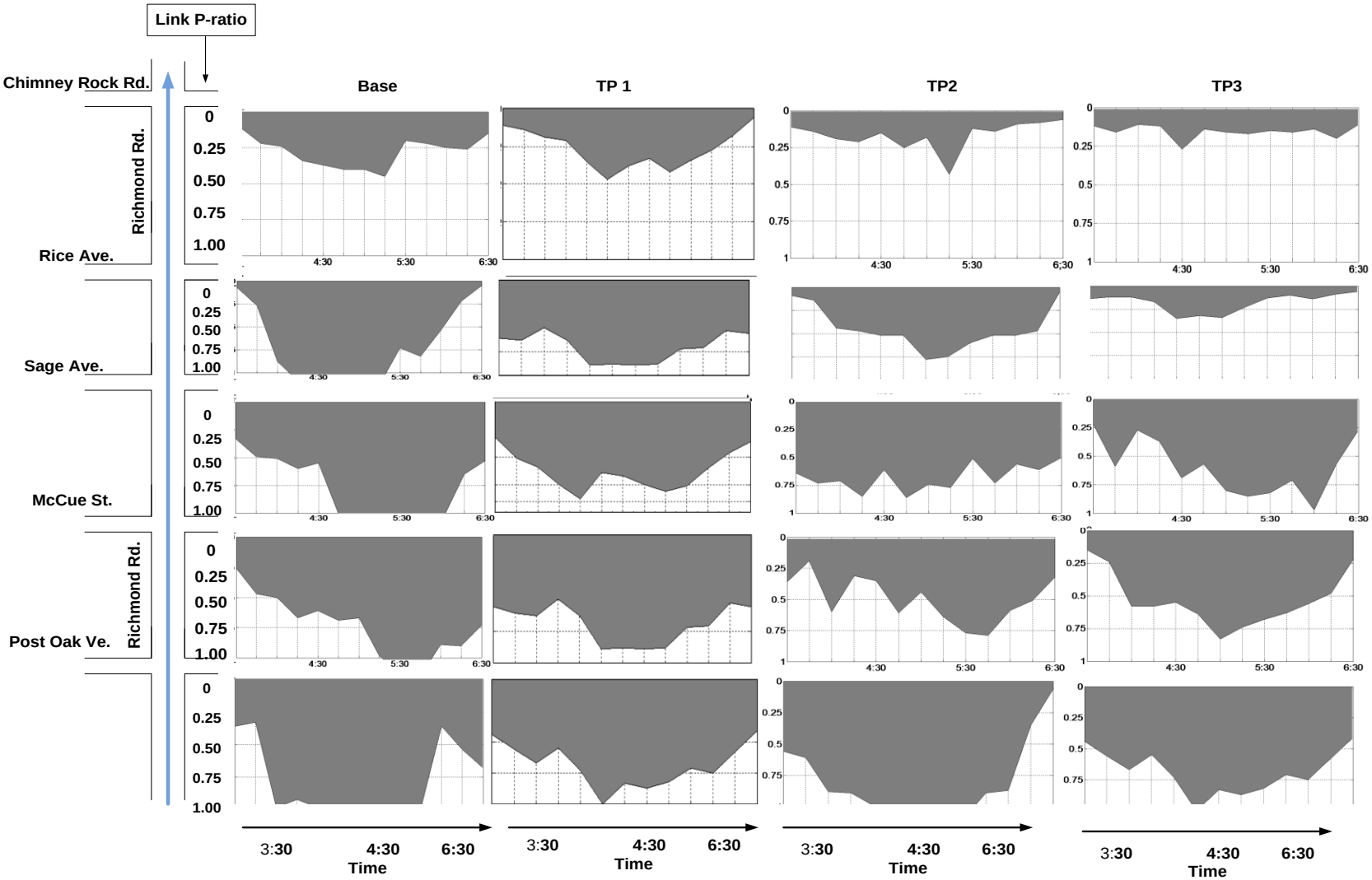


Figure 67: Richmond Ave. WBT P-ratios during the analysis period (3:30-6:30 P.M.) for the base and tested control plans

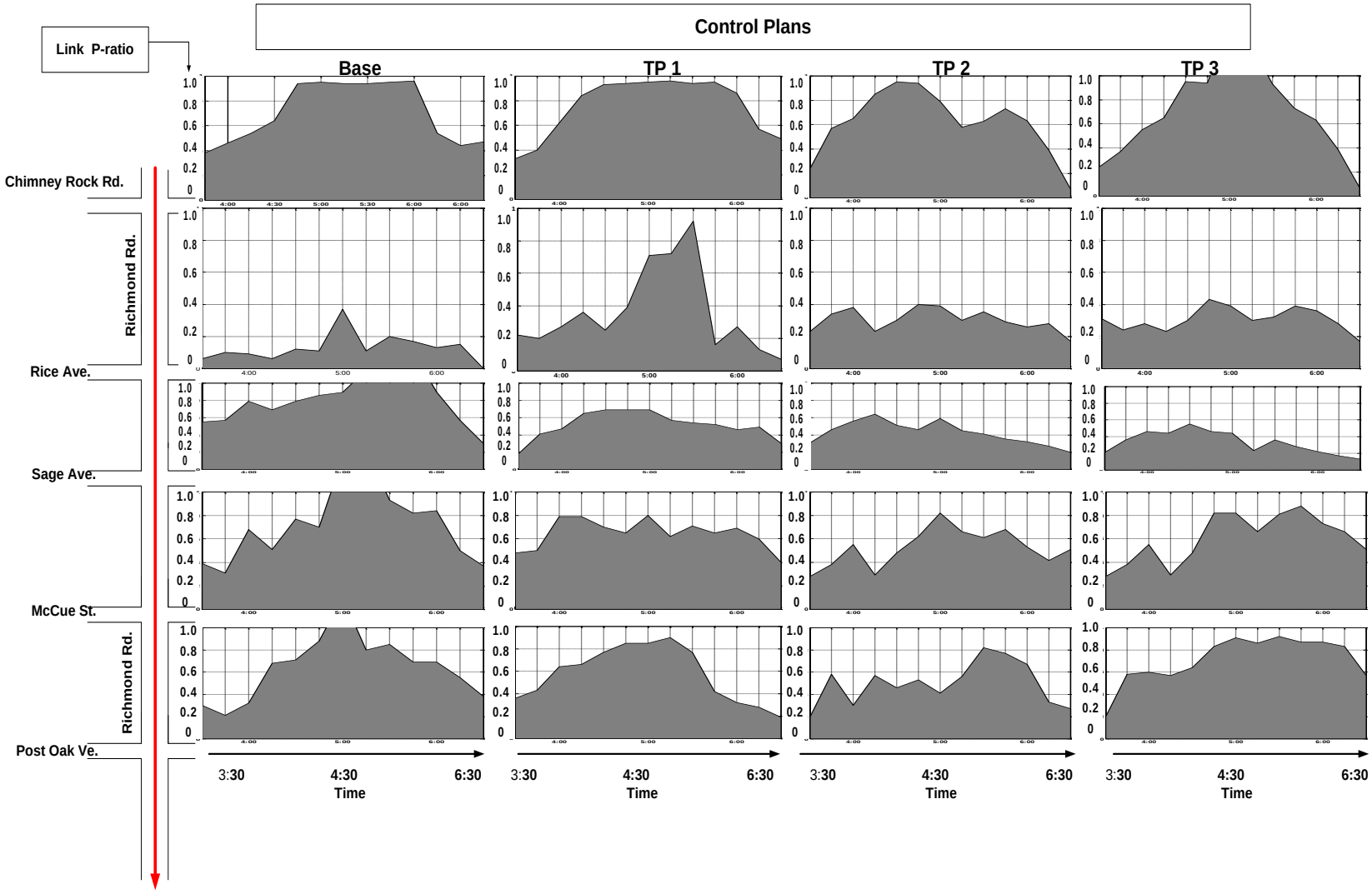


Figure 68: P-ratios on EBT P-ratios during the analysis period (3:30-6:30 P.M.) for the base and tested control plans

9.7.3 *Network reserve capacity percentage*

It measures the available space of queues on the receiving links in the network. Low percentage values indicate higher chances of queue backup and possible spillback.

$$Reserve\ queuing\ capacity = 1 - P\ ratio$$

Figure 69, Figure 70, Figure 71, and Figure 72 show the queuing capacities percentages available on the receiving network links as time progress. This measure is very significant as it indicates the inadequacy in queues allocation for the applied control plan. The reserve queuing capacity for side streets is an indicator of the network *external metering* performance, while the for main arterial movements' reserve queuing capacity indicate the performance of network *internal metering*. Figure 69 shows that the network under base timing control with positive reserve queuing capacity for main arterial links, however, the previous figures indicates links spillback occurred for the main arterial movements. This is probably due to the fact that the existing capacity of the link between the intersections of Chimney Rock Rd, & Rice Ave. (i.e., length 1700 ft.). This sizeable queuing capacity in the network can accommodate several failing cycles without exceeding queuing capacity.

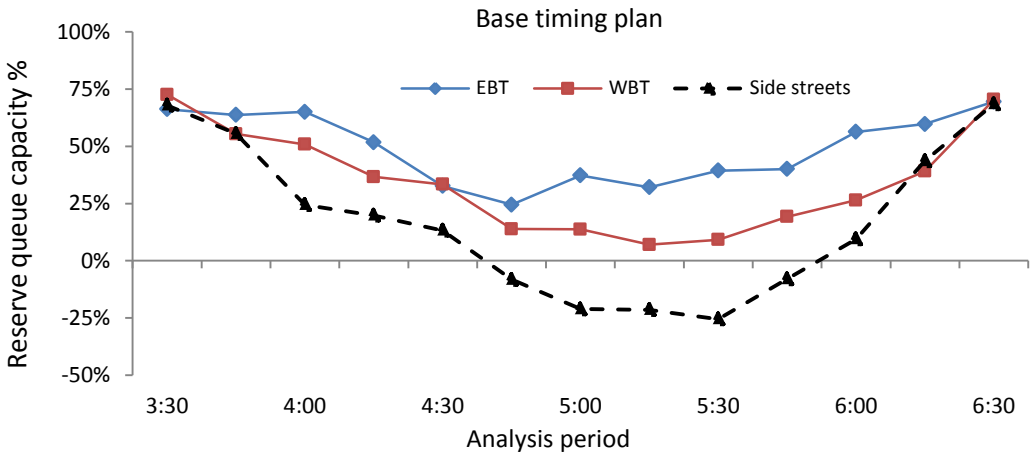


Figure 69: Network reserve queuing capacities percentages

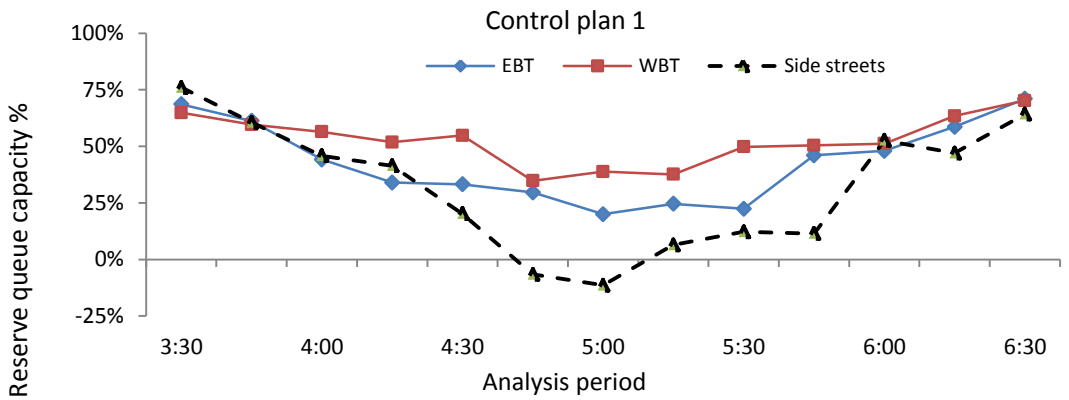


Figure 70: Network reserve queuing capacities percentages

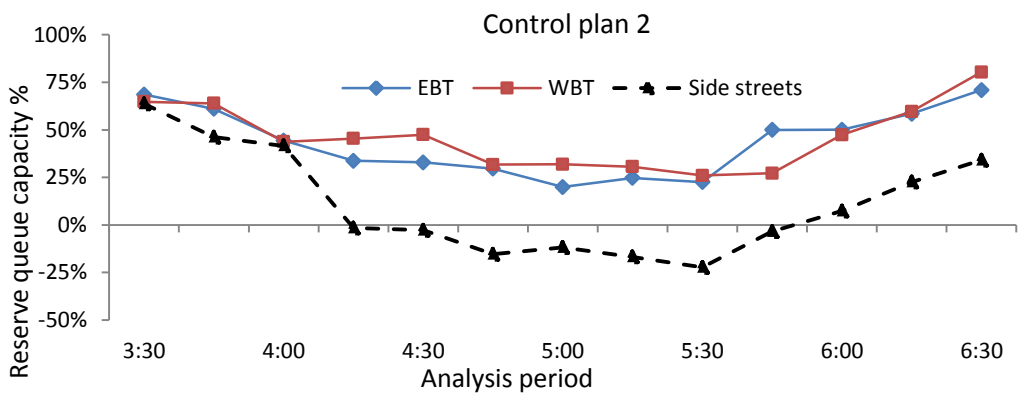


Figure 71: Network reserve queuing capacities percentages

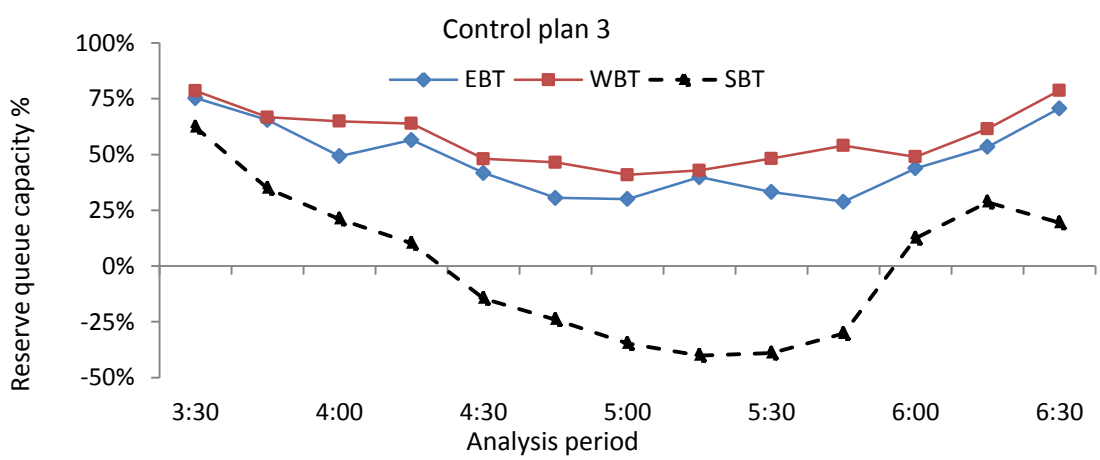


Figure 72: Network reserve queuing capacities percent

Figure 73 shows the simulation result's P-ratios for both critical and non-critical links. The diagrams indicate the developed control plans (TP1, TP3, and TP3) managed to keep the P-ratios of critical links under 1 as intended in the formulations, while non-critical links P-ratios exceed the value of spillback (i.e., P > 1). Figure 7 indicates the effectiveness of the developed control plans in allocated queues in the network's links.

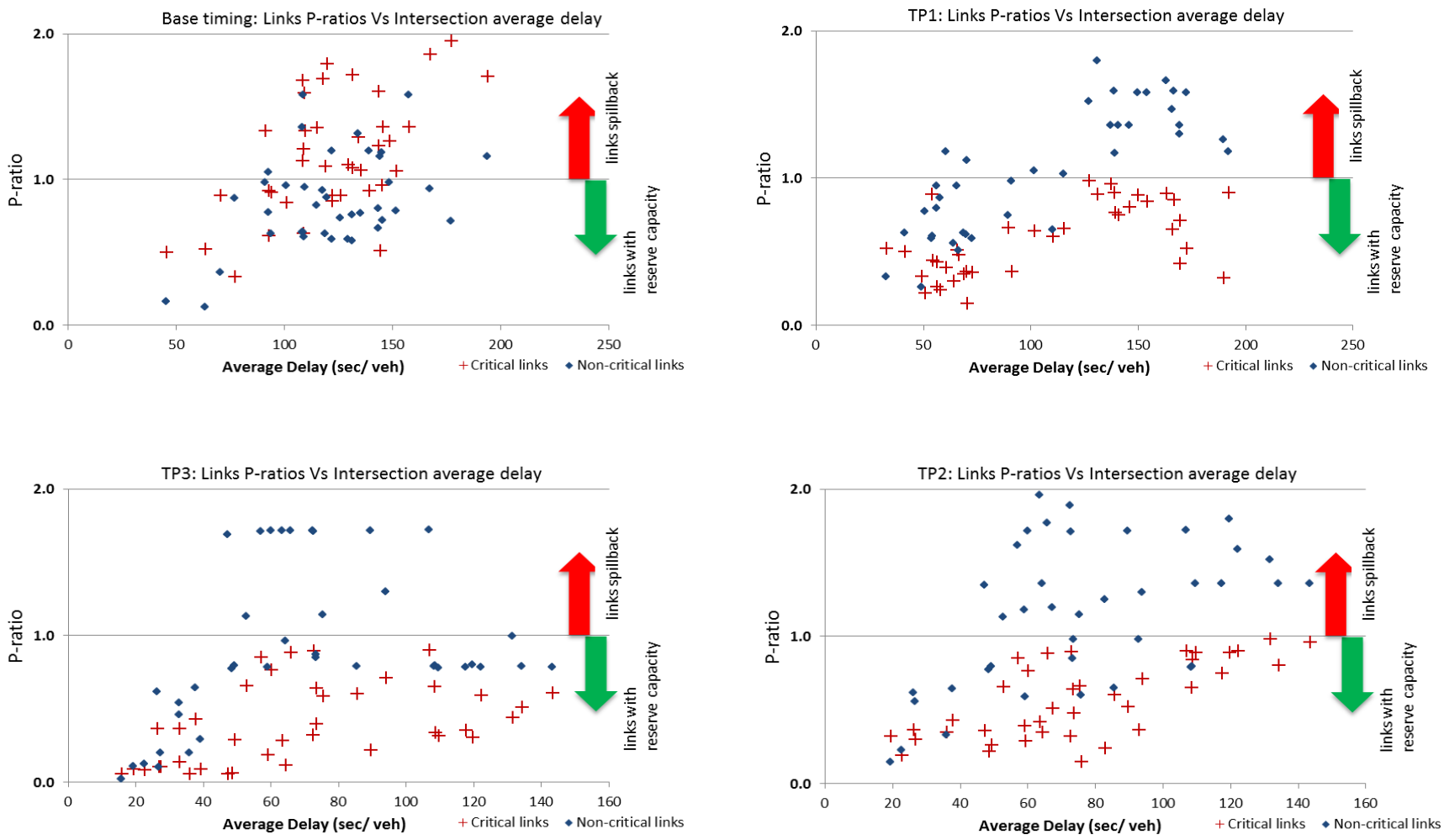


Figure 73: The P-ratios of critical/non-critical links of the base timing and tested control plans TP1, TP2, and TP3

9.7.4 *Throughput analysis*

Definition:

Network output rate: Measured by vehicles per hour, this performance measure quantifies the utilization of network's supply resources, green times and links' storage capacity. Higher output rate indicates that more vehicles discharge per unit time through the network's intersections. It also means that more traffic is processed through the network, which implies the waste of greens is minimized. Thus, it can be a measure of applied control plan efficiency.

Traffic held back: Measured by number of vehicles, this measure is the number of vehicles that want to enter into the system (arterial and cross streets) but was metered out the system. It is estimated from traffic demand arrival rates at the entry point of the system and the volumes allowed into the system. This performance measure quantifies the delay and disruption caused outside the system by the applied control plan.

Network traffic load: Network traffic load is the number of vehicles existing in the network during a certain time interval. For a vehicle that is about to enter the system, this is the number of vehicles that the system will have to process ''out'' before the entering vehicle makes it out of the system. It is a measure of the potential ease with which a vehicle can travel the system. The following figure illustrates the concept of network traffic load over time.

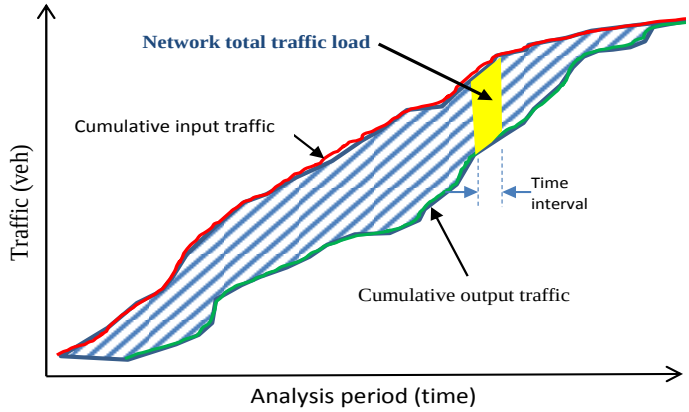
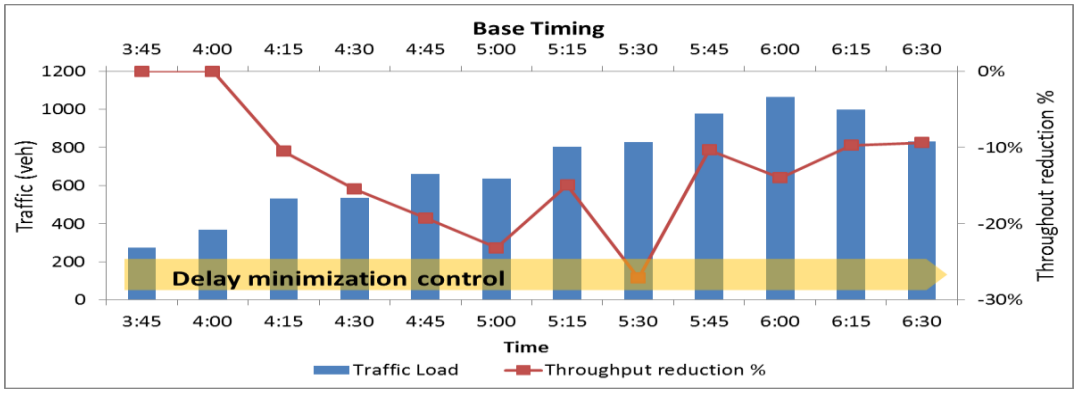
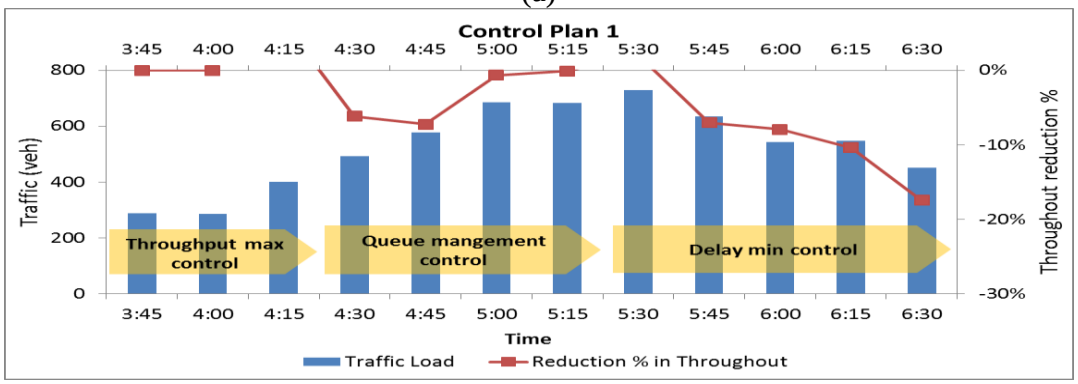


Figure 74: Traffic loading

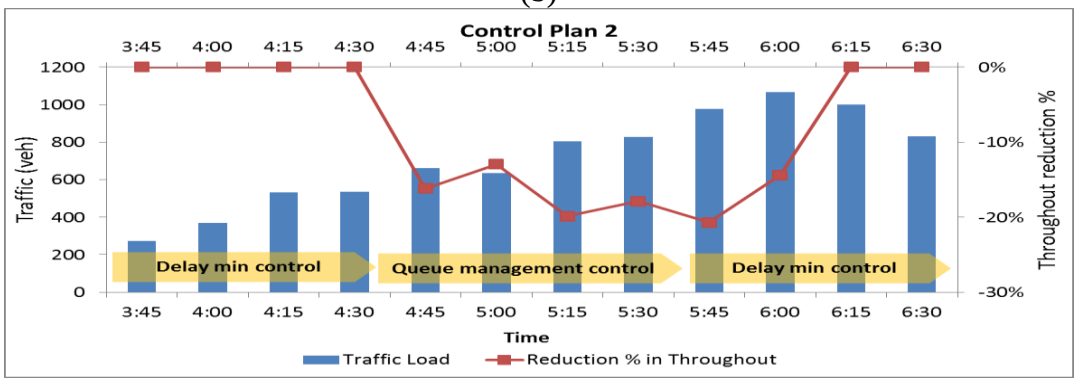
Figure 75 shows the percentage of network's throughput reduction profile a long with total network loading for the applied control plan. The figures also depict the control plan objectives and their switching points. The figure indicates that different control objective will result in different loading patterns in the network. The base timing plan, which has been developed based on peak volumes, shows an early deterioration in throughput rate. The reduction in throughput reached nearly 30% during the processing regimes. Under the base timing, the network continues discharging in lower rate (i.e., 10% reduction in throughout) despite the drop in arrival demand during the recovery regimes.



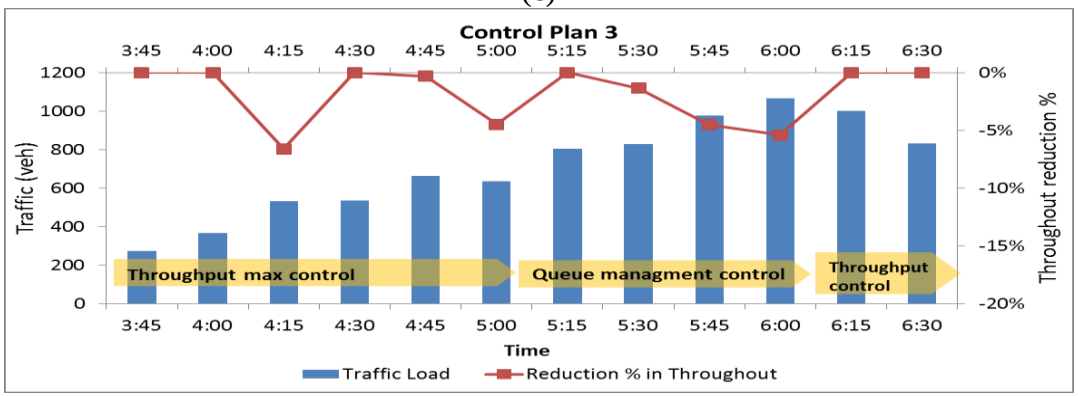
(a)



(b)



(c)



(d)

Figure 75: Network traffic load, throughput reduction %, and control objective for the tested control plans.

Held back traffic: Measured by number of vehicles, it measures the number of vehicles that want to enter into the system (arterial and cross streets) but was metered before entering the system. It is estimated from traffic demand arrival rates at the entry point of the system and the volumes allowed into the system. This performance measure quantifies the delay and disruption caused outside the system by the applied control plan. Figure 76 shows the held back traffic profile for applied control plan Vs base timing plan during the analyzed peak period. Figure 76, Figure 77, and Figure 78 show the percentage of held back traffic before and after the peak point. Base timing plan hold 58% of the metered traffic after the peak, while other control plans start holding back traffic before the peak.

2. Control Plan 1

CP1 applied capacity maximization during the loading phase. This strategy kept the traffic loading at low level at the early stage of network saturation. It can be inferred that CP1 would cause the least impact outside the system, and on side streets approaches. The traffic load decreased significantly during the recovery phase as Figure 75 (b) shows. Intuitively, it is rather obvious that the rapid drop in network throughput rate during recovery phase is due to side streets discharge. Since during recovery regime, delay minimization control allocates extra greens to movements with excessive queues formations (side streets).

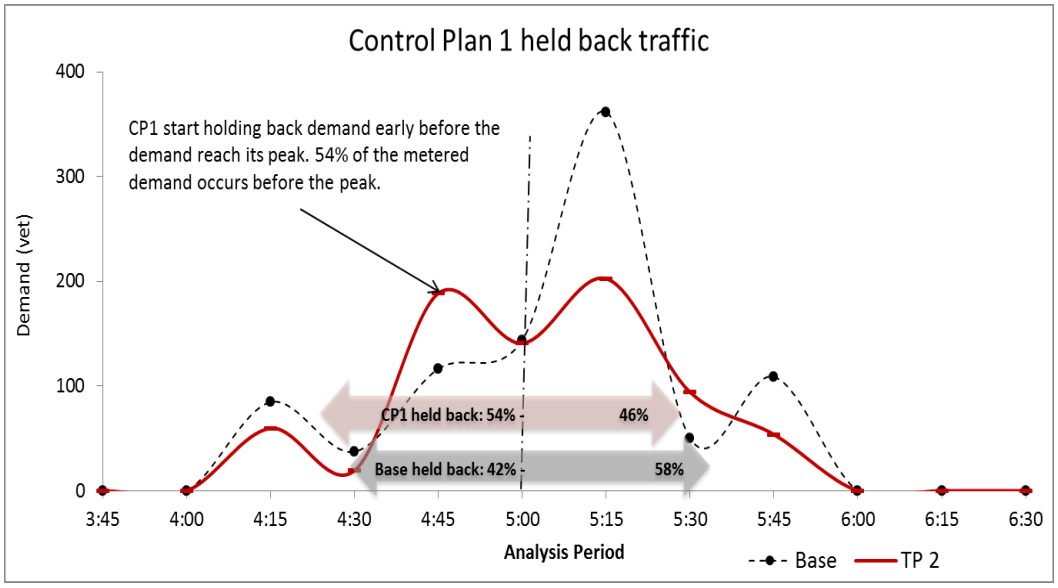


Figure 76: Control plan1 held back demand profiles

CP1 held back most of the traffic during the process phase as more than 574 vehicles got delayed at network entry points. This amount is compared with only 24 vehicles endure metering during the loading phase and 148 vehicles during the recovery phase. As shown in Figure 76, CP1

started metering arrival demand early than the base timing plan. CP1 hold back more than 54% of the metered vehicles before the arrival demand reach its peak comparing with only 42% for base timing plans.

3. Control Plan 2

CP2 has the most traffic load at any time since it applied delay minimization control objective during both loading and recovery demand phases. No capacity maximization control was applied to assist clearing the network (i.e., no priority to arterial forward progression). Traffic load continued to grow with time even during the recovery phase. Most importantly, CP2 substantially reduced the network throughput rate up to by 20% as Figure 75 illustrates. The implementation of delay minimization control in early stage of saturation period comes in expense of network throughput during the process and recovery phases. The conditions during this control plan can be characterized by sluggish movements through the network most of the times. This means that traffic in the network was experiencing frequent stops. Therefore, vehicles progression over multiple links in one cycle is hardly achieved.

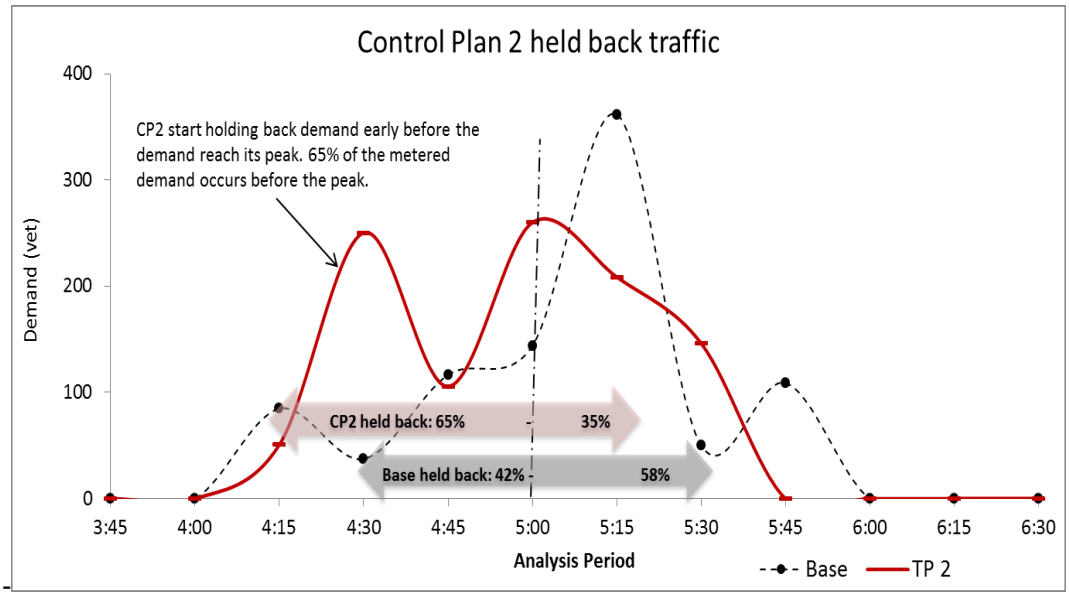


Figure 77: Control plan 2 held back demand profiles

CP2 held back most of the traffic during the processing regime as more than 720 vehicles got delayed at network entry points. This amount is compared with only 301 vehicles endure metering during the loading phase and no vehicles were metered during the recovery phase. As shown in Figure 77, CP2 started metering arrival demand early than the base timing plan. CP1 hold back more than 65% of the metered vehicles before the arrival demand reach its peak comparing with only 42% for base timing plans. Based on the high number of vehicles held back, it seems that CP2 would cause the most disruption to the neighboring facilities.

4. Control Plan 3

Oversaturated control plan 3 (TP3) shows significantly lower network content as time progresses. In this control strategy (i.e., capacity maximization-Queue management-capacity maximization) the effect of the high priority given to arterial traffic throughput is clearly shown in the relatively lower level of traffic load during the process phase (5:15 – 6:15 P.M,). It is clear that the control plan manage to keep the same level of traffic load until the end of the control period. This means that the quality of arterial traffic progression was maximize during the analysis period. Figure 75 shows the traffic load during the analysis period along with the throughput reduction percentage. It is quite interesting to realize the significant low reduction in network throughput rate. This was brought about by applying throughput maximization during the loading phase and recovery phase.

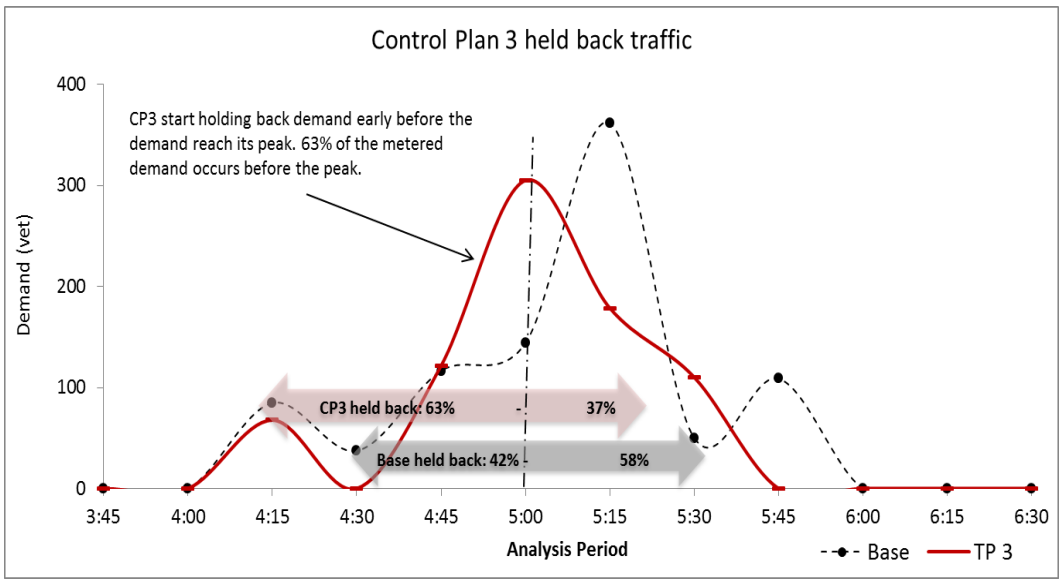


Figure 78: Control plan 2 held back demand profiles

CP3 held back most of the traffic during the loading phase as 494 vehicles got delayed at network entry points. This amount is compared with 289 vehicles endure metering during the processing phase and no vehicle was held back during the recovery phase. As shown in Figure 78, CP3 started metering arrival demand early than the base timing plan. CP3 hold back more than 63% of the metered vehicles before the arrival demand reach its peak comparing with only 42% for base timing plans. This control plan had the highest proportion of the network held-back-traffic during early stage of the operation period, CP3, where very high emphasis was given to clearing queues from the main arterial and providing for through forward progression; the network output rate was the highest among the other control plans. In other words, the high degree of progression achieved by this strategy had to be bought at the expense of ‘‘not very efficient’’ use of arterial green time, and, by implication, more disruption outside the system and to the side streets movement.

9.7.5 *Comparison of network loads and Throughput flows*

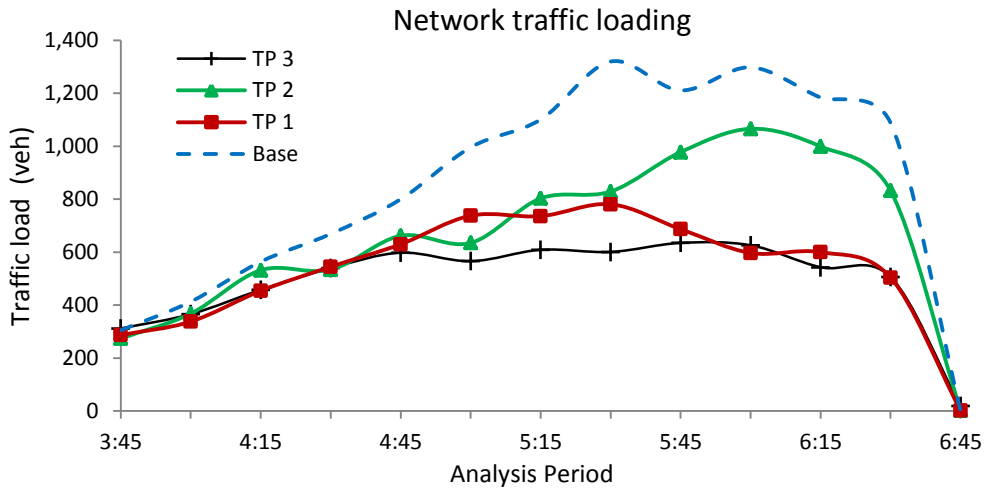


Figure 79: Comparison of network traffic loading

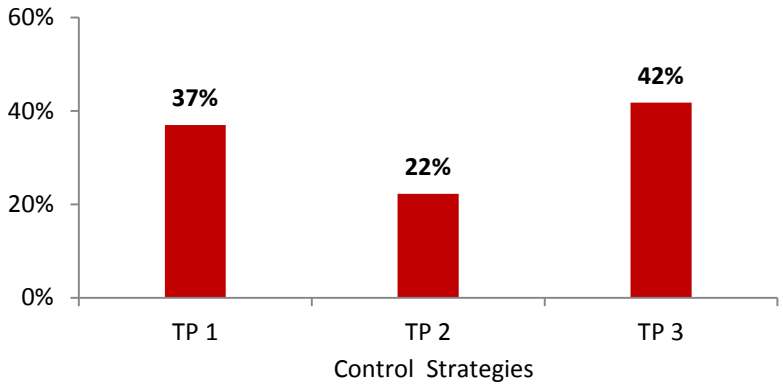


Figure 80: Comparison of network total loading reduction %

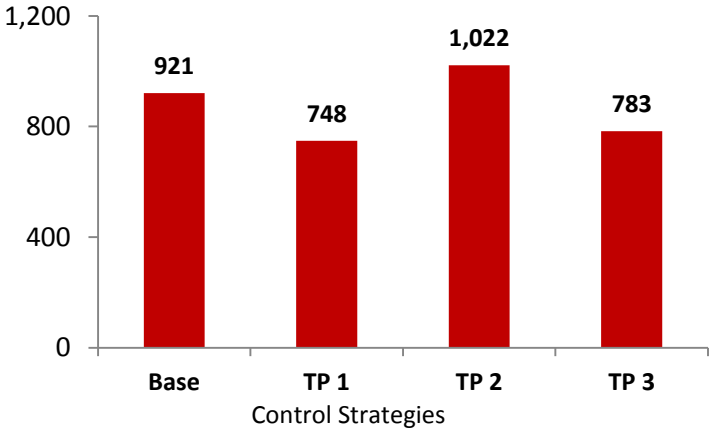


Figure 81: Total network held back demand

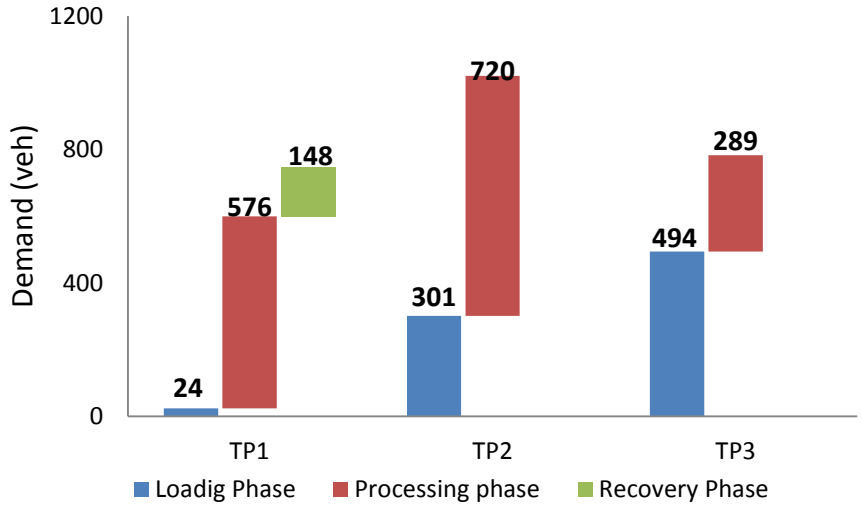


Figure 82: Breakdown of held back demand

9.7.6 *Travel time analysis of results*

OTS control plans for Post Oak network under congested conditions was also analyzed in terms of travel time through the network prime routes. The critical routes travel times results are presented in the following section. The developed control plans improve the performance for the network critical routes and negatively impact the non-critical routes.

Figure 83: route (A), WBT on Richmond Ave.

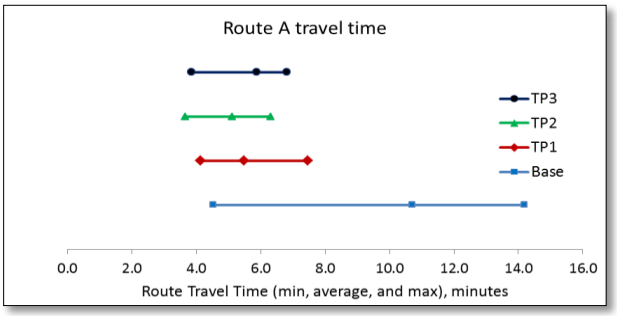


Figure 84: Network critical routes travel times (max, average, and min)

Figure 85: critical route (B), EBT on Richmond Ave.,

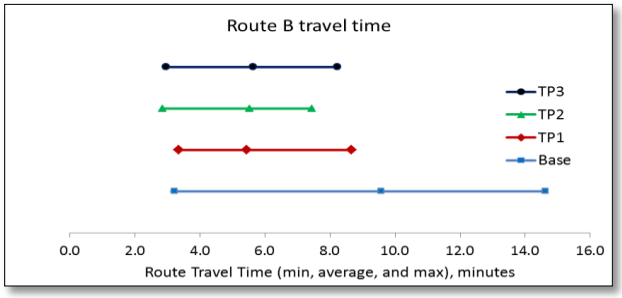


Figure 86: Route B travel times (max, average, and min)

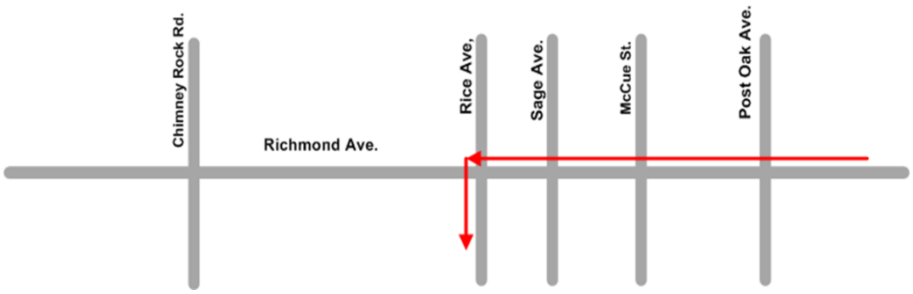


Figure 87: Route (H), WBT on Richmond Ave. then SLT on Rice Ave.

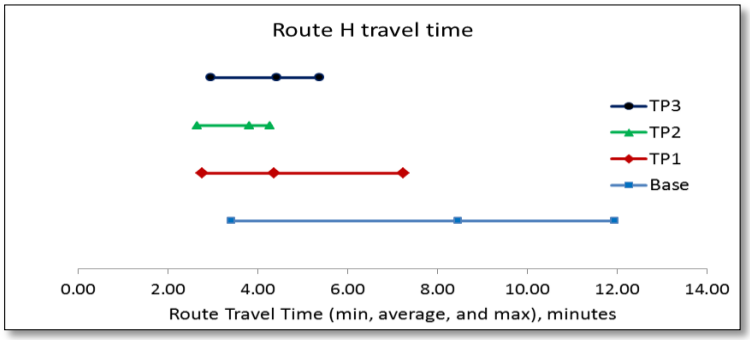


Figure 88: Route H travel times (max, average, and min)

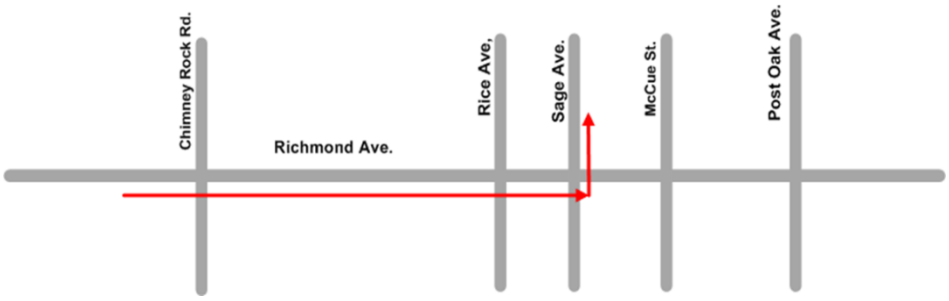


Figure 89: Route (G), EBT on Richmond then NLT n Sage Ave.

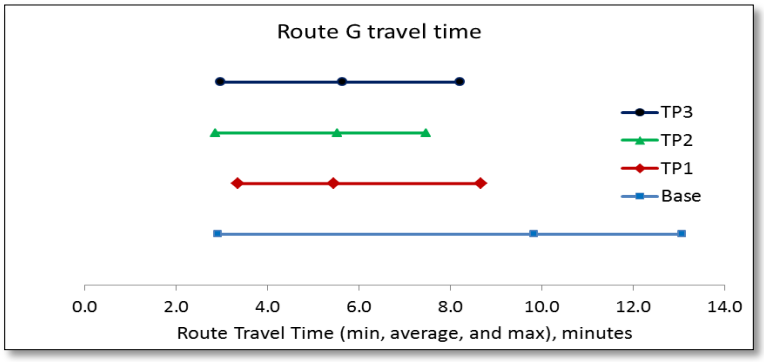
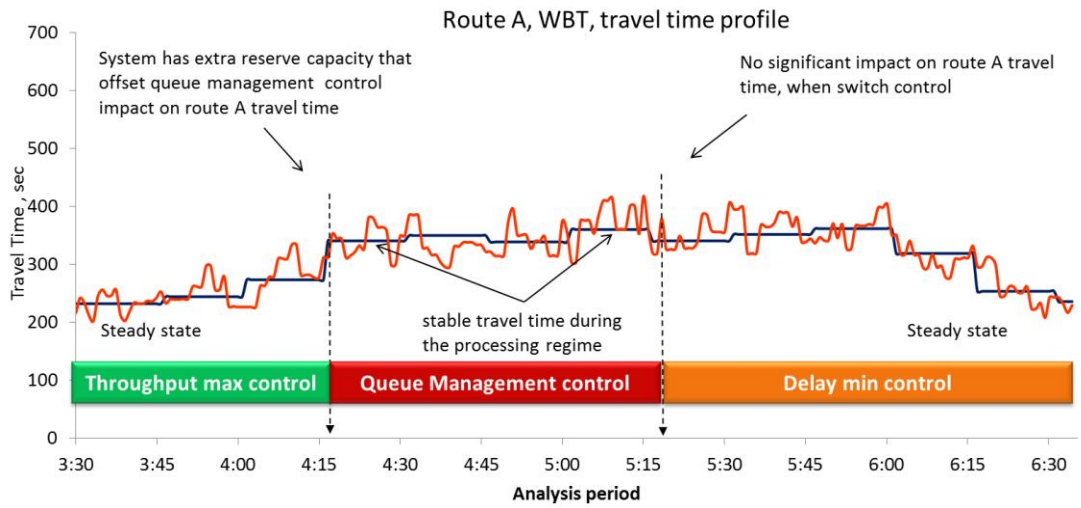


Figure 90: Route G travel times (max, average, and min)

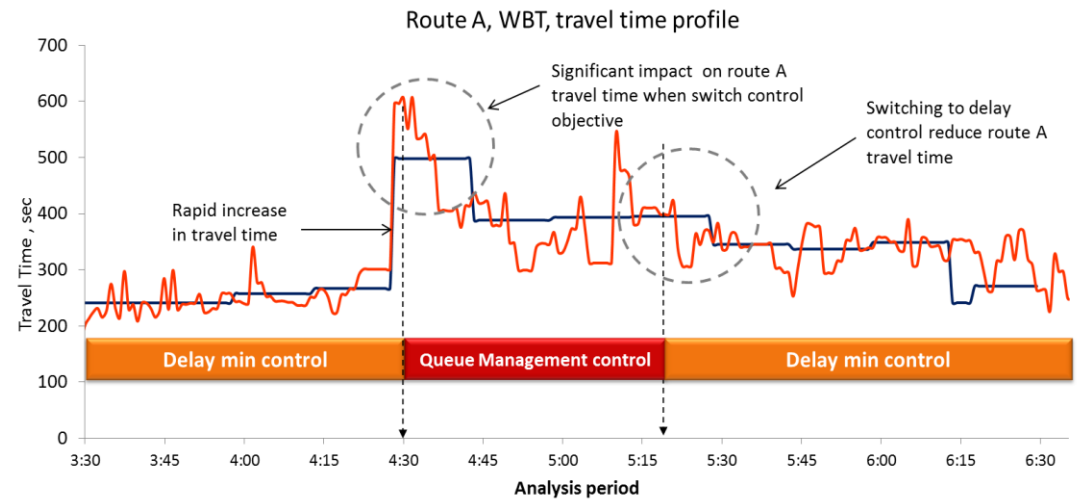
[Figure 84-Figure 90] demonstrates how the travel times in the network critical routes changed significantly when using the optimized oversaturated timing plans. It is clear from the above figures that the OTS timing plan not only produces a more stable travel time, but also had a lower average value of travel time than the base timing plan. When comparing average travel time, OTS timing plans (TP1, TP2, and TP3) reduce route (A) travel time by 50%, 48, and 47%, respectively. The OTS offsets design (i.e., oversaturation offset) was set to favorer progression on routes (A) and (H). The left turns phase re-services operation on Rice Ave. and Sage Ave. have contribute in lowering travel times of routes (G) and (H).

Travel time profiles analysis

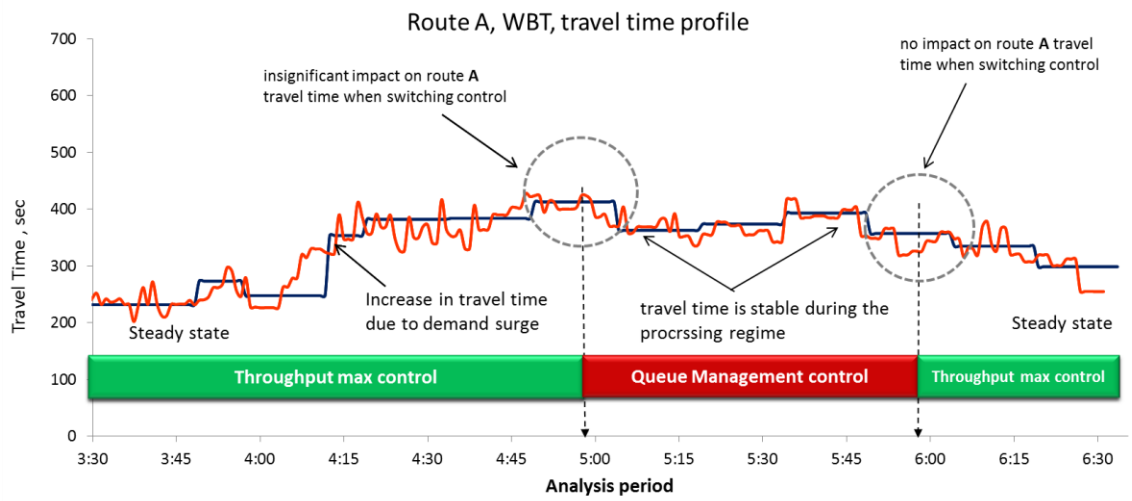
The travel time profiles depicting in the following figures were generated by aggregating vehicles travel time every 1 minute and 15 minute.



(a)



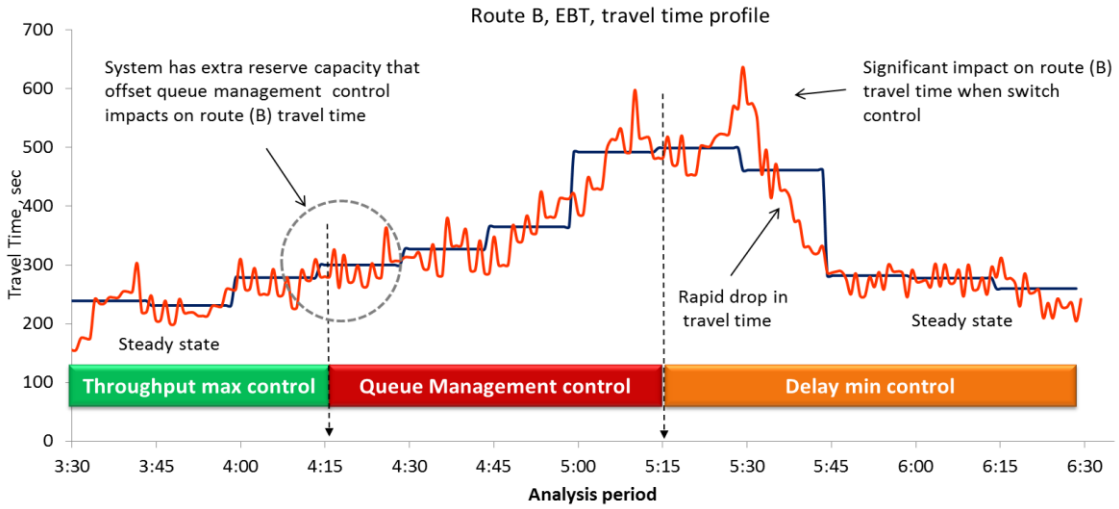
(b)



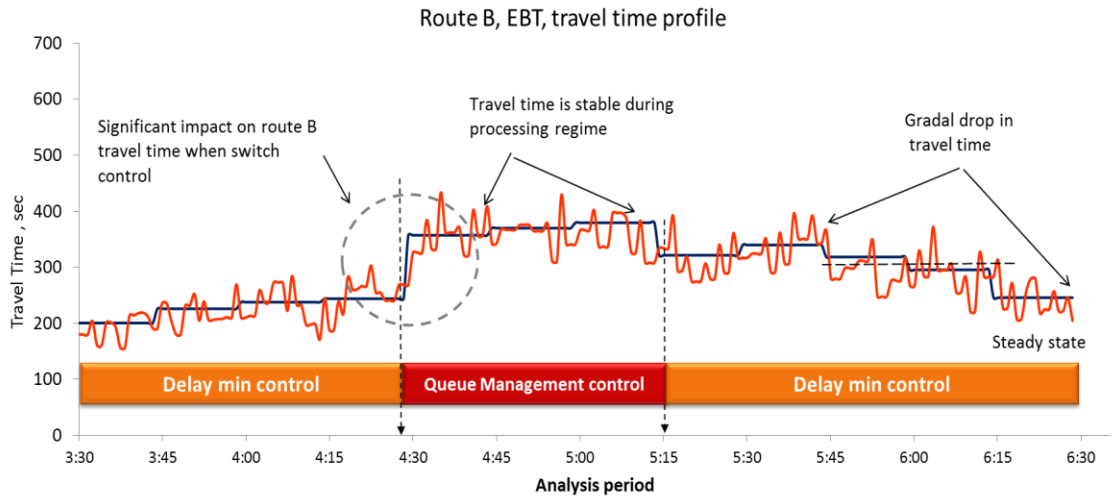
(c)

Figure 91: Route (A) travel time profile for (a) Control Plan 1, (b) Control Plan 2, and (c) Control Plan 3

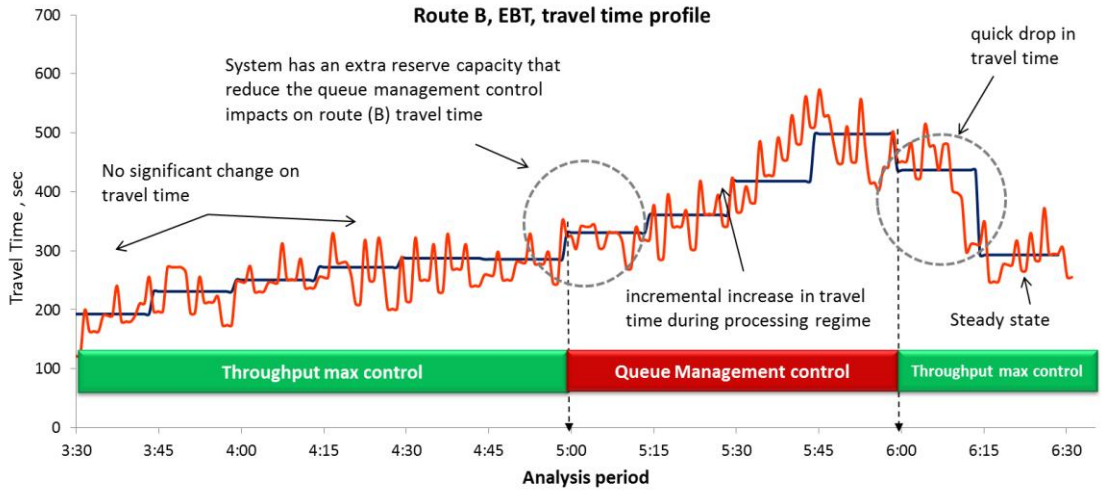
Figure 91 (a) and (b) show that no significant impact occurs instantly on route (B) travel time when control objective switch from throughout maximization to queue management control. One explanation is the fact that applying throughput maximization during loading regime provides significant reserve capacities on network's critical route that postpones oversaturation symptoms on critical routes. However, route (b) travel time increase gradually during the process regime as network gets saturated. For both control plans 1 and 3, route (B) travel time reached as high as 10 min. Figure 91 (b) shows the impact of applying delay minimization control during the loading regime on route (B) travel time. The travel time increase rapidly once queue management control commence. However, travel time keeps stable most of the time during the processing regime. Figure 91 also illustrates that applying delay minimization control during the recovery regime enable the system return to *steady-state* operation *faster* than applying throughput maximization control.



(a)



(b)



(c)

Figure 92: Route (B) travel time profile for (a) Control Plan 1, (b) Control Plan 2, and (c) Control Plan 3

9.7.6.1 *Travel time profile discussion:*

Impact of switch control objective on critical route travel time

The control plan 1 starts by maximizing the network capacity then shift to queue management during the oversaturated phase, then shift to delay maximization during recovery phase. While control plan 2 starts by delay minimization then shift to queue management during the oversaturated phase, then shift to delay maximization during recovery phase. From Figure 92, it is clear that no significant impact on route (A) travel time when switching to queue management control from throughput maximization control unlike switching from delay minimization control as shown in Figure 92 as travel time increased rapidly. These realizations can be explained by that fact that capacity maximization control supplies the network with extra reserve capacities on the critical routes. On the other hand, delay minimization control balances the queues formations on network links. Therefore, the extra reserve capacity help to postpone the adverse impacts of queue management control operations. It is quite interesting that travel time of route (A) continue Fluctuating considerably during queue management control phase after the rapid escalation in the travel time as shown in.

Impact of early switch of control objective on critical routes travel time

Figure 93 illustrates the impact of control switch from throughput maximization control to queue management control on route (A) travel time. Controls plan 1 switch control objective from throughput to queue management at 4:15 P.M. while control plan 3 continues throughput maximization until 5:00 P.M. Route (A) travel time increase. Figure 93 revels that route (A) travel time increase by 5.2%, 5.8%, and 6.2% at 4:15, 4:30, and 4:45, respectively, when switch to queue management control. Route (A) endure insignificant impact when control switch at 4:15 P.M.

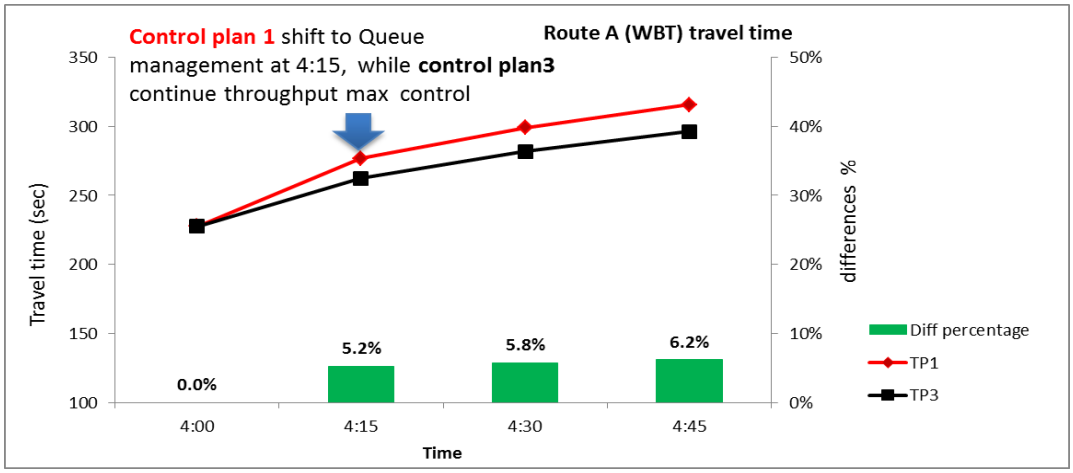


Figure 93: The impact of control objective switching time on route (A) travel time

Figure 94 illustrates the impact of early switching from throughput maximization control to queue management control on route (B) travel time. Figure 94 revels that route (A) travel time increase by

8.4%, 30.7%, and 32.6 % at 4:15, 4:30, and 4:45, respectively, when switch to queue management control. These results indicate that Route (A) endure significant impact when control switch to queue management at 4:15 P.M. in fact, route (B) travel tine decrease as throughput maximization control continues. This outcomes support the literature and practitioners' recommendations of applying throughput maximization control strategies through the loading regime as long as possible. In other hand, delay minimization control strategies were found to be more effective than strategies designed for throughput maximization during the recovery period.

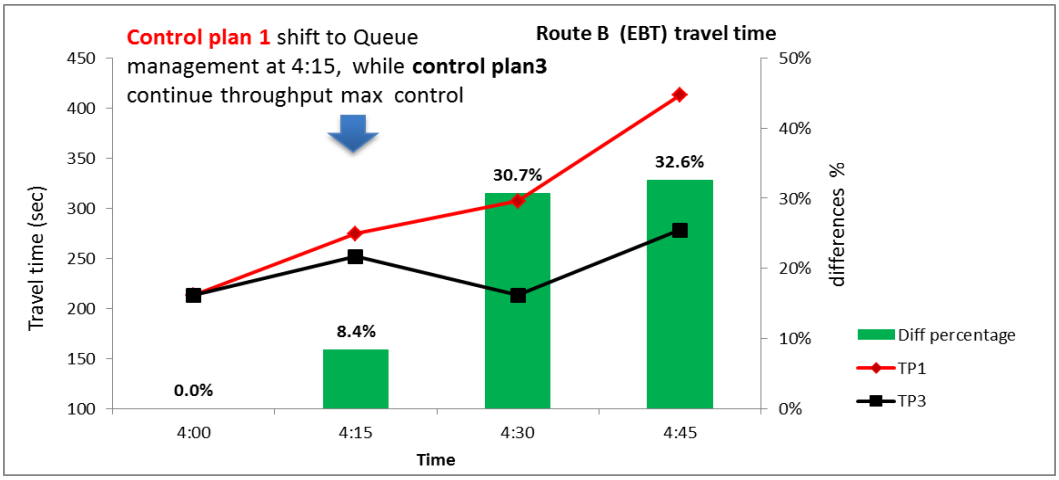


Figure 94: The impact of control objective switching time on route (B) travel time

Reliability analysis:

The ability to measure the variability of travel time is important for assessing the control plan performance. Reliability is often defined as a probability or a percentage of acceptable performance, while in transportation it is typically reported as on-time performance. In the Future-SHRP Reliability Research Program, it is indicated that travel time reliability can be defined in terms of how travel times vary over [93]. This concept of variability can be extended to any other travel time-based metrics such as average speeds and delay. Network travel time reliability mentioned are the Buffer Time Index, 95th percentile of travel times, percent "on-time performance", and travel time window [94]. The following is a demonstration of the effect of oversaturation control plan on the network routes travel time distributions. The routes travel times distributions were compared for base case timing and the developed control plans. The differences in distributions of the travel times are noticeable. As Figure 95 illustrates, the critical routes travel times distribution is less spread-out than the base timing. The figures also demonstrate that OTS control plans not only produces a more stable travel time, but also had a lower average value of travel time than the base timing plan. When comparing average travel time, OTS timing plans (TP1, TP2, and TP3) reduce route (A) travel time by 50%, 48, and 47%, respectively. The OTS offsets design (i.e., oversaturation offset) was set to

favorer progression on routes (A) and (H). The left turns phase re-services operation on Rice Ave. and Sage Ave. have contribute in lowering travel times of routes (G) and (H).

Network critical routes travel time distribution

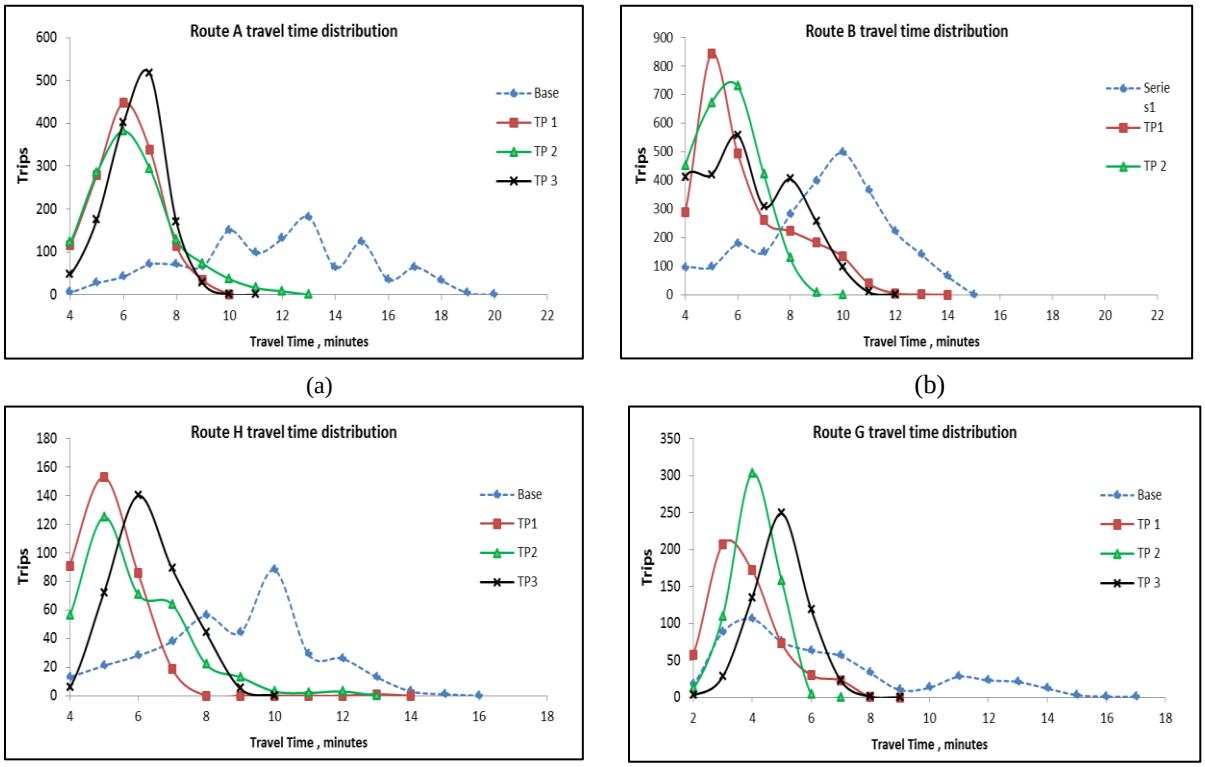


Figure 95: Network critical routes (A, B, G, and H) travel time distributions

Network non-critical routes travel time distributions

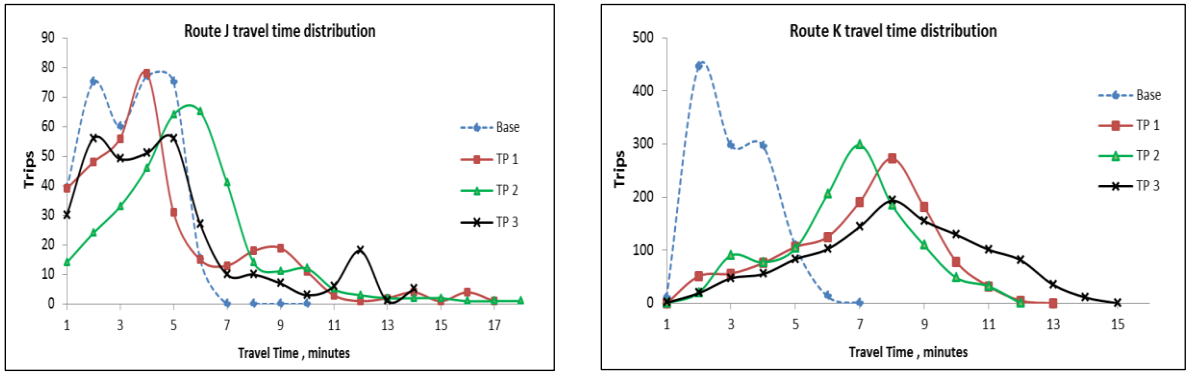


Figure 96: Network critical routes (J and K) travel time distributions

Travel time index (TTI) is a comparison between travel time conditions in the peak period to the free flow conditions. It measures of congestion level. A TTI value of 1.4 indicates that peak period travel time take 40% longer than that of under saturated conditions. The travel time index includes both

recurring and nonrecurring congestion. The TTI can be used as an indicator of the length of extra time spent in the transportation system during trip [94].

$$\text{TTI} = \frac{\text{Max travel time}}{\text{free flow travel time}}$$

Buffer time index (BTI) express the amount of extra "buffer" time needed to be on-time 95 percent of the time. BTI is a measure of travel time reliability/ unreliability [94]. As it increases travel time become more unreliable,

$$\text{Btffer time index} = \frac{95^{\text{th}}\text{Max travel time}}{\text{Average travel time}}$$

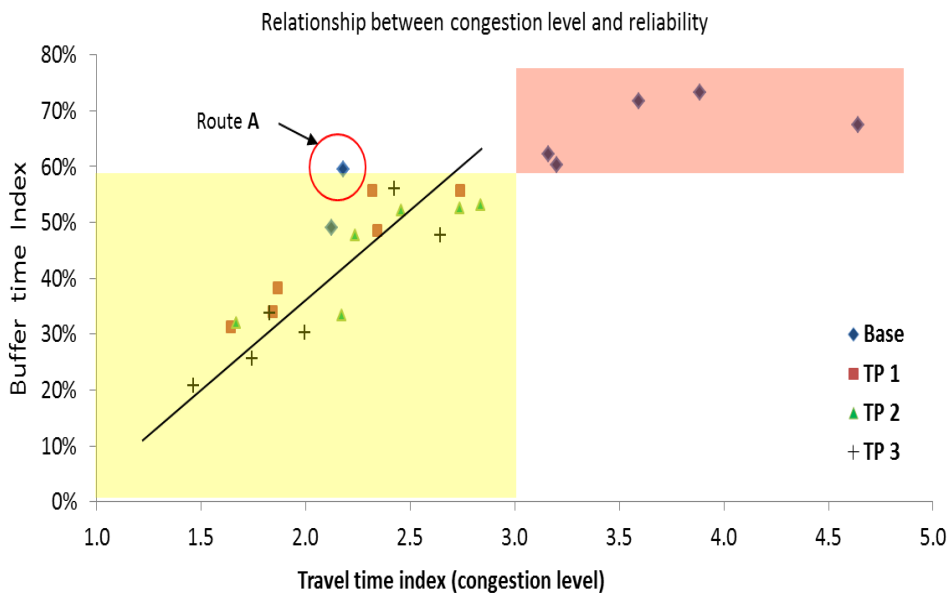


Figure 97: Relationship between congestion level and reliability

Figure 97 shows the relationship between congestion levels and reliability measure of the base and tested timing plans. In Figure 97 each point represents a route travel time performance, The plot reveals that as average congestion level increase, travel time become less reliable. The base case timing plan produced higher level of congestion in the network with reliability index of more than 60% and congestion level more than 3. While the OTS control plans produce consistency levels of congestion with reliability index between 20% and 60%. Although conditions vary from route to another, the general relationship between congestion level and reliability is present in this data set as the trend line in Figure 97 shows.

9.8 Summary and Conclusion

This chapter presents a developed timing procedure that employs several of oversaturation timing principles such as oversaturation offsets design, multi-period optimization, un-served demand carry-over, system dynamics, switching of control objectives, and critical links assignment. Applying this procedure resulted in a series of optimal timing plans with different control objectives thorough the oversaturation period. The simulation results shows that the generated control plans successfully manage the queues along the network critical links by applying both internal and external metering. Several performance measures were reported to assess the overall merits of each control plan as compared with its intended objective and the base timing (i.e., delay minimization control).There are several directions for future research:.

The simulation results of Post Oak network show significant improvements due to the application of the proposed timing procedure. Three control plan sequences were tested for a specific critical routing scenario. The first control plan applied multiple control objectives, that is, capacity maximization, queue management, and delay minimization during the loading, processing, and recovery periods, respectively. The second control plan applied the following control objectives: delay minimization, queue management, and delay minimization during the loading, processing, and recovery periods, respectively. While the third control plan applied throughput maximization during the loading and recovery regimes. The following are the summary of the major findings:

- Control plan 1 performed better in term of total delay, average delay, and average speed. However, control plan 3 had the best performance in term of system throughput as it lowers the network loading significantly during the processing regime.
- The study case results demonstrate that the developed control plans are able to manage queues and control traffic efficiently in the primary directions and keep them from spilling back into upstream links. This was expected since the arterial primary directions employ offsets that are designed carefully to avoid both spillback and starvation
- The reserve queuing capacity for side streets is an indicator of the network *external metering* performance, while the for main arterial movements' reserve queuing capacity indicate the performance of network *internal metering*. The results indicate that the network has a sizeable queuing capacity that can accommodate several failing cycles without exceeding queuing capacity (between the intersections of Chimney Rock Rd, & Rice Ave.).
- The results indicate that throughput maximization control provides network with extra reserve capacities on the critical routes during loading regime. While, delay minimization control balances the queues formations on all network links. Therefore, starting the control plan with capacity maximization control could be more beneficial during oversaturation regime.

- Throughput maximization control provides the network with extra reserve capacities on the critical routes during the loading regime. Delay minimization control balances the queues formations on all network links. Therefore, starting the control plan with throughput maximization control will reduce the amount of time that the critical routes remain oversaturated during the processing regime
- The developed control plans had significantly reduced the travel time on network critical routes A, B, G, and H, as well as significantly improve travel time reliability.
- Travel time on critical routes continues to be better the longer that the throughput maximization control period carries on . This implication supports the earlier literature and practitioners recommendations of applying throughput maximization control strategies through the oversaturated period.
- Delay minimization control was found to be more effective during the recovery period as it clears residual queues in non-critical links that buildup during the processing regime.

10 Conclusions and Future Directions

10.1 Introduction

The research presented in this dissertation defines a rational approach methodology for mitigating oversaturated conditions at signalized intersections using traffic control strategies. The methodology discussed the following components: diagnosis of the type and causes of the oversaturated condition, identification of the appropriate operational objective(s)/ control strategy(s) based on the observed condition(s), developing the corresponding optimal timing plans, and finally evaluation of the proposed optimal timing plan via a multi-objective framework.

In this research, three regimes of operation were defined and considered in oversaturated network: (1) loading, (2) processing, and (3) recovery. The research also concludes that it is necessary to understand the nature of the traffic patterns on the network before designing "optimal" control strategies. During oversaturated conditions, the temporal and spatial extent of the congestion is always changing and different movements and routes may become critical, causing different types of detrimental effects over time. A multi-objective evaluation methodology was developed to identify combinations of mitigating strategies that can improve system throughput in complex situations. Finally, two test cases were executed and showed that there are tangible performance improvements that are achievable in certain types of oversaturated scenarios.

10.2 Oversaturation Definitions

A substantial set of definitions were established in order to develop a taxonomy that defines a particular oversaturated scenario in terms of ***spatial extent*** (intersection, route, network), duration (intermittent, persistent, pervasive), ***causation*** (demand, incidents, timings, etc.), ***recurrence***, and ***symptoms*** (storage blocking, starvation, etc.). In particular, oversaturated conditions were defined as the presence of a residual queue on an approach movement after the end of the green time for particular movement. Higher level definitions for the entire network saturation states were then developed. The traffic system is not perpetually oversaturated, but rather it evolves in three phases of operation: loading phase, oversaturated "processing" phase, and recovery phase. Each mitigation strategy identified was then compared to each of these attributes to identify when each strategy would be appropriate to apply.

10.3 Critical Routes Identification

One of the major finding was the importance of identifying the critical routes through a network of intersections. While mathematical approaches akin to O-D estimation were explored, the identification of critical routes was determined to be more of an ad-hoc inspection of where the

oversaturated conditions grow and dissipate over time; since volumes are not easily measurable during oversaturation with traditional detection.

A proposed method to investigate the potential of advanced signal processing techniques in determining the transportation network critical routes was developed by performing wavelet decomposition of observed peak hour's counts profiles. This process revealed the structural breaks and identified volatility within networks significant movements. The counts profiles were filtered using *Daubechies* wavelets at detailed level 3 of decompositions which found to be the best representation of these time series. The obtained routes pattern matches the field observation during peak period. Using traditional O-D estimation methods during congested periods can be challenging, as congestion results in low volume counts on links and long travel times and critical routes need to be determined using judgment calls by the local staff after on-site observation of vehicle flows. Finally, it is very important to emphasize the criticality of this part of the analysis in the overall timing strategy framework in oversaturated conditions

10.4 Oversaturation Control Strategies Development

A wide variety of potential mitigation strategies were identified and categorized according to the oversaturation taxonomy. These control strategies include metering, phase re-service, negative offset, simultaneous offset, oversaturation offsets, cycle time adjustment, left-turn treatment, etc. After identifying the critical routes for a specific scenario, combinations of mitigations strategies can be considered based on their potential effect and level of application.

The multi-objective strategy are built on previous work by Akcelik [32], Abu-Lebdeh, Lieberman [10, 43, 51, 95, 96], Ross, Abbas [61, 89], and Rathi[29, 40], specifically in generating the principles by which the cycle length, splits, offsets, and control objectives of the methodology were developed. Then, an evaluation methodology was developed to compare the performance of various mitigating strategies under different assumptions about the critical routes in a particular test case. This methodology essentially compares all possible mitigating strategies against all possible realizations of critical route flows. The performance of each mitigating strategy is then compared for both delay and throughput measures. Although the proposed methodology shows promise, there is still significant additional work needed to fully develop a comprehensive closed-form, analytical procedure for which a set of common pattern parameters of traditional traffic controllers (cycle, split, offset, sequence, time of day schedule, etc.) could be generated.

10.5 Test Case 1

The first test case analyzed an oversaturated scenario on Reston Parkway in Herndon, VA. The studied network is an arterial that intersects with the heavily traveled Dulles Toll Road. Combinations

of Phase re-service, cycle time adjustments, green time allocation, negative offsets, and metering strategies were tested. Two sets of mitigation strategies were developed, one set of strategies apply upstream metering while the other set of strategies apply phase-re-service at downstream intersection. This case study revealed several finding such as: short cycle lengths with simultaneous offsets would minimize total system delay. Medium-length cycle times were found to maximize throughput. Strategies using metering that were focused on maximizing throughput in general decreased total delay by 35% and increased total throughput by 11%. Strategies that were focused on minimizing total delay could reduce delay by up to 35%, but increased throughput by only 7%. Strategies using phase re-service that were focused on minimizing total delay could reduce delay by up to 29%, but decreased throughput by only 1%. This test case illustrated the importance of considering different objectives when different assumptions or information about critical routes are available. In this test case, the loading, processing, and recovery regimes were not considered. Only one set of strategies were applied during the entire duration of the scenario.

10.6 Test Case 2

- In this case study of Post Oak network in Houston, TX, a timing procedure that employs several of oversaturation timing principles such as oversaturation offsets design, multi-period optimization, un-served demand carry-over, system dynamics, switching of control objectives, and critical links assignment was presented. Applying this procedure resulted in a series of optimal timing plans with different control objectives thorough the oversaturation period. In this test case, three regimes of operation (loading, processing, and recovery) were explicitly considered in designing the timing plans. The simulation results showed that the generated control plans successfully managed the queues along the network critical links by applying both internal and external metering. Several performance measures were reported to assess the overall merits of each control plan as compared with its intended objective and the base timing (i.e., delay minimization control). The study case results demonstrate that the developed control plans are able to manage queues and control traffic efficiently in the primary directions and keep them from spilling back into upstream links. The results also indicate that throughput maximization control provides network with extra reserve capacities on the critical routes during loading regime. While, delay minimization control balances the queues formations on all network links. Therefore, starting the control plan with capacity maximization control could be more beneficial during oversaturation regime.

10.7 Research Conclusion Summary

- The importance of identification of the critical routes through a network of intersections cannot be over-emphasized.
- The research methodology for developing timing plans that explicitly consider oversaturated conditions shows promise. However, the complexity of issues that must be considered during oversaturated conditions is still daunting for the development of closed-form solutions.
- Once the oversaturated condition grows beyond just a single intersection, traditional operation with "minimize delay" strategies will only tend to exacerbate the situation further since these methods will over-allocate green time to minor approaches.
- The traditional thinking such as "more green is always better" can work directly against the throughput objective because if the downstream link is already significantly queued, the upstream traffic will not be able to move anyway (i.e., starvation).
- The research identified three phases of operation during oversaturated conditions
    - Loading
    - Processing
    - Recovery
- During the loading phase, shifting the traditional operational objectives from minimizing delays to maximizing throughput can provide a measurable improvement in performance on the approaches, routes and networks that will shortly become oversaturated. In addition, early application of mitigation strategies is easier to conceptualize when the causal factors are recurrent.
- During the oversaturated phase, the traffic volumes and route proportions are such that queues and congestion are not going to be dissipated until either the traffic volumes are reduced, the route proportions are changed (i.e. drivers' avoid the area, adjust their routes, decide to travel later, etc.), or both. Queue management strategies should be applied during this phase of operation to help the system return to steady-state operation sooner during the recovery phase than continuing to apply the "normal" operational strategies designed for under-saturated conditions.
- During the recovery phase, mitigation strategies can be considered especially effective in returning the system to steady state. The test cases performed in this research generally indicate that the "recovery" period is where the most substantial performance improvements can be achieved.
- The mitigation strategies that are necessary to alleviate the oversaturated condition are likely to degrade the performance of other movements and non-critical routes (in some cases, causing those other routes to become oversaturated).The degree to which the other movements and routes

are affected is directly proportional to the aggressiveness of the mitigation applied. For example, implementing a green flush strategy will increase the side-street queues substantially, while only modifying the offsets and cycle time may not have as significant an impact on non-critical movements.

- Mitigation of oversaturated conditions will frequently involve trade-offs between the storage of traffic queues from the oversaturated movements to other less utilized movements.
- Counter-intuitively, the same control strategy that provides user-optimal delay minimization in under-saturation conditions can work against the minimization of total delay when one or more approaches become oversaturated.
- It may be necessary in some instances to induce phase failures and overflow queuing on side streets in order to maximize the flow rates on heavily oversaturated movements.
- Holistic performance measures (such as total system input or output) are much more difficult to directly measure in the field than traditional measures. A straight-forward assessment approach that follows traditional "before" and "after" principles should be augmented with the characterization of the three regimes of oversaturated operation: loading, processing, and recovery. Estimation of these regimes of operation in the before and after conditions can be helpful in determining a measurable improvement without needing to measure holistic system performance such as total system input or output.

10.8 Recommendations for Future Research

There are several needs for future research that were identified, including the following:

- There is a need to develop additional test cases that will illustrate the performance of certain mitigations strategies in specific situations. While the real-world cases that were tested in this research are instructive, in some cases it is difficult to extrapolate specific results from a particular test to other scenarios.
- All of the tests that were performed were done in simulation environment. It is well known that simulations have certain limitations in representing real-world behaviors during oversaturation. Field testing and application of mitigation strategies in s experimental sites (for example, those that were tested in simulation studies) could certainly provide valuable validation of the simulation results.
- There is a need to develop practitioner guidance with a "rules of thumb" and/or more concrete design principles for individual types of mitigation strategies. Instead of the "status quo" that considers control mitigations strategies development process as an art more than science.

- Development of an off-line timing tool(s) that can evaluate mitigation strategies in general network structures like traditional tools that are focused on under-saturated condition will be extremely helpful to practitioners.
- Development of adaptive algorithms and logic that directly consider oversaturated conditions in actuated-coordinated systems would be of benefit to the industry.

References

1. Davies, P., et al., *Assessment of Advanced Technologies for Relieving Urban Traffic Congestion*, 1991: United States. p. 110p.
2. Webster, F.V., *Traffic signal settings*, 1958. p. 44.
3. Troutbeck, R. and W. Kittelson, *Overview of the 1997 HCM update.* ITE Journal (Institute of Transportation Engineers), 1998. **68**(Compendex): p. 28-30, 32.
4. Roess, R.P., McShane, William R. and . *Traffic engineering*. Prentice Hall polytechnic series in traffic engineering2004, Englewood Cliffs, N.J.: Prentice-Hall.
5. Little, J.D.C., *The synchronization of traffic signals by mixed-integer linear programming.* Operations Research, 1966. **14**(4): p. 568-594.
6. Gazis, D.C., *OPTIMUM CONTROL OF A SYSTEM OF OVERSATURATED INTERSECTIONS.* Operations Research,, 1964. **12**(6): p. 815-831.
7. Green, D.H., *Control of oversaturated interactions.* Operational Research Quarterly, 1967. **18**(2): p. 161-173.
8. Longley, D., *A Control Strategy for Congested Computer Controlled Traffic Network.* Transportation Research, 1968. **2**: p. 391-408.
9. Pignataro, L.J., et al., *TRAFFIC CONTROL IN OVERSATURATED STREET NETWORKS.* 1978(Compendex).
10. Abu-Lebdeh, G. and R. Benekohal, *Design and evaluation of dynamic traffic management strategies for congested conditions.* Design and evaluation of dynamic traffic management strategies for congested conditions, 2003. **37**: p. 109-127.
11. Benekohal, R.F. and S.-O. Kim. *Arrival-based uniform delay model for oversaturated signalized intersections with poor progression*. 2005. National Research Council.
12. Dion, F., H. Rakha, and Y.S. Kang, *Comparison of delay estimates at under-saturated and over-saturated pre-timed signalized intersections.* Transportation Research, Part B (Methodological), 2004. **38B**(Copyright 2004, IEE): p. 99-122.
13. Fambro, D.B. and N.M. Rouphail, *Generalized delay model for signalized intersections and arterial streets.* Transportation Research Record, 1996(Compendex): p. 112-121.
14. Kang, Y.-S. *Delay, stop and queue estimation for uniform and random traffic arrivals at fixed-time signalized intersections*. 2000; Available from: http://scholar.lib.vt.edu/theses/available/etd-04202000-12070029.
15. Kim, S.-O. and R.F. Benekohal, *Comparison of control delays from CORSIM and the Highway Capacity Manual for oversaturated signalized intersections.* Journal of Transportation Engineering, 2005. **131**(Compendex): p. 917-923.
16. Cronje, W.B., *COMPARATIVE ANALYSIS OF MODELS FOR ESTIMATING DELAY FOR OVERSATURATED CONDITIONS AT FIXED-TIME TRAFFIC SIGNALS.* Transportation Research Record, 1986(Compendex): p. 48-59.
17. Cronje, W.B., *DERIVATION OF EQUATIONS FOR QUEUE LENGTH, STOPS, AND DELAY FOR FIXED-TIME TRAFFIC SIGNALS.* Transportation Research Record, 1983(Compendex): p. 93-95.
18. Dunne, M.C. and R.B. Potts, *Algorithm for traffic control.* Operations Research, 1964. **12**(6): p. 870-881.

19. Michalopoulos, P.G. and G. Stephanopoulos, *Oversaturated signal systems with queue length constraints. I. Single intersection.* Transportation Research, 1977. **11**(Copyright 1978, IEE): p. 413-21.
20. Michalopoulos, P.G. and G. Stephanopoulos, *OVERSATURATED SIGNAL SYSTEMS WITH QUEUE LENGTH CONSTRAINTS - 2. SYSTEMS OF INTERSECTIONS.* Transportation Research, 1977. **11**(Compendex): p. 423-428.
21. Patten, C.J.D., et al., *Using mobile telephones: Cognitive workload and attention resource allocation.* Accident Analysis and Prevention, 2004. **36**(3): p. 341-350.
22. khorso, K., *Development and Evaluation of Real Time Control Algorithm for Critical Intersections* 1980, University of Minnesota
23. Miller, A.J., *Computers in analysis and control of traffic.* Traffic Quarterly, 1965. **19**(4): p. 556-572.
24. Weinberg, M.L., Goldstein, H., McDade, T., Wahlan, R.,, *Digital-Computer Controllered TRaffic Signal System for Small City*, 1966, National Cooperative Highway Research Program, Washington
25. Ross, D.W., R.C. Sandys, and J.L. Schlaefli, *A COMPUTER CONTROL SCHEME FOR CRITICAL-INTERSECTION CONTROL IN AN URBAN NETWORK.* Transportation Science, 1971. **5**(Compendex): p. 141-160.
26. May, A., F. Montgomery, and D. Quinn. *Control of Congestion in Highly Congested Networks*. in *Proc. CODATU IV Conference*. 1988. Jakarta.
27. Robertson, D.I., *Transyt: A Traffic Network Study Tool*, 1969: United States. p. 44p.
28. Gordon, R.L., *TECHNIQUE FOR CONTROL OF TRAFFIC AT CRITICAL INTERSECTIONS.* 1969. **3**(Compendex): p. 279-288.
29. RATHI, A.K., *A CONTROL SCHEME FOR HIGH TRAFFIC DENSITY SECTORS.* Transportation Research Part B, 1988. **22 B**: p. 8l-101.
30. Beaird, S., T. Urbanik, and D.M. Bullock. *Traffic signal phase truncation in event of traffic flow restriction*. in *Traffic Signal Systems and Regional Systems Management 2006*. 2006. 2001 Wisconsin Avenue NW, Green Building, Washington, DC 20007, United States: National Research Council.
31. Smaglik, E.J., et al., *Event-based data collection for generating actuated controller performance measures.* Transportation Research Record, 2007(Compendex): p. 97-106.
32. Rouphail, N.M. and R. Akcelik. *Preliminary model of queue interaction at signalised paired intersections*. in *Proceedings of the 16th ARRB Conference, November 9, 1992 - November 12, 1992*. 1992. Perth, Aust: Publ by Australian Road Research Board.
33. Prosser, N.a.D., M. *A Procedure for Estimating Movement Capacities at Signalized Paired Intersections*. in *2nd International Symposium on Highway Capacity,*. 1994. Sydney, Australia.
34. Messer, C., *Extension and Application of Prosser-Dunne Model to Traffic Operation Analysis of Oversaturated, Closely Spaced Signalized Intersections.* Transportation Research Record: Journal of the Transportation Research Board, 1998. **1646**(-1): p. 106-114.
35. Kovvali, V.G., et al., *Program for Optimizing Diamond Interchanges in Oversaturated Conditions.* Transportation Research Record 2002. **1811**: p. 166-178.
36. Herrick, G.C. and C.J. Messer, *Strategies for Improving Traffic Operations at Oversaturated Diamond Interchanges*, 1992: United States. p. 87p.

37. Tian, Z., T. Urbanik, and R. Gibby. *APPLICATION OF DIAMOND INTERCHANGE CONTROL STRATEGIES AT CLOSELY-SPACED NTERSECTIONS*. in *proceedings of the 81st Transportation Research Board Annual Meeting*. 2006.
38. Messer, C.J. and N.A. Chaudhary, *Evaluation of Arlington, Texas Diamond Interchange Strategy*, 1989: United States. p. 42p.
39. Kim, Y. and C.J. Messer, *Traffic Signal Timing Models for Oversaturated Signalized Interchanges. (Interim Report March 1989-January 1992)*, 1992: United States. p. 123p.
40. Rathi, A.K., *Traffic metering. An effectiveness study.* Transportation Quarterly, 1991. **45**(Compendex): p. 421-421.
41. Lieberman, E.B., J. Chang, and E.S. Prassas, *Formulation of real-time control policy for oversaturated arterials.* Transportation Research Record, 2000(Compendex): p. 77-78.
42. Girianna, M. and R. Benekohal. *DYNAMIC SIGNAL COORDINATION FOR NETWORKS WITH OVERSATURATED INTERSECTIONS*. in *proceeding of the 81st Transportation Research Board Annual Meeting*. 2002. Washington, DC.
43. Abu-Lebdeh, G., k. Ahmed, and R. Benekohal, *Signal Coordination and Arterial Capacity in Oversaturated Conditions.* Transportation Research Record, 2000. **1727**: p. 68-76.
44. Girianna, M. and R.F. Benekohal, *Using genetic algorithms to design signal coordination for oversaturated networks.* Journal of Intelligent Transportation Systems: Technology, Planning, and Operations, 2004. **8**(Compendex): p. 117-129.
45. Quinn, D., *A Review of Queue Management Strategies*, 1992, ITS University of Leeds.
46. Daganzo, C.F., *A Pareto optimum congestion reduction scheme.* Transportation Research, Part B (Methodological), 1995. **29B**(Copyright 1995, IEE): p. 139-54.
47. Lo, H. and A. Chow, *Control Strategies for Oversaturated Traffic.* JOURNAL OF TRANSPORTATION ENGINEERING @ ASCE, 2004. **130:4**: p. 466-478.
48. Lo, H.K., *A novel traffic signal control formulation.* Transportation Research, Part A (Policy and Practice), 1999. **33A**(Copyright 1999, IEE): p. 433-48.
49. Lo, H.K., Y.C. Chan, and A.H.F. Chow. *A new dynamic traffic control system: Performance of adaptive control strategies for over-saturated traffic*. in *2001 IEEE Intelligent Transportation Systems Proceedings, August 25, 2001 - August 29, 2001*. 2001. Oakland, CA, United states: Institute of Electrical and Electronics Engineers Inc.
50. Tang-Hsien, C. and S. Guey-Yin, *Modeling and optimization of an oversaturated signalized network.* Transportation Research, Part B (Methodological), 2004. **38B**(Copyright 2004, IEE): p. 687-707.
51. Abu-Lebdeh, G., R. Benekohal, and B. Al-Omari, *Models for Right-Turn-on-Red and Their Effects on Intersection Delay.* Transportation Research Record: Journal of the Transportation Research Board, 1997. **1572**(-1): p. 131-139.
52. Park, B., C.J. Messer, and T.U. II, *Enhanced Genetic Algorithm for Signal-Timing Optimization of Oversaturated Intersections.* Transportation Research Record, 2000. **1727**: p. 32-41.
53. Shelby, S., *Design and Evaluation of Real-time Adaptive Traffic Signal Control Algorithms* 2001, University of Arizona.

54. McTrans, *TRANSYT-7F Users Guide*, U.o. Florida, Editor 2008, McTrans Center: Gainesville FL.
55. *Traffic Analysis Software Tools*, 2000: United States. p. 44p.
56. Rakha, H., *A simulation approach for modeling real-time traffic signal controls*, 1995, Queen University.
57. Bretherton, R.D. and G.T. Bowen. *Recent enhancements to SCOOT-SCOOT Version 2.4*. in *Third International Conference on Road Traffic Control (Conf. Publ. No.320), 1-3 May 1990*. 1990. London, UK: IEE.
58. Martin, P.T. and S.L.M. Hockaday, *SCOOT-an update.* ITE Journal (Institute of Transportation Engineers), 1995. **65**(Compendex): p. 44-48.
59. Lan, C.J., Messer, C. J., Chaudhary, N. A., & Chang, E. , *Compromise approach to optimize traffic signal coordination problems during unsaturated conditions.* Transportation Research Record, 1992. **1360**: p. 112-120.
60. Anderson, J.M., T.M. Sayers, and M.G.H. Bell. *Optimization of a fuzzy logic traffic signal controller by a multiobjective genetic algorithm*. in *Proceedings of the 1998 9th International Conference on Road Transport Information & Control, April 21, 1998 - April 23, 1998*. 1998. London, UK: IEE.
61. Abbas, M.M., H.A. Rakha, and P. Li. *Multi-objective strategies for timing signal systems under oversaturated conditions*. in *18th IASTED International Conference on Modelling and Simulation, MS 2007, May 30, 2007 - June 1, 2007*. 2007. Montreal, QC, Canada: Acta Press.
62. Nguyen, S. and C. Dupuis, *EFFICIENT METHOD FOR COMPUTING TRAFFIC EQUILIBRIA IN NETWORKS WITH ASYMMETRIC TRANSPORTATION COSTS.* Transportation Science, 1984. **18**(Compendex): p. 185-202.
63. Willumsen, L.G. *ESTIMATION OF TRIP MATRICES FROM VOLUME COUNTS: VALIDATION OF A MODEL UNDER CONGESTED CONDITIONS*. in *10th Summer Annual Meeting - Planning and Transport Research and Computation. Proceedings of Seminar Q: Transportation Analysis and Models.* 1982. Coventry, Engl: PRTC Education & Research Services Ltd.
64. Sherali, H.D., R. Sivanandan, and A.G. Hobeika, *A linear programming approach for synthesizing origin-destination trip tables from link traffic volumes.* Transportation Research Part B: Methodological, 1994. **28**(3): p. 213-233.
65. Bell, M.G.H., *The estimation of origin-destination matrices by constrained generalised least squares.* Transportation Research, Part B (Methodological), 1991. **25B**(Copyright 1991, IEE): p. 13-22.
66. Carey, M., C. Hendrickson, and K. Siddharthan, *A method for direct estimation of origin/destination trip matrices.* Transportation Science, 1981. **15**(Copyright 1981, IEE): p. 32-49.
67. Cascetta, E. and S. Nguyen, *A unified framework for estimating or updating origin/destination matrices from traffic counts.* Transportation Research Part B: Methodological, 1988. **22**(6): p. 437-455.
68. McNeil, S. and C. Hendrickson, *REGRESSION FORMULATION OF THE MATRIX ESTIMATION PROBLEM.* Transportation Science, 1985. **19**(Compendex): p. 278-292.
69. O'Neill, W.A., *Origin-destination trip table estimation using traffic counts*1987, Buffalo.: University of New York.

70. Van Aerde, M., H. Rakha, and H. Paramahamsan. *Estimation of Origin-Destination Matrices: Relationship Between Practical and Theoretical Considerations*. 2003. National Research Council.
71. Bell, M.G.H. and C. Cassir. *Use of the Path Flow Estimator in multi-modal networks*. in *Proceedings of the 1998 Conference on Traffic and Transportation Studies, ICTTS, July 27, 1998 - July 29, 1998*. 1998. Beijing, China: ASCE.
72. Bell, M.G.H. and S. Grosso, *The path flow estimator as a network observer.* Traffic Engineering and Control, 1998. **39**(Compendex): p. 540-549.
73. Wardrop, J.G., *Some theoretical aspects of road traffic research.* Institution of Civil Engineers -- Proceedings, 1952. **1**(Part 2): p. 325-362.
74. LeBlanc, L.J. and K. Farhangian, *Selection of a trip table which reproduces observed link flows.* Transportation Research Part B: Methodological, 1982. **16**(2): p. 83-88.
75. Cascetta, E., *ESTIMATION OF TRIP MATRICES FROM TRAFFIC COUNTS AND SURVEY DATA: A GENERALIZED LEAST SQUARES ESTIMATOR.* Transportation Research, Part B: Methodological, 1984. **18 B**(Compendex): p. 289-299.
76. Holschneider, M., *Wavelets : an analysis tool*. Vol. xiii, 423 p. :. 1995, Oxford : New York :: Clarendon Press ; Oxford University Press.
77. Samant, A. and H. Adeli, *Feature extraction for traffic incident detection using wavelet transform and linear discriminant analysis.* Computer-Aided Civil and Infrastructure Engineering, 2000. **15**(4): p. 241-250.
78. Karim, A. and H. Adeli, *Comparison of fuzzy-wavelet radial basis function neural network freeway incident detection model with California algorithm.* Journal of Transportation Engineering, 2002. **128**(1): p. 21-30.
79. Martin, R.A., K.L. Hoy, and R.H. Peterson, *Computer simulation of tolylene diisocyanate-polyol reaction.* Industrial and Engineering Chemistry -- Product Research and Development, 1967. **6**(4): p. 218-222.
80. Drew, D.R., *METHODOLOGY FOR TRANSPORTATION POLICY FORMATION AND EVALUATION.* Modeling and Simulation, Proceedings of the Annual Pittsburgh Conference, 1977. **8**(Compendex).
81. Denney, R.W., L. Head, and K. Spencer, *Signal Timing Under Saturated Conditions*, 2008: United States. p. 80p.
82. Khosla, K. and J. Williams, *Saturation Flow at Signalized Intersections During Longer Green Time.* Transportation Research Record: Journal of the Transportation Research Board, 2006. **1978**(-1): p. 61-67.
83. Adam, Z., M. Abbas, and Y. Li. *Critical routes identification method using wavelet filtering*. in *13th International IEEE Conference on Intelligent Transportation Systems, ITSC 2010, September 19, 2010 - September 22, 2010*. 2010. Funchal, Portugal: Institute of Electrical and Electronics Engineers Inc.
84. Burrus, C.S., *Introduction to wavelets and wavelet transforms : a primer*, ed. R.A. Gopinath and H. Guo. Vol. xiv, 268 p. :. 1998, Upper Saddle River, N.J. :: Prentice Hall.
85. Daubechies, I., *Ten lectures on wavelets*. Vol. xix, 357 p. :. 1992, Philadelphia, PA :: Society for Industrial and Applied Mathematics.
86. Mallat, S.G., *Theory for multiresolution signal decomposition: the wavelet representation.* IEEE Transactions on Pattern Analysis and Machine Intelligence, 1989. **11**(7): p. 674-693.

87. Daubechies, I., *Orthonormal bases of compactly supported wavelets.* Communication on Pure and Applied Mathematics, 1988. **41**: p. 909-996.
88. Roess, R.P., & McShane, W. R., *Traffic engineering*. Second ed1990, Englewood Cliffs, N.J: Prentice-Hall.
89. Abbas, M., Z. Adam, and D. Gettman, *Development and Evaluation of Optimal Arterial Control Strategies for Oversaturated Conditions.* Transportation Research Record: Journal of the Transportation Research Board, 2011. **2259**(-1): p. 242-252.
90. Manual, H.C., *Highway capacity manual*, 2000, Washington, DC.
91. Rakha, H., et al. *Systematic verification, validation and calibration of traffic simulation models*. 1998. Citeseer.
92. PTV-AG, *VISSIM 5.2 user Manual*. Vol. II. 2010: PTV.
93. *Interim Planning Activities for a Future Strategic Highway Research Program. Study 4. Capacity. Research Plan*, 2003: United States. p. 286p.
94. *Traffic Congestion and Reliability: Trends and Advanced Strategies for Congestion Mitigation*, 2005: United States. p. 24p.
95. Abu-Lebdeh, G. and R. Benekohal, *Genetic Algorithms for Traffic Signal Control and Queue Management of Oversaturated Two-Way Arterials.* Transportation Research Record, 2000. **1727**: p. 61-67.
96. Abu-Lebdeh, G., K. Ahmed, and R. Benekohal. *Modeling of Traffic Signal Output for Design of Dynamic Intelligent Control in Congested Conditions*. in *proceedings of the 83rd Transportation Research Board Annual Meeting*. 2004. Washington, DC.

Appendix

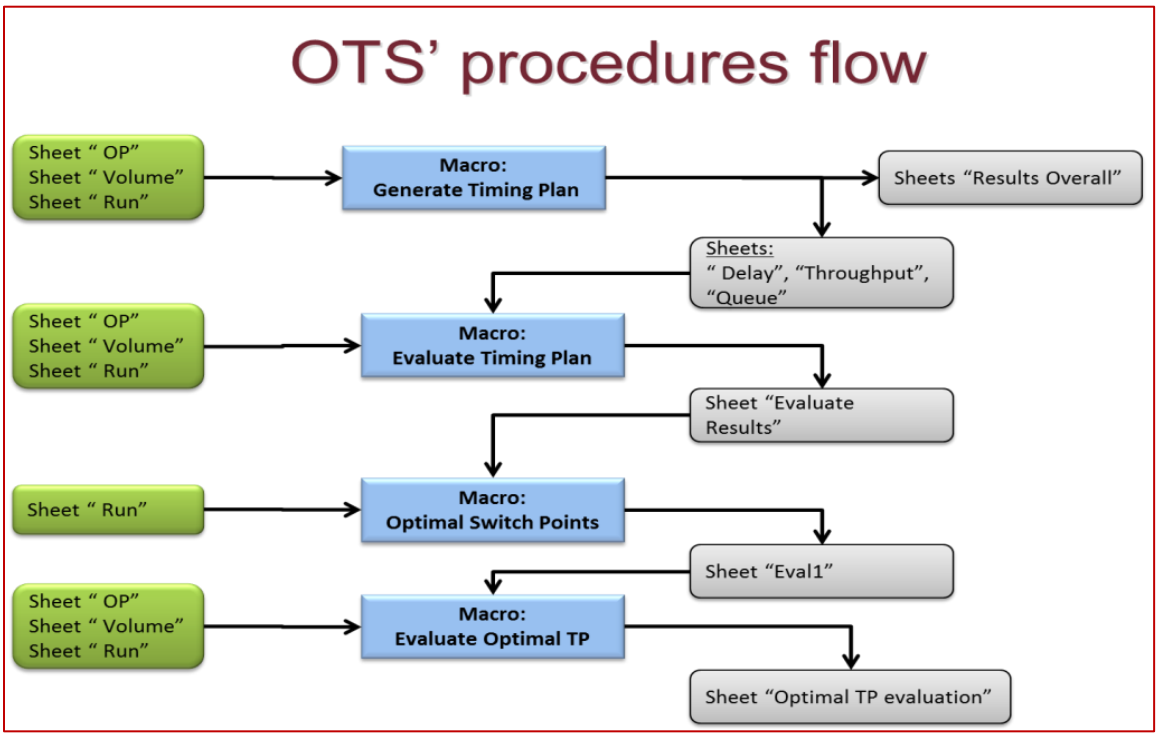


Figure 98: Oversaturation timing spreadsheet flow logic; input sheets, output sheets, and macros

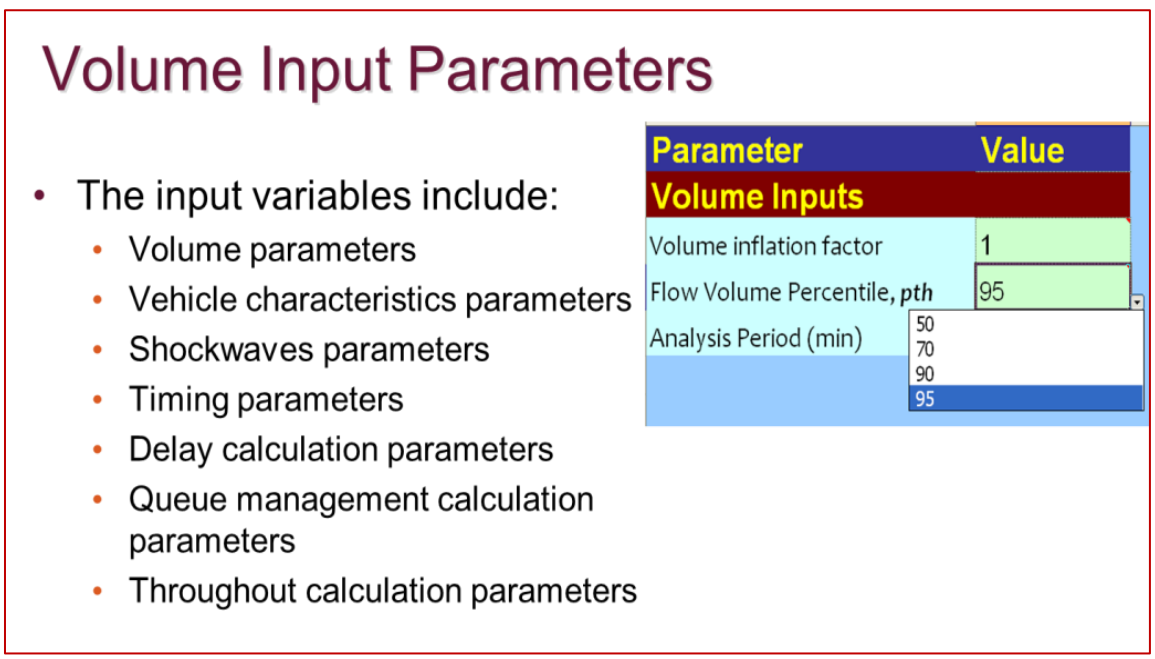


Figure 99: OTS volume parameters input cells

- **Volume inflation factor:**
- Count volume can be increase to account for the actual demand
- The factor can be used to conduct sensitivity analysis for the given counts

Parameter	Value
Volume Inputs	
Volume inflation factor	1
Flow Volume Percentile, *pth*	95
Analysis Period (min)	

50
70
90
95

- **Flow volume percentile**
- Insert the confidence (Pth) level for the arrival of the given volume count during a particular cycle length.
- And therefore estimate the expected queue length during this period
- **Analysis period**
- Insert the time interval for the count volume (e.g., 15min, 30 min or 1 hour)

Figure 100: OTS volume parameters input cells

- The shockwaves parameters
  - Queue discharge wave speed in ft/sec
  - Queue discharge rate per hour
  - Main approach speed in mph
  - Start-up loss time in sec
  - Saturation headway in sec

Shockwave Parameters	
Queue discharge rate (qs), veh/lane/hour	1800
Desired approaches speed (V), mile/hour	35
Queue discharge wave speed (u), ft/s	
Start-up loss time (l), sec	
Saturation headway (h), sec	

20
25
30
35
40
45
50
55

- These parameters are essential in determine the over-saturation offset values

Figure 101: OTS shockwaves parameters input cells

- Regarding the delay calculations, user can select from the following options:
  - Consider the residual queues in delay calculations
  - Consider the progression factor in the calculation or ignore it
  - If considering the progression factor, what is the arrival type?
  - Whither to include upstream filtering adjustment factor or not ?

Delay Calculation Parameters	
Residual queue carry over	Yes
Progression factor	Include
Arrival type	AT3
Upstream filtering/metering adjustment factor, I (side street)	0.5

Figure 102: OTS Delay calculation parameters input cells

Queue Management Parameters

- Approaches' queues lengths can be weighted and added to the corresponding delay to form a ***synthetic queue management*** formula
- Queue-to-delay weight can be inserted
- If queue length exceed its capacity, penalty can be imposed on the queue length
- User can define these parameters in the following table

1.7 Queue Mangt. Calculation Parameters	
Weighted delay formula	No
Critical movements factor	Yes
Links weights factor	Yes
Delay-to-queue weight	10
Spillback penalty	No
Spillback penalty value	999

By setting weighted delay formula to "No", the formula will calculate the total queues length

By setting critical movement factor to "yes", only critical movements queues will be considered in the queues calculations

Figure 103: OTS queue management parameters input cells

Optimizer run parameters: Multi-objective

- If multi-objective control option is selected, user has to define the objective and the starting period of each phase period in the following table

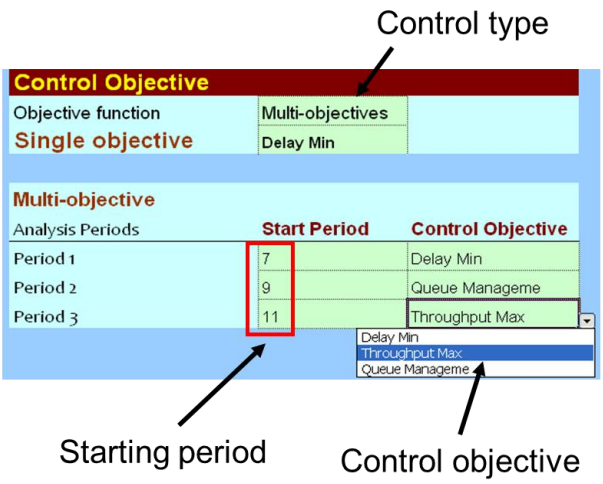
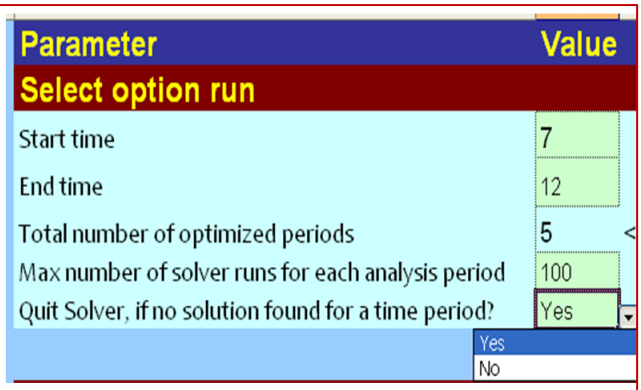


Figure 104: OTS optimizer (objective function) parameters input cells

- Select the optimization runs option:
- Start Time: The beginning of optimization period (i.e., first volume)
- End Time: The last of optimization period (i.e., first volume)



- For non-linear problem, Excel solver tend to converge to non feasible solutions.
- To overcome this problem, the OTS will continue the runs until it converge to a feasible solution. These may go to infinity.
- Therefore, user can insert the maximum number of solution trials for the optimization procedure.
- After reaching the max number of runs, Solver quit and report it frailer to converge to a feasible solution

Figure 105: OTS optimizer (multi-period) parameters input cells

Optimization parameters: Initial solution

- User can select the method that generate the initial solutions
- Two generation methods are available in the OTS:
  - Totally random
  - v/c based and scaled
- When select v/c method, the variable ***"IniSolGen"*** = 1, in case of selecting Random method, ***"IniSolGen"*** = 2.

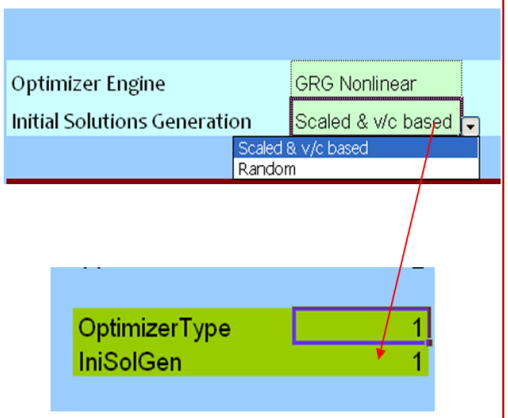


Figure 106: OTS optimizer initial solution parameters input cells

Optimization parameters: Timing constraints

- The user can choose the following option in the optimization process:
  - Optimize the cycel length (not recommended)
  - Constraints the min & max cycle length green splits
  - Constraints the min & max green splits
  - Relax the ring barrier
  - Constraints the approaches P-ratio

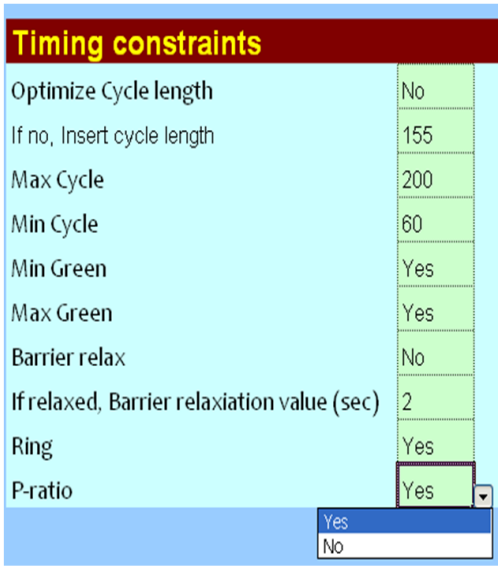


Figure 107; OTS timing control constraints input cells

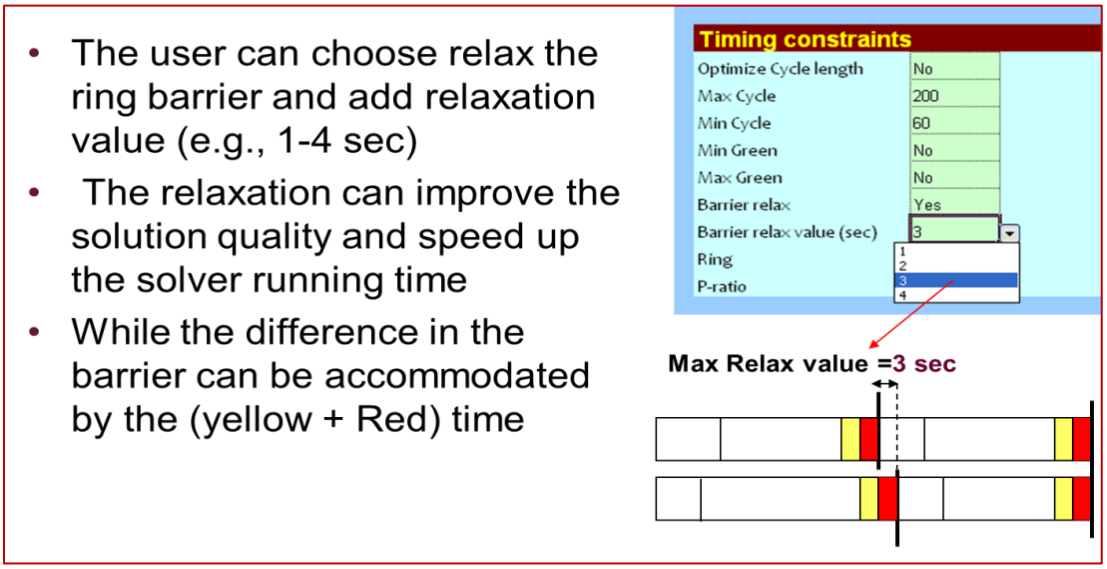


Figure 108: OTS timing control constraints (barrier relaxation) input cells

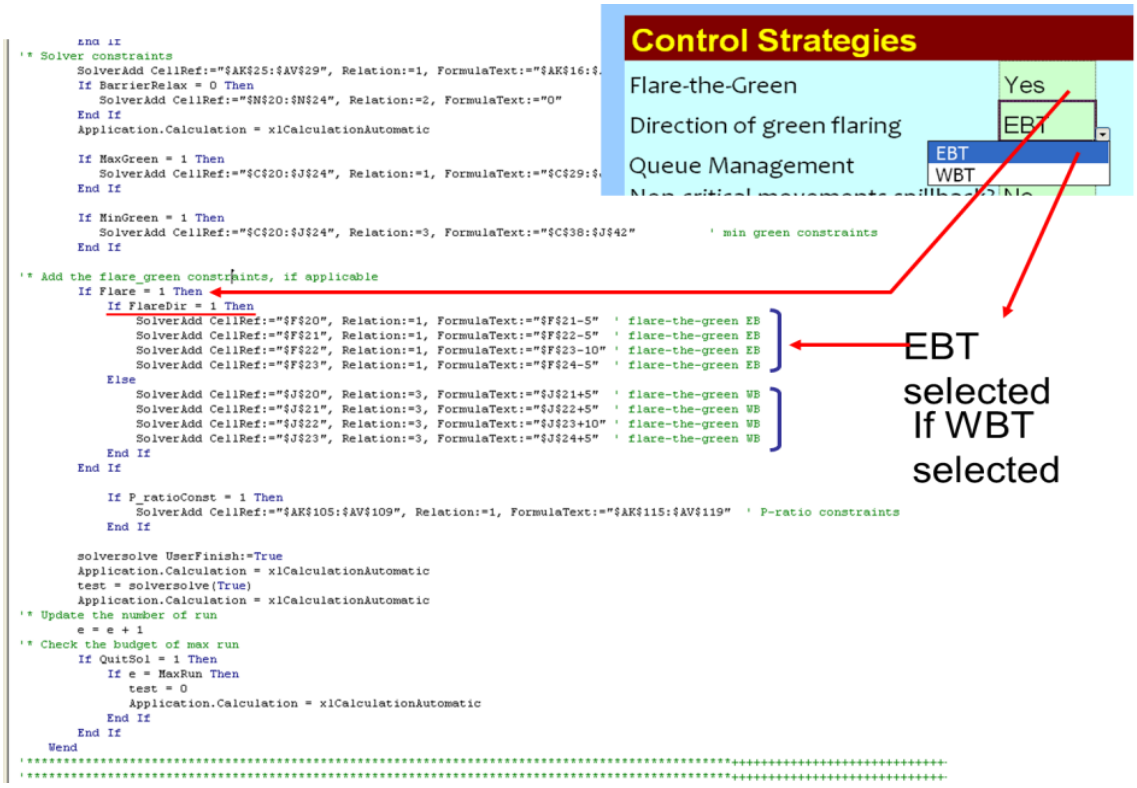


Figure 109: OTS control strategy (flare-the-green) parameters input cells

Offset design

- After optimizing the intersection splits, main approaches offsets can be determine according to the user setting
- OTS user can select the type of offset design form the following:
  - Spillback-avert
  - Starvation-avert
  - Progression
  - ect..
- Two options for simultaneous offset design are available in OTS:
  - Start of green
  - End of green
- For range option, user has to specify the link sensitivity factor from (0 -1)

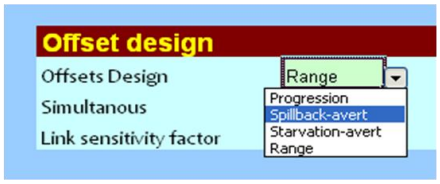
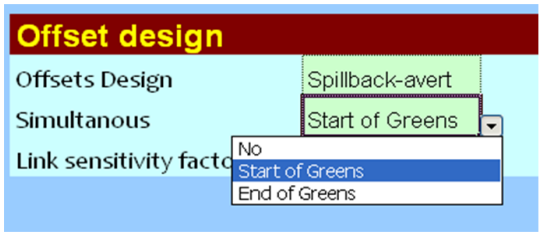
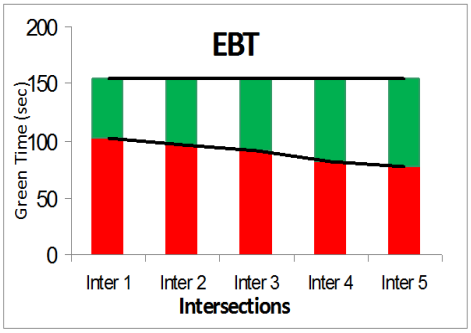


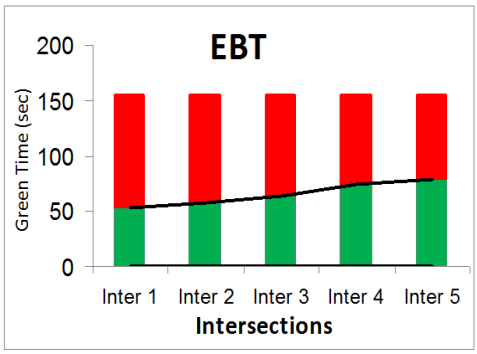
Figure 110: OTS offset design parameters input cells

- Two options for simultaneous offset design are available in OTS:
  - Start of green
  - End of green
- The figure shows the offset design
- Offset design was set to be:
  - Green is flared EBT
  - Simultaneous offset
  - Start of greens & End of green





Green is flared EBT with simultaneous offset End of green



Green is flared EBT with simultaneous offset start of green

Figure 111: OTS offset design parameters input cells

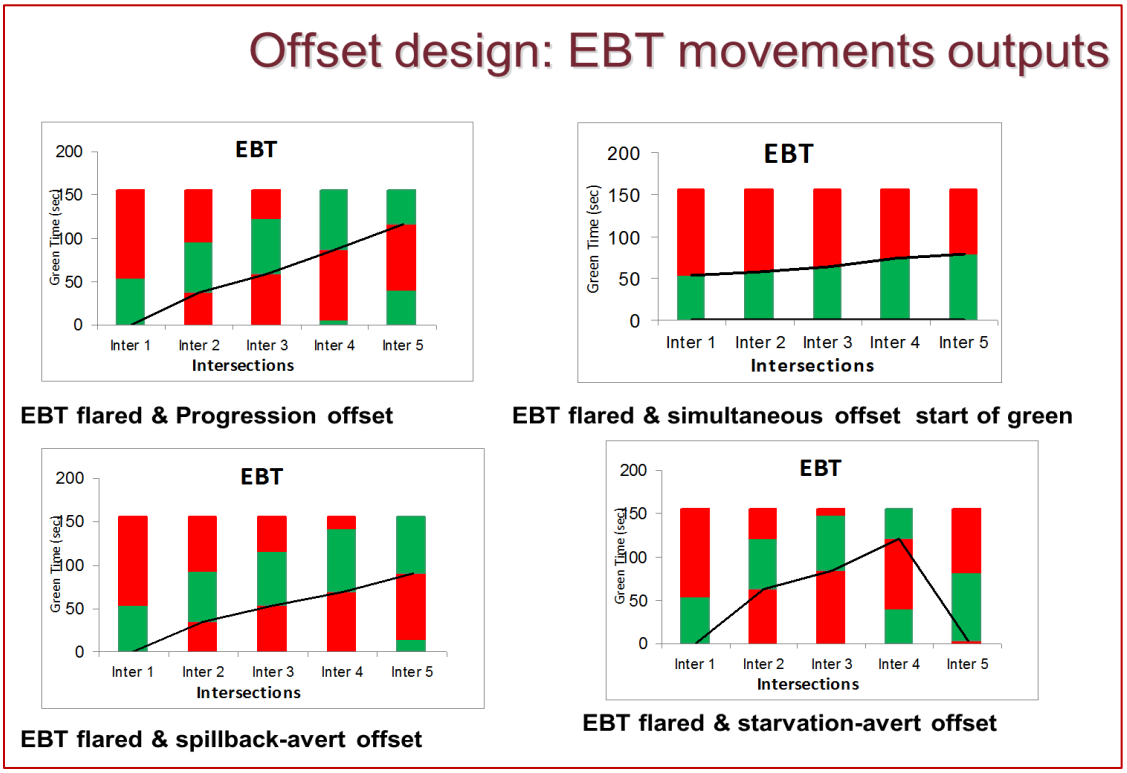


Figure 112: OTS EBT offsets time-space diagrams outputs

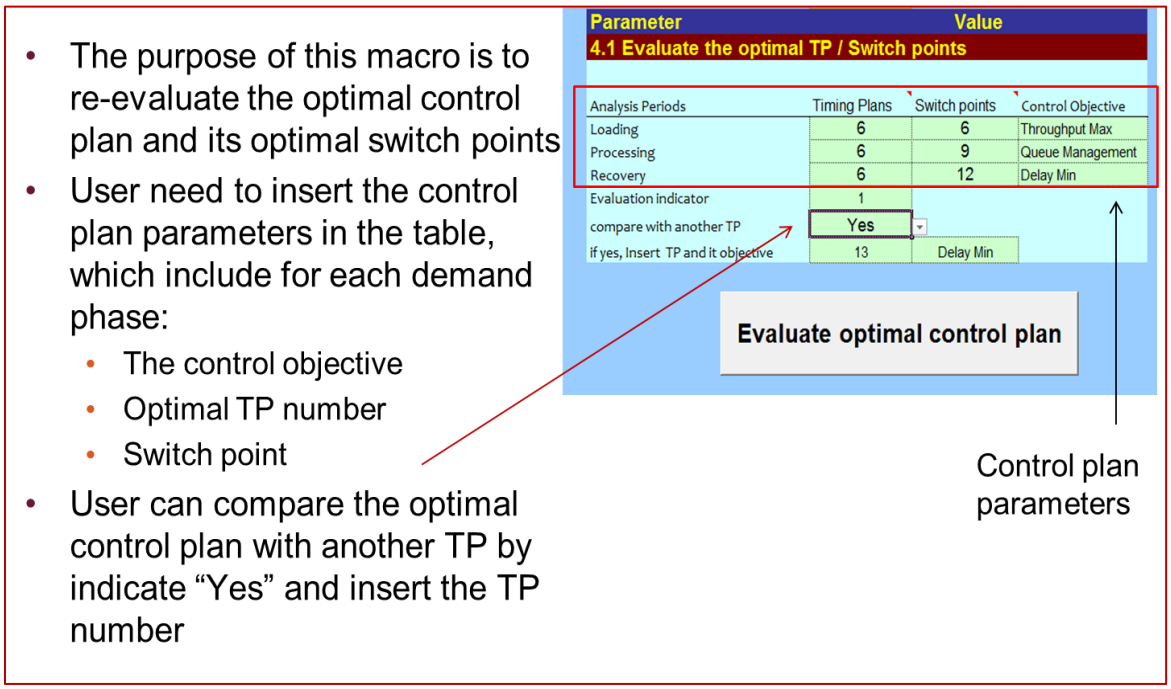


Figure 113: OTS timing plans evaluation parameters input cells

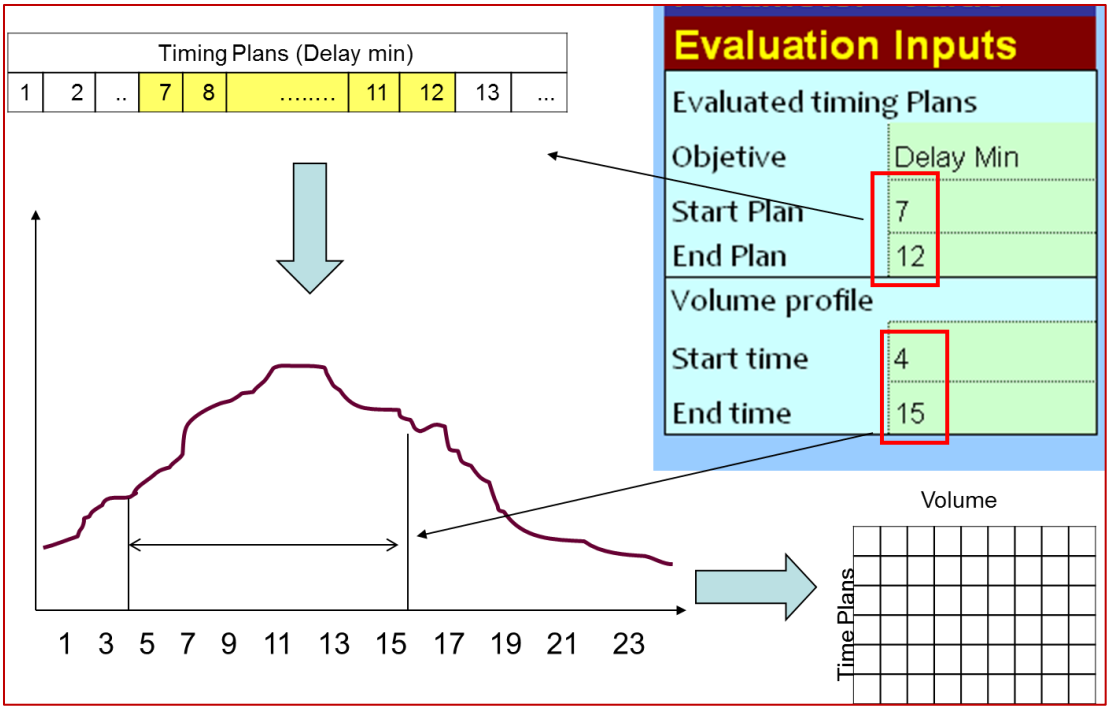


Figure 114: OTS optimal timing plans evaluation parameters input cells

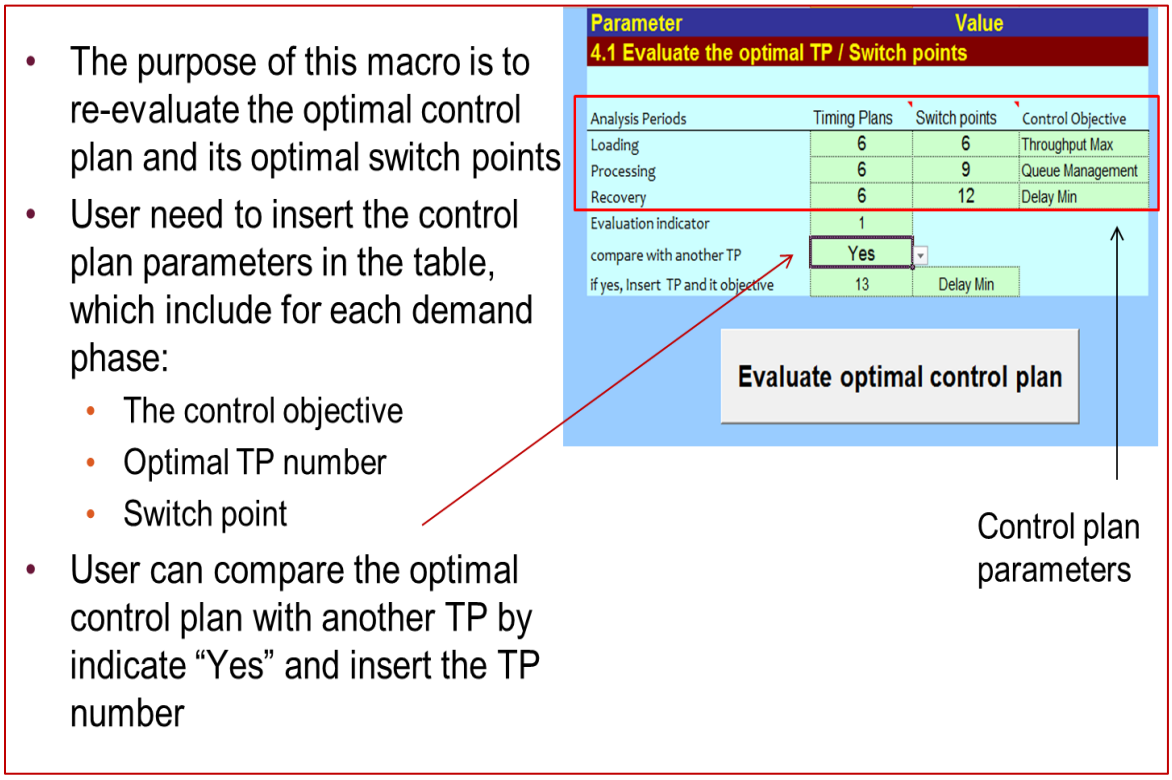


Figure 115: OTS control plans selection & switching parameters input cells

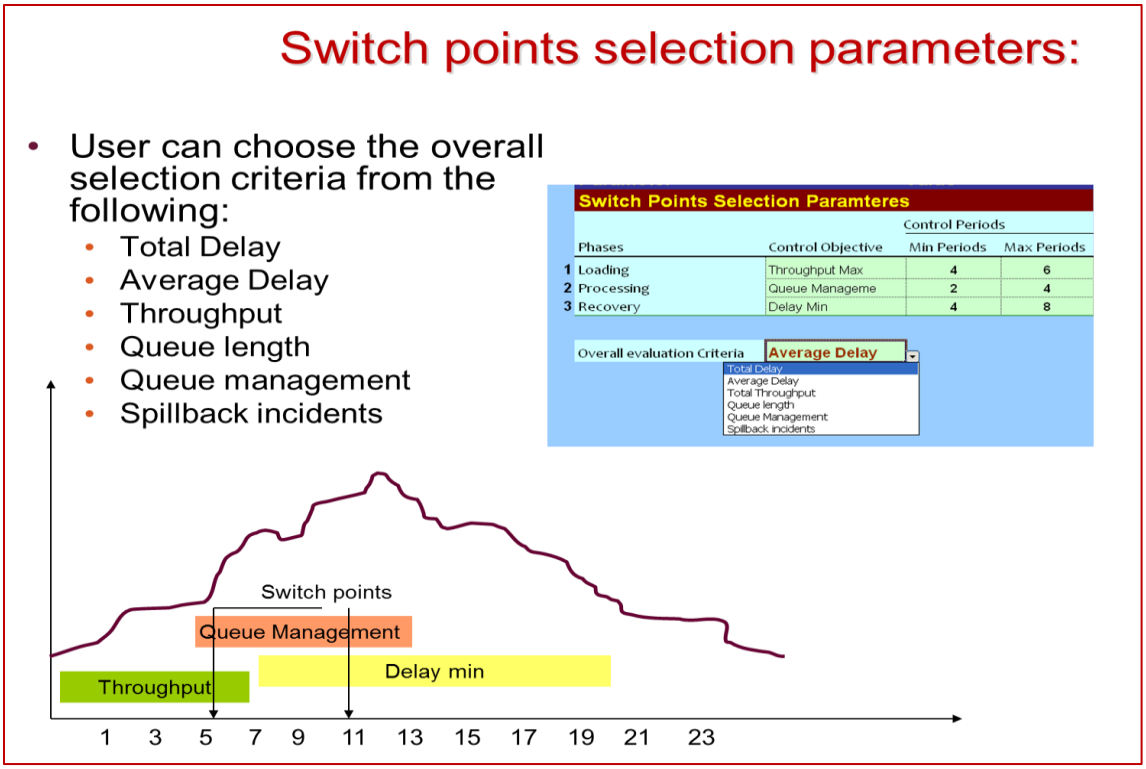


Figure 116: OTS control plans selection & switching parameters input cells

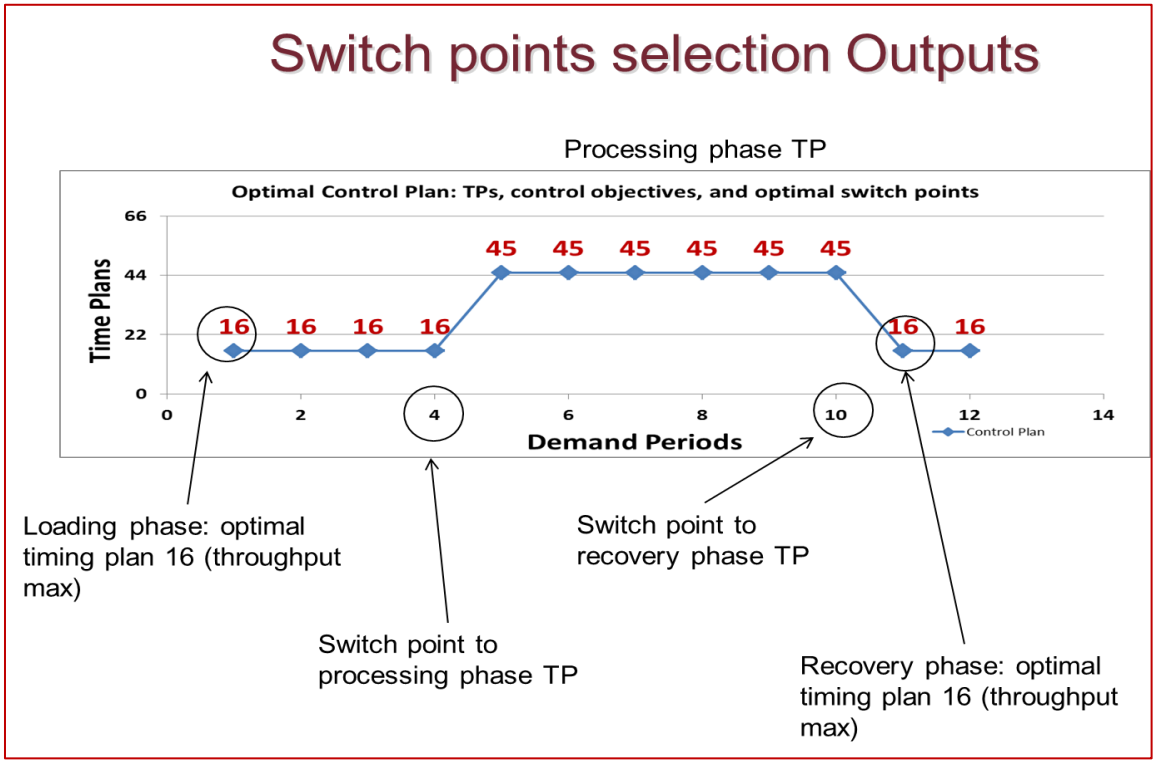


Figure 117: OTS control plans selection & switching output diagram