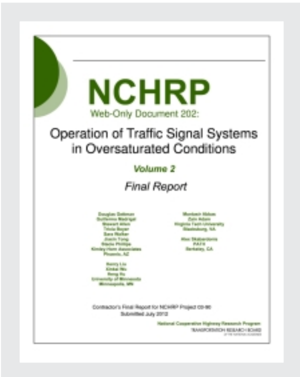


Operation of Traffic Signal Systems in Oversaturated Conditions, Volume 2 – Final Report

DETAILS

0 pages | 8.5 x 11 | PAPERBACK
ISBN 978-0-309-43314-3 | DOI 10.17226/22289

AUTHORS

Gettman, Douglas; Abbas, Montasir; Liu, Henry; and Alex Skabardonis

ACKNOWLEDGMENT

This work was sponsored by the American Association of State Highway and Transportation Officials (AASHTO), in cooperation with the Federal Highway Administration, and was conducted in the National Cooperative Highway Research Program (NCHRP), which is administered by the Transportation Research Board (TRB) of the National Academies.

THE NATIONAL ACADEMIES

Advisers to the Nation on Science, Engineering, and Medicine

The **National Academy of Sciences** is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. Upon the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Ralph J. Cicerone is president of the National Academy of Sciences.

The **National Academy of Engineering** was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. C. D. Mote, Jr., is president of the National Academy of Engineering.

The **Institute of Medicine** was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, upon its own initiative, to identify issues of medical care, research, and education. Dr. Victor J. Dzau is president of the Institute of Medicine.

The **National Research Council** was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both Academies and the Institute of Medicine. Dr. Ralph J. Cicerone and Dr. C. D. Mote, Jr., are chair and vice chair, respectively, of the National Research Council.

The **Transportation Research Board** is one of six major divisions of the National Research Council. The mission of the Transportation Research Board is to provide leadership in transportation innovation and progress through research and information exchange, conducted within a setting that is objective, interdisciplinary, and multimodal. The Board's varied activities annually engage about 7,000 engineers, scientists, and other transportation researchers and practitioners from the public and private sectors and academia, all of whom contribute their expertise in the public interest. The program is supported by state transportation departments, federal agencies including the component administrations of the U.S. Department of Transportation, and other organizations and individuals interested in the development of transportation. **www.TRB.org**

www.national-academies.org

Contents

List of Figures

List of Tables

List of Equations

Author Acknowledgments

This research reported herein was performed under NCHRP Project 03-90 by Kimley-Horn and Associates, Inc., with Virginia Tech University, University of Minnesota, and Dowling & Associates. Kimley-Horn and Associates, Inc. was the primary contractor. Dr. Douglas Gettman, Associate, was the Project Director and Principal Investigator; Ms. Tricia Boyer, P.E., Associate, Ms. Stacie Phillips, P.E., Mr. Stewart Allen, MSSE, Mr. Guillermo Madrigal, EIT, Mr. Jiaxin Tong, EIT, and Ms. Sara Walker, P.E., participated in the research. The other authors of this report are Dr. Montasir Abbas, Associate Professor of Civil Engineering at Virginia Tech and co-Principal Investigator; Zain Adam, Ph.D. Candidate at Virginia Tech; Dr. Henry Liu, Associate Professor of Civil Engineering at the University of Minnesota and co-Principal Investigator; Heng Hu and Xinkai Wu, Ph.D. candidates at the University of Minnesota, and Dr. Alexander Skabardonis, Dowling & Associates. Mr. Eric Nelson, P.E., of Advanced Traffic Solutions, LLC and Mr. Tod Eidson, MSEE, of Fourth Dimension Traffic, Inc. provided support and software services.

Abstract

This research project included work in several areas related to the operation of traffic signal systems in oversaturated traffic conditions. Quantitative measures were developed that characterize the intensity of oversaturated conditions and can be calculated based on high-resolution data from advance detectors and second-by-second phase timing data. The technique measures the length of the overflow queue at the beginning of the red light and estimates the amount of green time that is wasted either by dispersing this overflow queue or by vehicles that cannot proceed due to a downstream blockage. A heuristic procedure was developed to process the measurements of oversaturation and compute green time modifications on an oversaturated route. A methodology was developed to design timing plans that can improve system throughput and manage queues in oversaturated conditions. The methodology considers three regimes of operation of an oversaturated scenario: (1) loading, (2) processing, and (3) recovery. Individual timing plans are designed for each regime and the switching time between the plans is also determined by the methodology. It was found that application of strategies that maximize throughput in the loading and recovery phases of operation can result in performance improvements over timing plans designed for undersaturated operation. Using timing plans that are designed to minimize the degree of saturation on critical routes (i.e. manage queues) are most effective during the processing regime. An online logic tool was also developed that can directly use the quantitative oversaturation intensity measures for real-time, traffic-responsive application of pre-configured mitigation strategies. Finally, practitioner guidance was developed to assist in the process of matching mitigation strategies with specific oversaturated conditions scenarios using a systems engineering approach. Six simulation test cases were executed to evaluate and document the benefits of the techniques. In general the research found that there are tangible performance improvements that are achievable in certain types of oversaturated scenarios by applying mitigation strategies that consider overflow queues.

Executive Summary

Traffic congestion continues to grow significantly in the United States and throughout the world. Agencies tasked with managing traffic control systems are continually challenged with moving traffic in congested conditions and situations where the traffic demand exceeds the capacity of the system. Under this condition of oversaturation, typical traffic control strategies do not work as efficiently as necessary, particularly since the objectives are decidedly different when mobility is restricted. The results of the Traffic Signal Operation Self-Assessment surveys (http://www.ite.org/selfassessment/) indicate the majority of agencies involved in the operation and maintenance of traffic signal systems are stretched thin and challenged to provide adequate service to drivers in their jurisdictions.

In this project, Kimley-Horn and Associates, Inc. (KHA), the University of Minnesota, and Virginia Tech University performed research on the mitigation of oversaturated traffic conditions on arterials and networks with traffic signal systems. Dr. Alex Skabardonis provided advisory support. This research was divided into four components:

- Development of quantitative metrics for oversaturated conditions
- Identification of appropriate operational objective(s) based on the observed condition(s)
- Development of a methodology for generating timing plan strategies to address specific oversaturated scenarios
- Development of an online tool to relate measurement of oversaturated conditions with pre-configured mitigation strategies

In addition to these four areas of emphasis, the research also resulted in a rational guide for practitioners to identify oversaturated scenarios and apply appropriate strategies using systems engineering. The research focused on traffic control plans (cycle, split, offset, etc.) that can be implemented by state-of-the-practice traffic signal systems. The research did not address methodologies or issues related to freeway operations, geometric reconfiguration, re-routing, traveler information, or other strategies that seek to influence travel demand, departure time choice, or route choice. Nor does the guide explicitly address strategies or oversaturated conditions for modes of travel other than vehicles (including transit, pedestrians, bicycles, or trains).

There were seven tasks in this research project:

- Literature Review
- Identification of Diagnosis Methods and Definitions
- Identification of Strategies and Objectives
- Synthesis of Interviews with Experts

- Interim Report
- Preparation of Practitioner Guidance
- Application of Guidance and Strategies on Test Scenarios

Summary of Task 1: Literature Review

Development of strategies to handle oversaturated conditions is not a new concept; however literature that could directly be used for the purpose of this project was very limited. The research team found a variety of work in both diagnosis and strategy development. In diagnosis we focused on the review of techniques for measuring queues and the degree of saturation. Research on queue estimation is dominated by input-output modeling methods. These methods are limited to estimating queues up to the point of the input detector, but not. For arterial streets this requires installation of exit-side detection in order to measure a queue that is the length of the link. Such detector installation can be cost-prohibitive. Methods for measuring the degree of saturation can identify the saturation level up to the point of saturation, but estimates of saturation above “1.0” have not been shown to be reliable, except for those estimates used by SCOOT and SCATS which are not published. In this research, we developed a queue estimation methodology and quantitative measures of oversaturation that can be measured from advance detectors given high-resolution information on the phase timing.

In the review of strategies we looked at previous research on adaptive control systems, “optimal” control formulations, and various other approaches. Features of adaptive control systems are described in the literature in a qualitative manner. Concepts can be leveraged, but specific algorithms are not typically described quantitatively. Notably the features of SCOOT and SCATS that handle oversaturated conditions were found to be if…then type rules with thresholds that change some parameters or impose additional constraints on certain decision variables in the formulation. Descriptions of these features are not accompanied by research indicating their ability to be effective in the real world or even in simulation. Optimal control formulations found in the literature all require information on traffic volumes, queue lengths, or both. Volume information is the most difficult to obtain during oversaturation using state-of-the-practice detection systems, which makes most of the optimal formulations difficult to apply directly. In this research, we extended these “optimal” formulation concepts to consider roll-over of unserved volume from one time period to the next time period.

Summary of Task 2: Definitions and Diagnosis

A substantial set of definitions were developed. In particular, oversaturated conditions were defined as the presence of an overflow queue on a traffic movement after the termination of the green time for that movement. Higher level definitions were then developed for approaches, phases, routes, arterials, and networks. A taxonomy was developed to describe a particular oversaturated scenario in terms of spatial extent (intersection, route, network), duration (intermittent, persistent, pervasive), causation (demand, incidents, timings), recurrence, and symptoms (storage blocking, starvation). In Task 3, each mitigation strategy that was identified was compared to each of these taxonomy attributes to identify when and where each strategy would be appropriate to apply. This taxonomy helps to guide thinking, but it is not prescriptive in nature.

For diagnosis, a methodology was developed to estimate queue length from second-by-second occupancy data from advance detectors and second-by-second data for phase timing. This methodology allows measurement of queues that grow substantially upstream of the detector location. In addition, two quantitative measures of oversaturation intensity, TOSI and SOSI, were developed. These measures (TOSI and SOSI) quantify the relationship between the length of the overflow queue at an intersection approach or movement with the available green time. SOSI measures how much green time is wasted when vehicles cannot move due to downstream blockage. TOSI measures how much green time is spent dissipating overflow queuing from the previous cycle. Oversaturated routes, arterials, intersections, networks, and so on are then defined as having TOSI and/or SOSI > 0 on the constituent approaches and movements at the same time. The characteristics of how TOSI and SOSI change over time are described in Task 2 in field tests in Minneapolis on TH55.

Summary of Task 3: Development of Objectives and Strategies

In Task 3, we define the three broad operational objectives that cover the regimes of potential operating conditions, (1) minimizing (user) delay, (2) maximizing throughput, and (3) managing queues. The minimize delay objective drives strategies and operational principles that assume undersaturated operation and encapsulates all objectives that might be considered "effective" (or perhaps only acceptable) during undersaturated operation. Minimization of *user* delay is the traditional basis for most actuated coordinated signal timing in North America. Avoiding phase failures is an "equitable" traffic management policy; one that over-emphasizes the importance of light traffic movements such as left turns and side streets. It does not minimize *total* delay, it rather minimizes each driver's perception of being delayed at the signal.

"Progression" is an objective that blurs the lines between "minimize delay" and "maximize throughput". Progression is achieved in signal systems by arranging for the green times to be consecutively opened (by way of setting offsets) in a desired travel direction to allow vehicles to continue through a sequence of intersections without stopping. By carefully setting the offset values, the objective of minimizing delay (equity treatment for all users) can be satisfied at individual intersections while still meeting the objective of progression and providing the "consistent user experience" that drivers' tend to expect on arterial roads. This objective is hindered when residual queues begin to form on the movements that the control scheme is trying to progress. Offsets that were designed (i.e. forward progression offsets) assuming that no queues were present will tend to exacerbate the situation.

Minimizing delay is not appropriate when the situation is oversaturated since it is no longer possible to avoid phase failures. Thus, maximizing the number of vehicles actually *served* by the intersection, with respect to the vehicles presented to the intersection (the load), is a more appropriate objective. This keeps as much of the system operational as possible, perhaps delaying movements or phases where the total traffic demand is quite low. From an equity perspective, strategies that maximize throughput might be considered to "punish" light movements to benefit the greater good. This is done by moving much heavier phases for longer amounts of time and more frequently than would be expected by the typical cycle-failure minimizing actuated control approach.

Throughput maximization strategies goals either increase input, increase output, or both. At some point, however, no further revision to the signal timing will increase maximum throughput, and queues will continue to grow until demand diminishes. When growing residual queues can no longer be relieved by maximizing throughput, then the practitioner's only choice becomes arranging the operation of signals within a network to prevent the queues from increasing the problem. This objective is denoted "queue management".

Since so little is known or published about methodologies for mitigating oversaturated conditions), a research methodology was developed to compare the performance of various mitigation strategies optimized for the three objectives. In oversaturated traffic scenarios, the methodology also explicitly considers the identification of the critical routes through a network of intersections. Mathematical approaches similar to O-D estimation were explored in this project and are considered experimental. The multi-objective strategy development and evaluation methodology built on previous work by Akcelik, Abu-Lebdeh, Lieberman, and Rathi in generating the principles by which the green times, offsets, and objective functions of the methodology were developed. These principles were extended by the research team in the methodology to address the critical routes within the network. However, 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.

This evaluation methodology compares all possible mitigation strategies against all possible realizations of critical route flows for both delay and throughput measures. Using Pareto analysis, non-dominated strategies are identified for each time period of each test case. Typically it was found that no specific mitigation is optimal for both minimizing delay and maximizing throughput. This research process is depicted in Figure 1. This methodology was applied to two test cases in Task 7. In the first test, one timing plan was applied to the oversaturated scenario through the entire simulation. In the second test, the optimization methodology was extended to consider the regimes of operation (loading, processing, and recovery). Three timing plans were applied and evaluated in the test case.

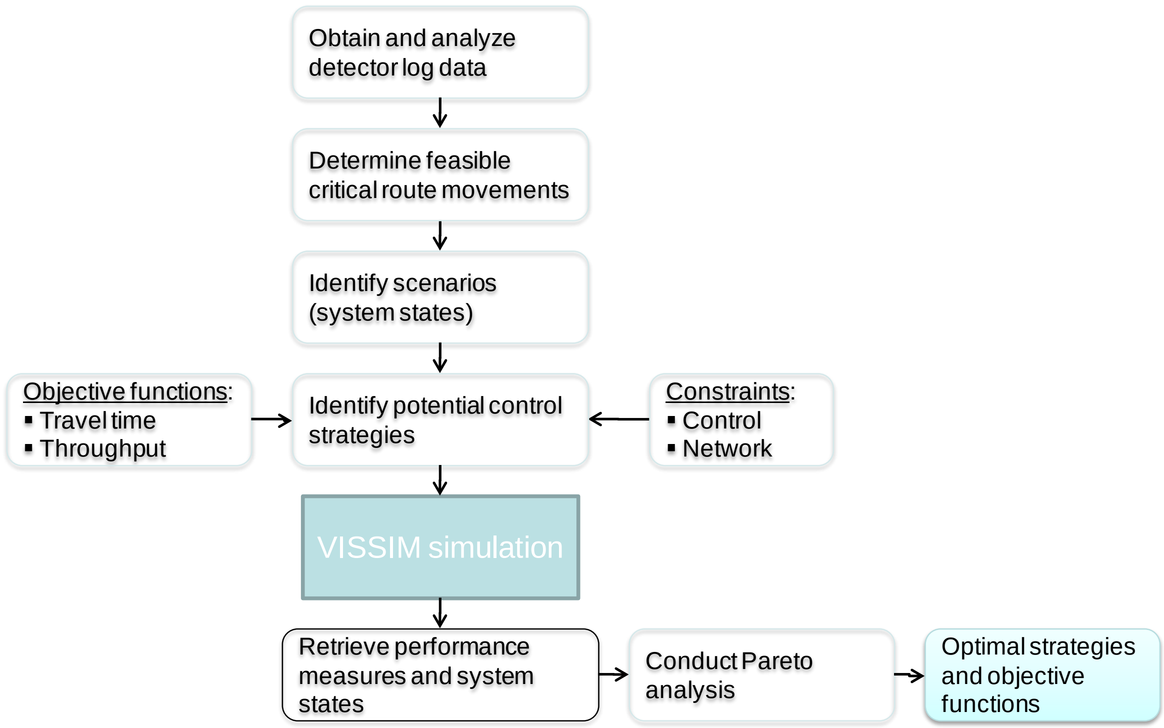


Figure 1. Research methodology for development and evaluation of mitigation strategies

In Task 3, we also qualitatively described a wide variety of potential mitigation strategies and allocated them against to the taxonomy developed in Task 2. The strategies included phase reservice, negative offsets, cycle time adjustment, left-turn type, phase sequence, dynamic lane allocation, green flush, and phase truncation. Various combinations of these strategies were tested on the two test cases in Task 7.

Summary of Expert Practitioner Interviews

In Task 4, we interviewed a number of expert practitioners. All expert practitioners viewed the management of oversaturated conditions as one of the most important issues they face from day to day. They frequently relied on personal experience and trial-and-error application than running models or performing extensive analytical analysis.

The focus of most, if not all, of the expert practitioners was first and foremost to make simple changes to splits or phase sequences to minimize delay before applying more complex approaches. Many of the expert practitioners stated their approach to implement strategies (where possible) was to prevent oversaturation from occurring rather than reacting to the issues after the fact.

Many of the expert practitioners stated the use of lower cycle times and reducing the number of phases at oversaturated locations to improve performance. Following is a list of the strategies for individual intersections cited by the expert practitioners:

- Re-allocate split time to the oversaturated phase.
- Decrease the cycle time when more than one phase is oversaturated. More opportunities to service the queue because there is a 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. One expert practitioner mentioned attempting to run three intersections on one controller. Another expert offered a rule-of-thumb that if two intersections were less than seven seconds travel time apart, they should be run on one controller.
- Use lead-lag left turns to provide more green time for oversaturated through movements.
- Use phase reservice.
- Use adaptive control methods. This was primarily mentioned as an approach to delay the onset of congestion and reduce the total amount of time that an intersection is oversaturated by managing conditions beforehand.
- Use queue detection loops to increase the max time (select an alternative MAX2 value via logic) for an oversaturated phase.
- Use queue detection to decrease phase split time (select an alternative MAX2 value via logic) when downstream link is congested.
- Use split phasing based on volume detection.
- Run intersection free.
- Do not run intersection free (this contradicts the bullet above) as this tends to exacerbate the problems for adjacent intersections.
- Change the barrier structure for certain plans by time of day (omit phases, combine movements, remove protected phases when not needed).
- Run fixed time, short cycles during construction.
- Run fixed time at interchanges – early return to green at other locations can actually harm progression.
- Use lane control signals or signs to provide flexible capacity increases by TOD / plan.
- Time detector extension times appropriately to provide swift operation.
- Allow pedestrian times to exceed cycle time for infrequent pedestrian operation on side streets.

Strategies for managing oversaturation on arterials and grids were less commonly reported by the expert practitioners than strategies for individual intersections. Most expert practitioners stated they applied both negative and simultaneous offsets for queue management on arterials. Double cycling and combinations of harmonics (3:2, 3:1, etc.) were reported as methods they previously used for handling atypical demand at one or more intersections.

At the network level, most expert practitioners reported metering as the predominant approach for alleviating gridlock conditions in critical areas. Identifying metering locations was suggested to be done on a case-by-case basis. Some expert practitioners did not have CBD experience to draw from, but in general the least amount of time during the interviews was spent discussing oversaturation issues related to grids. Many of the expert practitioners mentioned having tried traffic-responsive methods in the past with mixed results, particularly due to waffling or difficulty in setting up the parameters so that the actions taken by a traffic-responsive system were reasonable.

Few experts reported having access to performance measures from central systems that were adequate for diagnosing problems with oversaturation. All agreed that, in general, more diagnostics from both the local controller and the central system would be helpful to improve operations. One expert practitioner described a smart archiving capability that would allow queries to be processed such as “provide the top 10 heaviest left turns in the system over the last six months” or “tell me which non-coordinated phases maxed out most often during peak period”.

All of the expert practitioners agreed that reliable detection systems are critical for adequate operation. Most indicated that in their jurisdictions a fully-actuated detection scheme was most common. Several experts had critical remarks about the limitations of video detection systems and indicated that a hybrid detector (inductive + radar, video + radar, infra-red, etc.) is probably the most likely candidate to overcome the limitations of a particular detection technology.

In general, the expert practitioner interviews confirmed there are a number of strategies that have been and can be applied to the management of oversaturated conditions. Most techniques were described for individual intersections. Expert practitioners tend to be more interested in solving problems than documenting effectiveness, so quantitative benefits of applying a particular strategy were not described by any expert practitioner. This lack of evidence of the effectiveness of any approach is exacerbated by the fact that automated data collection becomes problematic during congested conditions. It was clear that the experts provided an adequate solution for a given situation and then moved on to other pressing field problems instead of spending additional effort to document or measure the degree of effectiveness.

Summary of Practitioner Guidance

In Task 6, we developed a Practitioners Guide for applying mitigation strategies to specific oversaturated condition problems. This guide is provided as a supplement to this final report. The guide is a stand-alone document which includes some of the material from the final report, but includes original material as well. The intended audience for the guide is a practicing traffic engineer with responsibility for designing and implementing traffic signal system timing, phasing, sequencing, and scheduling. The guidance follows a systems engineering approach to problem resolution, starting with problem characterization as illustrated in Figure 2. The goal of the initial steps of this approach is to answer several basic questions:

- How many intersections and travel directions are affected? (**Spatial extent**)
- How long does the oversaturated condition last? How does it evolve over time? How does it dissipate during recovery? (**Temporal extent**)
- How frequently does the oversaturated condition occur? (**Recurrence**)
- What is the cause or causes of this oversaturated condition? (**Causes and Symptoms**)

Subsequent steps recommend that the practitioner identify the objectives and approximate regimes of operation such as the duration of loading, processing, and recovery regimes. The next step in the process is to match appropriate mitigation strategies with the size and extent of the scenario and the objective(s) that are intended to be met. The guide provides mitigation strategies and examples of effectiveness and rules of thumb for application for some strategies.

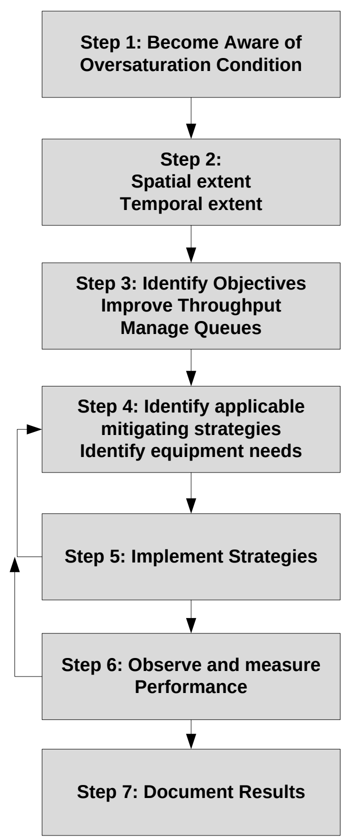


Figure 2. Process of identifying and addressing oversaturated conditions

The guide also covers the generic process of deploying the mitigation and evaluating the effectiveness.

Summary of Test Applications

In Task 7, we applied methodologies developed in this project to six test networks. Two of the tests were used in the development and testing of the methodology for developing mitigation strategies and testing those strategies using a multi-objective Pareto analysis. In the first test case we considered application of a single signal timing strategy for an entire oversaturated scenario. In the second test case, we explicitly considered the three regimes of the scenario in applying a sequence of three signal timing plans during the three operational regimes. Two other networks were used in development and testing of strategies directly related to TOSI and SOSI. TH55 was used to prove and refine the concepts of TOSI and SOSI, and test the forward-backward procedure in a relatively simple situation. The Pasadena, CA downtown network was used for developing and testing the forward-backward procedure in a stressed and complicated routing scenario. Finally, two other test cases were used to test a variety of mitigation strategies using engineering judgment and applying the guidance methodology developed in Task 6. The Windsor, ON network was also used to demonstrate the application of the if…then online mitigation strategy selection tool.

All of the test applications were simulated using Vissim with either the RBC or the Virtual D4 traffic controller. While route proportions and demand flows were changed over time, no dynamic traffic assignment was used (i.e. vehicles in the simulation did not react to the congestion conditions to change their route, change their destination, or forgo travel). The characteristics of these test cases are summarized in Table 1.

Table 1: Summary of attributes of test cases

Test Case	Config	Number of Ints	Spacing (ft)	speed (mph)	Typical phasing	Test duration	Causation	Types of symptoms	Recurrence	Critical locations	Types of mitigation	Components tested
Reston Parkway; Northern VA	Arterial with freeway interchange	14	500 to 3300	40	4, 6, 8	3 hours	Demand	All	Recurrent	2	Cycle, splits, offsets, phase reservice, gating	Timing plan development framework
Post Oak area of Houston, TX	Network	16	400 to 1800	30-40	2, 4, 6, 8	3 hours	Demand	All	Recurrent	8	Cycle, splits, offsets, phase reservice, gating	Timing plan development framework
TH55; Minneapolis, MN	Arterial	5	500 to 2600	55	4	1 hour	Preemption	Spillback, overflow queuing	Non-recurrent	2	Green extension, green truncation	Calculation of TOSI, SOSI, and queue length
Downtown grid; Pasadena, CA	Grid	22	400 to 1000	25-35	2, 4, 6, 8	2 hours	Light-rail; demand	Spillback, overflow queuing	Recurrent	5	Splits, offsets	Calculation of TOSI, SOSI, and queue length
Border Tunnel Entrance; Windsor, ON	Small Network	9	400 to 800	25-35	2, 4, 6	45 min, 1 hour, 2 hours	Incident	All	Non-recurrent	1	Many	Online feedback tool; TOSI/SOSI
Surprise, AZ	Arterial	6	2600	45	8, 6	1.5 hours, 3 hours	Planned event	All	Both	1	Many	Application of guidance

Test Cases for the Multi-Objective Pareto Analysis

Two scenarios were used to develop and test the strategy development methodology. The first test case applied a single mitigation timing plan to the entire oversaturated scenario. The second test case applied a sequence of three timing plans during the scenario to address the loading, processing, and recovery regimes. The first test case analyzed an oversaturated scenario on Reston Parkway in Herndon, VA. This scenario is an arterial that intersects with the heavily traveled Dulles Toll Road. Combinations of cycle, splits, and offsets designed for operation in oversaturated conditions were tested.

The timing plans were combined with either upstream metering on the critical route or with phase reservice for the northbound left turn at the critical interchange. In both cases it was found, in general, that short cycle lengths (e.g. 100s) with close to simultaneous offsets would minimize total system delay. Medium-length cycle times (e.g. 140s) were found to maximize throughput. In general, strategies that were optimized to maximize throughput and combined with upstream metering decreased total delay by 20% and increased total throughput by 15%. Strategies that were optimized to minimize total delay and combined with upstream metering could reduce delay by up to 40%, but increased throughput by only 7%. Strategies that focused on maximizing throughput and combined with phase reservice decreased total delay by 27% and increased total throughput by 22%. Strategies that focused on minimizing total delay with phase reservice could reduce delay by up to 63%, but increased throughput by only 5%.

In the second test case in the Post Oak area of Houston, TX, the three regimes of operation were explicitly considered in the timing plan development process. Two combinations of critical routes were evaluated. Timing plans were then designed with consideration of these critical movements and considering a sequence of timing plan changes at the beginning of the processing regime and then at the beginning of the recovery regime. The timing plan strategies considered green flaring, phase reservice, negative and simultaneous offsets, and harmonic (2:1 and 3:1) cycling at many of the intersections in the network.

In both strategies, the cycle time of the mitigations was reduced during the processing regime (from 150s to 100s or 90s) and then slightly increased during the recovery regime (from 150s to 160s). In addition, approximately half of the intersections in the network were double-cycled during the processing regime (80s or 75s cycle time) in order to manage the growth and interaction of long queues on the short network links in the interior of the network.

Both strategies provided modest 5-10% improvements over the baseline strategy for total delay. The (maximize throughput→manage queues→maximize throughput) strategy produced improved throughput on approximately 2/3 of the intersections and decreased throughput on the remaining intersections when compared to the baseline. The detriments to throughput at the intersections with reduced performance were less significant than the detriments produced by the (minimize delay→manage queues→maximize throughput) strategy. Those locations with throughput reductions were typically at the locations where SOSI was non-zero (portions of the green time were wasted because no vehicles could move). In addition, significant throughput improvements were found at the intersections that were double-cycled.

Test Cases for Direct Application of TOSI and SOSI to Re-allocate Green Time

In the test applications for direct application of the TOSI and SOSI measures, we found significant improvements were possible. In the TH55 in Minneapolis application, the TOSI and SOSI values were used to identify offset and green time re-allocation recommendations to improve the throughput performance. In the test case using the Pasadena downtown network, two scenarios were tested. The first was a single oversaturated route in one direction on an arterial. The FBP was developed and applied using the average TOSI and SOSI values measured along the route in the "do nothing" condition. The resulting adjustments to the green splits along the oversaturated route resulted in a 30% improvement to the throughput along the route. Notably because of the downstream blocking conditions along portions of the route, some of the green time splits were actually reduced, but the throughput performance along the route was still increased.

In the second test, two intersecting oversaturated routes were analyzed (one southbound and one westbound). The same FBP was applied to calculate the green time adjustments along both routes. In this test, the average throughput for southbound showed no appreciable improvement but the westbound route was improved by 10%.

Test Cases for the Application of the Practitioner Guidance and Online Evaluation of Mitigations

Two additional test applications focused on following the process defined in the practitioner guidance. The first test case evaluated mitigation strategies for handling oversaturated conditions on two critical routes competing for access to a single capacity-limited destination. In this test, a non-recurrent incident is generated at the entrance to a border-crossing tunnel. Six different mitigation strategies including metering, dynamic lane assignment, phase omits, and green time re-allocation were tested to see if more equitable allocation of green time could be provided for the two critical routes with minor effect on the non-critical routes. This test was also envisioned to explore improvement of performance to non-critical routes that were previously blocked by vehicles on the critical routes using the standard operations.

All of the mitigation strategies applied to this test case showed that more efficient use of available space could be achieved by applying metering and offset strategies to reduce the occurrence of SOSI > 0 along the critical route. The performance results for total travel time, delay, and throughput were largely inconclusive as most link-by-link performance was either not significantly different than the baseline, or performance improvements on some links were offset by performance detriments on other links.

The final test case focused on a heavily traveled arterial with pre-planned special event traffic overlaid on P.M. peak traffic that is already near oversaturated conditions. Eight different mitigation strategies were formulated and tested on this scenario. The mitigations included combinations of cycle time adjustment, green time re-allocation, negative and simultaneous offsets, dynamic lane allocation, and double cycling. In general, the results for this scenario indicated that all of the mitigations outperformed the baseline operation for both total travel time, throughput and delay measures. Most of the mitigation strategies improved the system recovery time by more than 20 minutes.

Conclusions and Directions for Further Research

Some key take-away findings were identified during the project. The first key finding was that identifying the critical routes through a network of intersections is the first a critical first step in identifying appropriate mitigations. The methodology we developed for designing timing plans that explicitly consider critical routes showed promise that alternative formulation for the optimization process can result in significant gains in total system performance. In addition, this methodology is one of the first we know of that can optimize both the timing plan parameters of individual timing plans and the sequence and duration of application of those plans during a scenario. There is still much effort necessary to bring this complicated and experimental methodology closer to being able to be applied by a typical practitioner much like they might run a tool like Synchro or Transyt.

The second key finding in this project was the development of the TOSI and SOSI quantitative measures and the development of the heuristic Forward-Backward Process to directly compute re-allocation of green times from those measurements. This process was developed and tested in an "offline" manner in this project and shown to make improvements to an oversaturated scenario. These experiments point towards online application of the FBP to continually re-compute the green time allocation in an adaptive manner.

The final key finding of the project was that it is important to consider operating the system differently during the three regimes of operation during oversaturated conditions:

- Loading
- Processing
- Recovery

During the loading regime, systems are best operated by continuing to minimize total delay or maximizing total throughput. When TOSI and SOSI become significantly > 0, strategies which manage queues on the critical routes are most effective. These strategies minimize the degree of saturation on critical routes in order to minimize SOSI and TOSI effects. Minimizing SOSI is the most important goal during the processing regime. Finally, in the recovery regime, strategies which maximize throughput are much more effective in clearing the overflow queues that were generating during the processing phase than other strategies.

Directions for Future Research

It is often said that good research generates more questions than provides answers. This project generated a number of significant directions for further research. While the goal of this project was initially to develop a guide for practitioners, there were simply too many unknowns to distill a limited amount of benefits information into a guide or generate a prescriptive unified theory of operations.

As such, the first future research need is to evaluate additional test cases that will illustrate the performance of certain mitigations in specific situations. In order to fully develop a comprehensive guide, at a minimum, an example application of each potential technique is needed. The second major need is to test and evaluate mitigation strategies in the real world. All of the tests that were performed during this project were done with simulation tools. It is well known that simulations have challenges in representation of real-world behaviors during oversaturation. Field testing and application of mitigations in real-world sites (among those that were tested in the simulation studies during this project) would certainly be a valuable research activity to follow this effort.

The practitioner guidance could be greatly improved by development of additional "rules of thumb" and more "cook book" type design principles for mitigations. This could not be achieved during this research due to the shortage of knowledge in this area. Furthermore, the combination of mitigations into comprehensive strategies is still more art than science and the methodology developed in this project is too cumbersome and complicated for use by a typical practitioner. Development of an offline analysis tool that can develop mitigation strategies in general network structures will be a valuable future research topic.

Additional research is needed on the role of thresholds, persistence time, and recovery time in the measurement of TOSI and SOSI for selection of mitigations in an online manner. In this research, we selected what seemed to be common-sense values for these parameters but did not apply any sensitivity analysis on the values of these inputs. Additional scenarios also need to be constructed and tested with the online tool for more comprehensive evaluation of the effectiveness of such a tool. To truly get such methods into real practice, system operators will have to "spec" and procure such features in upgrades or new installations of ATMS.

Finally, while it was found that real benefits can be achieved through the application of fixed-parameter timing plans, it was clear that online adaptive feedback control methods would improve the operation of oversaturated systems. In particular, it appears that offsets and green time on oversaturated critical routes need to be adjusted almost every cycle to mitigate TOSI and SOSI > 0. Negative offsets can be designed for a particular value of TOSI, but if the demand rate remains constant and the green time is left constant, the queue length will continue to grow until the link is filled. Development of adaptive algorithms and logic that directly consider oversaturated conditions in actuated-coordinated systems would be of benefit to the industry.

The FBP could be extended to a rolling-horizon formulation to take another step in that direction. Much additional research would be necessary to extend the basic heuristic of the FBP to consider phase sequencing, cycle time, protected/permitted lefts, and so on.

Chapter 1: Background and Motivation

Traffic congestion continues to grow significantly in North America and throughout the world. Agencies tasked with managing traffic control systems are frequently challenged with moving traffic in congested conditions and situations where the traffic demand exceeds the capacity of the system. The results of the 2005 and 2007 indicate Traffic Signal Operation Self-Assessment surveys, the majority of agencies involved in the operation and maintenance of traffic signal systems are stretched thin and challenged to provide adequate service to drivers in their jurisdictions.

Oversaturated traffic systems are the most complex and difficult traffic control problems. Under oversaturation, typical traffic control strategies do not work as efficiently as necessary, particularly since the objectives need to be decidedly different when mobility is restricted (e.g. “move someone, somewhere” rather than “give everyone equity treatment per cycle”). Many practitioners argue or conclude that “there is nothing that can be done when there is simply too much traffic”. This research project found that under many typical oversaturated conditions, mitigation strategies can be applied that have an appreciable effect on over-all system performance. There are two important clarifying factors. First, it is important to consider three different regimes (loading, mitigation, and recovery) of operation under which performance is measured and evaluated. Secondly, during oversaturated conditions, performance must be measured against different objectives than those that are appropriate during undersaturated operation. In particular, the objective to minimize user delay must be substituted with the objective to maximize system throughput.

Mitigation of oversaturated conditions frequently involves trade-offs between the storage of traffic queues from the oversaturated movements to other less utilized movements. This practice might be described by the idiom “borrowing from Peter to pay Paul”. Counter-intuitively, the same control strategy that provides user-optimal delay minimization in under-saturation can in some cases work against the minimization of total delay when one or more approaches become oversaturated. It may be necessary to induce cycle failures and residual queuing on side streets in order to maximize the flow rates on oversaturated movements in the main direction(s) of flow. This change in operational policy or strategy trade-offs may be challenging to communicate to organizations and citizens.

Development of strategies to handle oversaturated conditions is not a new concept. The research team found a variety of work in both diagnosis and estimation of oversaturation and control strategies and scenarios. In diagnosis and estimation we focused the review on techniques for measuring queues and degree of saturation and surrogates for the degree of saturation. While there has been quite a bit of work in the past on estimation of delay during oversaturated conditions and approaches for modeling oversaturated conditions primarily for Highway Capacity

Manual-type analysis, these research efforts are not directly applicable to this project. Research on queue estimation is dominated by input-output modeling approaches. These methods are limited to estimating queues up to the point of the input detector, but not beyond. For arterial streets this requires installation of exit-side detection in order to measure a queue that is the length of the link. Such detector installation can be cost-prohibitive. Methods for measuring the degree of saturation can identify the saturation level up to the point of saturation, but estimates of saturation above 1.0 have not been shown to be reliable, except for those estimates used by SCOOT and SCATS which are not published. In this research, we developed a queue estimation methodology and quantitative measures of oversaturation that can be measured from advance detectors given high-resolution information on the phase timing.

In the review of strategies and scenarios we looked at previous research on adaptive control systems, optimal control formulations, and various other approaches. Features of adaptive control systems are described in the literature in a qualitative manner. Concepts can be leveraged, but specific algorithms are not typically described quantitatively. Notably the features of SCOOT and SCATS that handle oversaturated conditions were found to be if…then type rules with thresholds that change some parameters or impose additional constraints on certain decision variables. Descriptions of these features are not accompanied by research indicating their ability to be effective in the real world or even in simulation. Optimal control formulations found in the literature all require information on traffic volumes, queue lengths, or both. Volume information is the most difficult to obtain during oversaturation using state-of-the-practice detection systems, which makes most of the optimal formulations difficult to apply directly.

Consolidated Results of Interviews with Expert Practitioners

The research team interviewed expert practitioners experienced with oversaturated conditions and who developed innovative solutions for addressing those conditions in the past. Each expert practitioner was interviewed for approximately one hour. The expert practitioner was provided a questionnaire that included the following general questions:

- What would you consider to be a definition of oversaturated conditions?
- What conditions cause you to take action to alleviate oversaturated conditions? How do you collect data/evidence on problems?
- What are typical techniques you use to address oversaturated conditions? Examples?
- Do we need better/different detection systems, algorithms, or approaches for detecting oversaturated conditions? What is the typical detection layout in your jurisdiction?

Feedback was varied from on several issues. It was clear that all participants viewed the management of oversaturated intersections as one of the most important issues they face day to day. Almost all expert practitioners indicated that they spent a significant amount of time dealing with oversaturation and that existing tools such as signal timing optimization software are not adequate for handling oversaturated conditions. They frequently relied on their personal experience and trial-and-error application than running models or performing extensive analytical analysis. In all cases, the expert practitioners brought up specific case studies as examples for where specific strategies were successful in addressing the problem. We have summarized detailed feedback from the expert practitioners on each of the topics above in the following sections.

Expert Practitioners Thoughts on Definitions

None of the expert practitioners were very concerned about identifying a precise definition of oversaturation. All agreed that demand greater than capacity was necessary, and that the condition needed to last at least a few cycles before it would be considered something they would spend any time to address with changes to signal timings. One expert practitioner stated that they considered at least one 15-minute period where the queue was persistent as a minimum level of congestion to call a facility oversaturated. To define the condition analytically as part of Task 2, we asked the expert practitioners if they considered an isolated intersection with a residual queue to be oversaturated even though there was no impact of the operation on other signal operations. Most agreed that an isolated condition would still be judged as oversaturated and worthy of signal timing revisions, but all considered that storing queues where there is adequate storage space (i.e. metering) was an effective strategy for many situations. Most also agreed that oversaturation is a condition that applies at the movement, approach, or phase level. One expert practitioner gave his definition at the intersection level, arguing that the changes necessary to address the oversaturated queue on one phase or movement will, by definition, degrade the performance on other movements and phases. So in his view, given the operating principles of modern signal controllers, the problem is at a minimum an intersection problem.

Several of the expert practitioners stated the definition must relate to the storage area of the approach, link, or movement. Since it is not enough to simply say that volume is greater than capacity of the green time, the ability of the traffic facility to store the excess queue will affect what strategies they would try to apply. None of the expert practitioners focused on the growth rate of the residual queue as an important indicator of oversaturation.

Expert Practitioners Thoughts on Diagnosis

Almost all of the expert practitioners relied on personal observation of the conditions in the field more often than using any analytical tools or reports. Expert practitioners in large systems (e.g. Los Angeles) and systems with CCTV deployment also relied on performance reporting tools (system detector data) and remote monitoring to verify field reports. It was common that practitioners would be notified by municipal staff and citizens when abnormal (non-recurrent) congestion and oversaturated conditions were persistent in the field.

Most expert practitioners noted certain locations were habitually oversaturated and other locations where oversaturation was recurrent but confined to peak travel periods. A third class of conditions was identified by some expert practitioners. This type of condition was either situational or caused by incident conditions such as downstream crashes. One situation cited was an on-ramp backup that would frequently cause ancillary effects at the interchange that typical signal timing cannot address. Determining the cause of the oversaturation was frequently cited as important as formulating a strategy and significant field time was typically spent in identifying the cause, and/or verifying the reports from citizens or other agency personnel.

Some expert practitioners indicated their agency uses a systematic approach, including both preventative maintenance of equipment and review of signal timing performance, to review operations throughout their jurisdiction. However, this was not common to all of the including both preventative maintenance of equipment and review of signal timing performance surveyed and the focus was on broad assessment of operational conditions rather than a specific focus on oversaturated conditions.

Expert Practitioners Thoughts on Strategies

The including both preventative maintenance of equipment and review of signal timing performance were not explicitly asked to identify differences in objective functions (maximizing throughput, managing queues) for different strategies, but common themes emerged that the expert practitioners acknowledged that the objective during oversaturation was indeed different. The focus of most, if not all, was to first make simple changes to splits or phase sequence to minimize delay before moving to more complex approaches unless the problem was identified immediately to be an arterial or grid problem. Most expert practitioners reported spending the most time on strategies at individual intersections or small groups of signals and could articulate approaches for intersection problems much more readily than describing arterial or grid methodologies. Many of the expert practitioners stated their approach to implement strategies (where possible) was to prevent oversaturation from occurring, rather than reacting to the issues after the fact.

Local Strategies

Many local strategies were referenced during the interviews. Most of the expert practitioners described local strategies in the context of one or more examples. A few described a systematic process to handling issues by attempting to address the problem with a sequence of actions. While most of the expert practitioners started their discussion of strategies at the local level, most indicated that in many situations there is always a need to consider the effects of the timings at an individual location and take those ancillary effects into account, if possible. Many of the expert practitioners addressed the use of lower cycle times and reducing the number of phases at oversaturated locations to improve performance. Following is a list of the strategies for individual intersections cited by the expert practitioners:

- Re-allocate split time to the oversaturated phase.

- Decrease the cycle time when more than one phase is oversaturated. More opportunities to service the queue because there is a 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. One expert practitioner mentioned attempting to run three intersections on one controller. Another expert offered a rule-of-thumb that if two intersections were less than seven seconds travel time apart, they should be run on one controller.
- Use lead-lag left turns to provide more green time for oversaturated through movements.
- Use phase reservice.
- Use adaptive control methods. This was primarily mentioned as an approach to delay the onset of congestion and reduce the total amount of time that an intersection is oversaturated by managing conditions beforehand.
- Use queue detection loops to increase the max time (select an alternative MAX2 value via logic) for an oversaturated phase.
- Use queue detection to decrease phase split time (select an alternative MAX2 value via logic) when downstream link is congested.
- Use split phasing based on volume detection.
- Run intersection free.
- Do not run intersection free (this contradicts the bullet above) as this tends to exacerbate the problems for adjacent intersections.
- Change the barrier structure for certain plans by time of day (omit phases, combine movements, remove protected phases when not needed).
- Run fixed time, short cycles during construction.
- Run fixed time at interchanges – early return to green at other locations can actually harm progression.
- Use lane control signals or signs to provide flexible capacity increases by TOD / plan.
- Time detector extension times appropriately to provide swift operation.
- Allow pedestrian times to exceed cycle time for infrequent pedestrian operation on side streets.

For queue detection strategies, several expert practitioners reported that this was implemented using detector delays from 7-30s (seconds) and special I/O logic. When the occupancy of the detector was 100% for the duration of the delay, the operational action was then taken. Several of the experts mentioned other uses of special I/O logic for certain conditions to improve efficiency of the controller operation. Some experts warned though that special configurations should be avoided if at all possible since this tends to reduce the maintainability of the location. It was suggested that large systems should have standardized controllers, cabinets, and detection systems, erring towards convention over configuration.

Arterial Strategies

Strategies for managing oversaturated arterials and grids were less commonly mentioned than strategies for individual intersections. Most expert practitioners stated they applied both negative and simultaneous offsets for queue management on arterials. None of the experts mentioned using an optimization tool for designing negative offset strategies or a systematic method for determining the settings. At least two of the experts indicated that it is more reliable to “tweak” offset values in the field based on observed conditions than relying on any software tools, particularly when residual queues are prevalent.

Double cycling and combinations of harmonics (3:2, 3:1, etc.) were mentioned as methods they previously used for handling atypical demand at one or more intersections. Two expert practitioners mentioned the use of coordination strategies in heavy flow conditions that do not attempt to progress flows for more than five or six intersections. One expert practitioner stated that this strategy seemed to match with driver expectations during peak hour operations. Another expert described an arterial strategy for selecting the loser intersection along the arterial and storing traffic on the approaches to this location, to the benefit of the other intersections in the system. Similarly another expert practitioner described the importance of maintaining throughput on major arterial routes with little regard to crossing route delays – particularly during peak periods.

One expert indicated that their central system allowed operators to send override commands to pre-defined groups of intersections on arterials for flushing queues (note that this capability is not unique and is available in many central systems). They described that this type of operation was typically used only for handling special events.

Network Strategies

At the network level, most expert practitioners mentioned metering as the predominant approach for alleviating gridlock conditions in critical areas. Identifying metering locations was primarily suggested to be done on a case-by-case basis. The experts could not offer any systematic methodology for locating metering locations but most indicated that quite a bit of time was spent on designing ingress and egress strategies for special events. At least one agency expert reported the deployment of a “traffic action team” in the field to supplement CCTV coverage and help keep traffic moving during significant special events. When gridlock occurs, several agency experts reported that they relied upon traffic officers to direct traffic.

Some practitioners did not have CBD experience to draw from, but in general the least amount of time during the interviews was spent discussing oversaturation issues related to grids. Those experts with CBD timing experience mentioned the use of simultaneous offsets and fixed timing to handle grid issues, particularly due to the lack of detection in most CBD areas. Similarly to the strategies for arterials, two experts offered the strategies for simultaneous offsets in grids to be implemented for three to five intersections, and then apply another set of simultaneous offsets in an alternating fashion (e.g. zero offsets alternating with 50% offsets). Many of the experts stated having tried traffic-responsive methods in the past with mixed results, particularly due to waffling or difficulty in setting up the parameters so that actions taken were predictable.

Automated Performance Measures and Central System Diagnostics

Few expert practitioners reported having access to performance measures from central systems that were adequate for diagnosing problems with oversaturation. Most did indicate they had used system detector reports including volume and occupancy information when they were available. One expert practitioner's system includes an arterial incident detection algorithm that helps to identify locations that are experiencing non-recurrent congestion. This method continuously compares current detector information with historical statistics in order to identify anomalous conditions. The expert practitioner noted that in a jurisdiction of appreciable size, automated methods such as this are necessary. The expert practitioners were split between a preference for automated handling of problems or central system features that would identify potential issues and bring them to the attention of staff. In general, all agreed that more diagnostics from either the local controller or the central system would be helpful to improve operations. One expert practitioner described a smart archiving capability that would allow queries to be processed such as "provide the top 10 heaviest left turns in the system over the last six months" or "tell me which non-coordinated phases maxed out most often during peak period."

Several expert practitioners stated that a dynamic max capability during free would be useful for a local controller. If the max for a phase was serviced several times in a row, the value would be increased (note that this feature is currently available in at least one controller firmware).

Detection Systems and Needs

All of the expert practitioners agreed that reliable detection systems are critical for adequate operation. Most indicated that in their jurisdictions a fully-actuated detection scheme was most common. Several experts had disapproving remarks about the limitations of video detection systems and indicated that a hybrid detector (inductive + radar, video + radar, infra-red, etc.) is probably the most likely candidate to overcome the limitations of a particular detection technology.

Several of the expert practitioners mentioned the need for detection systems that can monitor turning counts and the importance of deploying enough appropriate detection systems. One expert practitioner noted that long cycles were more susceptible to detector failures on side streets (wasted green time) and that shorter cycles could mitigate the effect of detector failures.

Two expert practitioners noted that adaptive control algorithms need to be advanced to the point where they can explicitly handle oversaturation and accommodate the (inevitable) failures in detection systems when they occur. European representatives mentioned reasonably effective algorithms for measuring very long queues from standard placements (25m from the stop bar) could be found in adaptive systems such as MOTION.

Summary of Expert Practitioner Interviews

In general, the expert practitioner interviews confirmed there are a number of strategies that have been and can be applied to management of oversaturated conditions. The discussions focused more on local intersection and arterials issues than on network and grid issues. This is not surprising since network problems are especially difficult to conceptualize and "get one's hands around" as stated by one expert practitioner. Expert practitioners tend to be more interested in solving problems than documenting effectiveness, so (like much of traditional traffic engineering) quantitative benefits of applying a particular strategy were not described by any expert. This lack of evidence of the approach effectiveness is exacerbated by the fact that automated data collection becomes problematic during congested condition. It was clear that the experts provided an adequate solution for a given situation and then moved on to other pressing field problems.

The expert practitioners were united in the opinion that central systems and field controllers need more comprehensive and robust performance monitoring and diagnostics. When analyzing oversaturated conditions, most of their data collection is done in the field with personal observations.. The experts were split in the opinion that existing detection technology was adequate or that new technologies were necessary. They were united in the opinion that proper site identification, installation, and technology choice was critical to obtaining reliable information from existing detection systems.

Summary of Motivation and Background

From these considerations and the guiding principles provided by the experts, the research was focused on the following four areas of emphasis:

- Development of quantitative measures of oversaturated conditions from traditional detection systems
- Development of a multi-objective methodology to develop and evaluate mitigation strategies and combinations of strategies for complex scenarios
- Development of an online tool to directly relate quantitative measures with selection of mitigation strategies
- Documentation of application test results of mitigation strategies to specific real-world test cases

In addition to these four areas, the research also resulted in a rational guide 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. The research did not address freeway operations, geometric reconfiguration, re-routing, traveler information, or other strategies that seek to influence travel demand, departure time choice, or route choice. Nor does the guide explicitly address strategies or oversaturated conditions for modes of travel other than private vehicles (including buses, pedestrians, bicycles, or trains).

There were seven tasks in this research project:

- Task 1: Literature Review
- Task 2: Development of Diagnosis Methods and Definitions
- Task 3: Development of Strategies and Objectives
- Task 4: Synthesis of Interviews with Experts
- Task 5: Interim Report
- Task 6: Preparation of Practitioner Guidance
- Task 7: Application of Guidance and Strategies on Test Scenarios

The detailed literature review will be presented in Appendix A. Chapter 2 presents the summary of the research approach for Tasks 2 and 3. Chapter 3 summarizes the findings of Task 7. Task 6 (guidance documentation) is a stand-alone document provided as a supplement to this document. Chapter 4 summarizes the project and provides conclusions and recommendations for future research.

Chapter 2: Research Approach

This research project resulted in several significant findings and developments:

- Development of new quantitative measures of oversaturation
- A research methodology for developing signal timing plans for oversaturated conditions
- A software tool that can select pre-planned mitigation strategies based on the new quantitative measures
- Evaluation of many combinations of strategies in real world examples
- Development of guidance and a process to help practitioners identify which strategies apply when and where

In this chapter, we will present the following:

- A series of definitions and framing concepts
- The motivation and theory for calculation of quantitative metrics of oversaturation
- The multi-objective methodology for developing and evaluating signal timing plans for oversaturated operation
- The design and concept of the online software tool for selection of strategies

In Chapter 3, we will summarize the findings and test cases of these research and development efforts:

- Testing and evaluation of the multi-objective development and evaluation methodology
- Testing and evaluation of the quantitative measures of oversaturation
- Testing and evaluation of a heuristic for green time re-allocation using quantitative measurement of oversaturation
- Testing and evaluation of the online strategy selection framework

Chapter 4 presents the findings, conclusions, and directions for future research.

Definitions

Oversaturated conditions can be described according to the following attributes:

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

The details of these five dimensions comprise a specific "scenario" of traffic conditions that warrant some mitigation strategies. To further clarify these dimensions, we present a series of definitions. The terms "***scenario***" and "***situation***" will be used interchangeably to describe the combination of spatial and temporal traffic conditions that describe the oversaturated traffic control problem. A ***strategy*** is a specific component or combination of traffic control actions applied to mitigate the symptoms of a ***scenario***. We will assume that the reader has an understanding of general North American traffic engineering terminology (cycle, split, offset, sequence, rings, movements, etc.) and will use these terms without definition. A glossary is included in Appendix B of this report.

As summarized in Chapter 1, a survey of expert practitioners was conducted. This survey produced a range of definitions for the concept of "oversaturation". All of the offered definitions considered that oversaturation is directly related to both the traffic demand exceeding the capacity of the intersection and the traffic control strategy in place. From these offered suggestions and our own experience, we settled on a definition of oversaturation which we tried to identify as the most basic "building block" definition. From this basic definition, further definitions will be presented.

First, we assert that the traffic *movement* is the lowest level building block of traffic control and operations at an intersection. Movements can have green time specifically allocated to them, such as a protected left turn. Movements can also be grouped together into phases for the purpose of allocation of green time, or movements can borrow green time from other movements or phases by using overlaps. Thus, we define:

> *A traffic* **movement** *is* **oversaturated** *when the traffic demand for the movement exceeds the green-time capacity such that a queue that exists at the beginning of the green time is not fully dissipated at the end of the green time for that movement.*

This basic definition of oversaturation does not immediately imply that a change in traffic control strategy is necessary or that any action is required at all. It simply describes the condition at its lowest common denominator in the context of traffic signal control. An example of this basic scenario is shown in Figure 1. The example on the left shows a queue of vehicles waiting to turn left at an intersection. Vehicles intending to turn left have been shown in green color. A "subject vehicle" is marked in the queue for reference. The illustration on the right then shows the resulting traffic condition after the left-turn green time has elapsed. The "subject vehicle" highlighted has made some progress towards the stop bar of the left turn bay, but did not proceed through the intersection on the green light. This illustrates the concept of an oversaturated movement at an intersection.

> *An* **overflow queue** *is defined as a minimum of one vehicle that is left over from a queue that could not be fully discharged during the previous green phase.*

Common sense dictates that a scenario where a queue of vehicles is dispersed and one or two vehicles are remaining after the termination of the green time is probably not a serious issue to address with alternative traffic control strategies, at least if it only lasts for one or two cycles. However, from general queuing theory we know that a sustained arrival rate (traffic demand) that exceeds the service rate (green time) of any process will result in queues that grow until the arrival rate is reduced. This can occur naturally as fewer vehicles arrive or the arriving traffic begins taking alternate routes because of the downstream congestion.

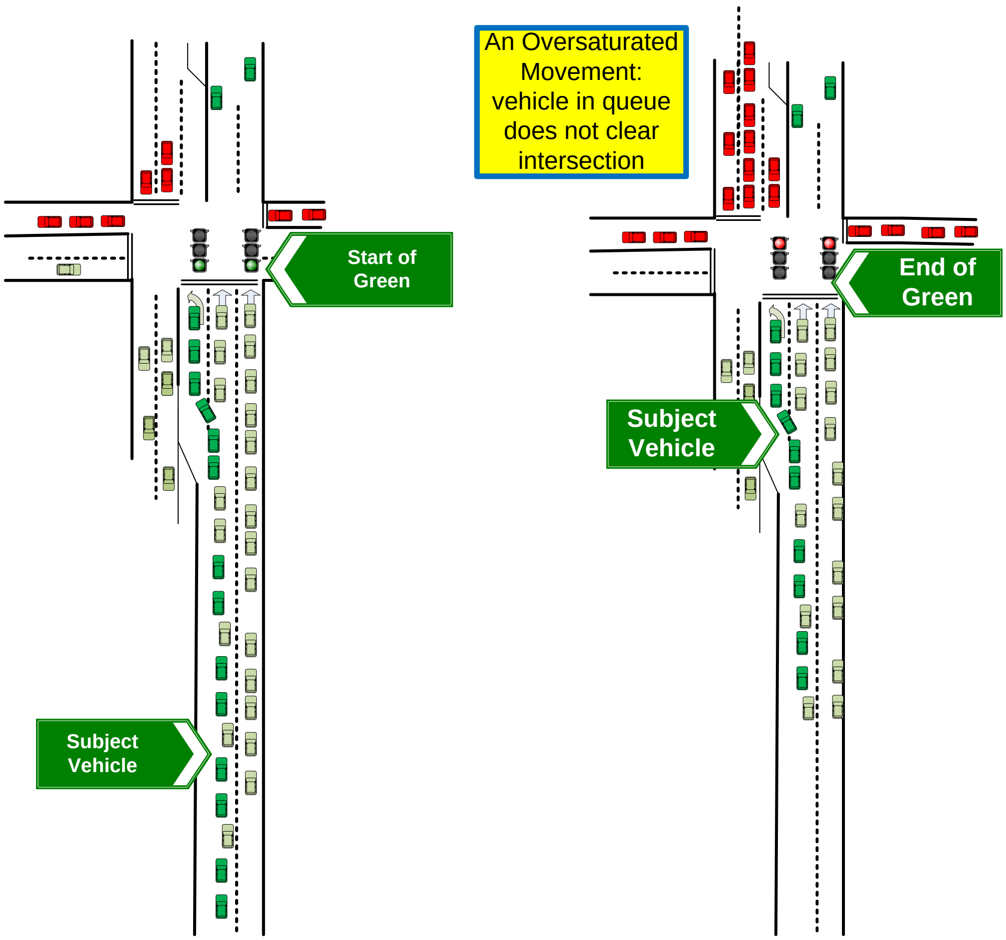


Figure 3. An oversaturated traffic movement

Additional qualifying conditions are necessary to extend this basic definition of oversaturation before changes in traffic control actions are typically required. These conditions can include:

- The degree to which the movement is oversaturated (i.e. the length of the overflow queue).
- The rate at which the oversaturation level is growing (the growth rate of the overflow queue).
- The effect of the oversaturation on other movements, approaches, and intersections.
- The length of time that the oversaturation persists.

Significantly oversaturated conditions existing on a single movement can be handled by re-timing the signal to shift green time from other under-saturated phases to the phase serving the oversaturated movement.

Extension of the Definition for Spatial Extent

From this basic condition of oversaturation on a movement, the next level of characterization of oversaturation is oversaturation on an approach to an intersection. An approach is defined as a combination of compatible traffic movements that serve traffic in the same direction of travel. A traffic movement is compatible with another movement if they do not inherently conflict (i.e. they could be served by the same traffic phase).

> *An **approach** is **oversaturated** if all movements of the approach are oversaturated or if an oversaturated movement causes "detrimental effects" to one or more of the other movements served by the approach.*

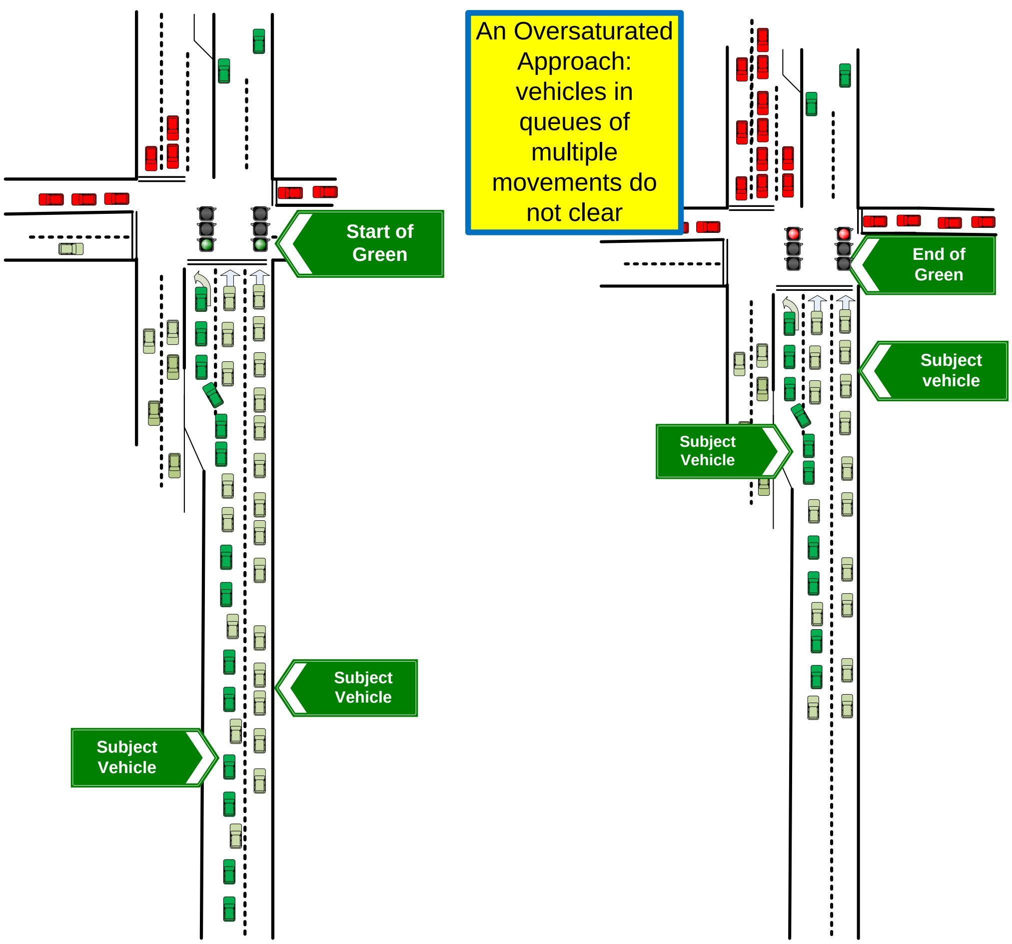


Figure 4. An oversaturated approach for both through and left-turn movements

Figure 4 illustrates the case where both the through movement and the left-turn movement are oversaturated. The “subject vehicle” tags in the figure indicate that both movements are oversaturated since neither vehicle proceeds through the intersection during the green signal. For the purpose of this illustration, and the definition it is not necessary that both movements are served by one traffic phase or separately by two phases. The oversaturation is defined on the “approach”.

Detrimental Effects

A ***detrimental effect*** or a ***symptom*** is a situation where the oversaturation on one movement causes reduction in the ability of traffic on a compatible movement (or any other movement) to utilize all of the green time allocated for that movement due to ***starvation*** or ***blocking***. Starvation is the condition where the light is green but there is no associated vehicle flow for that direction of travel.

Figure 5 illustrates the condition where the oversaturated condition on the left-turn movement creates starvation for the through movement because the vehicles that intend to turn left have blocked the ability of the through vehicles to proceed to the stop line. Thus, perhaps if the left turning movement was not oversaturated, the through movement would not have been impeded and the green time might have been adequate to satisfy the through demand.

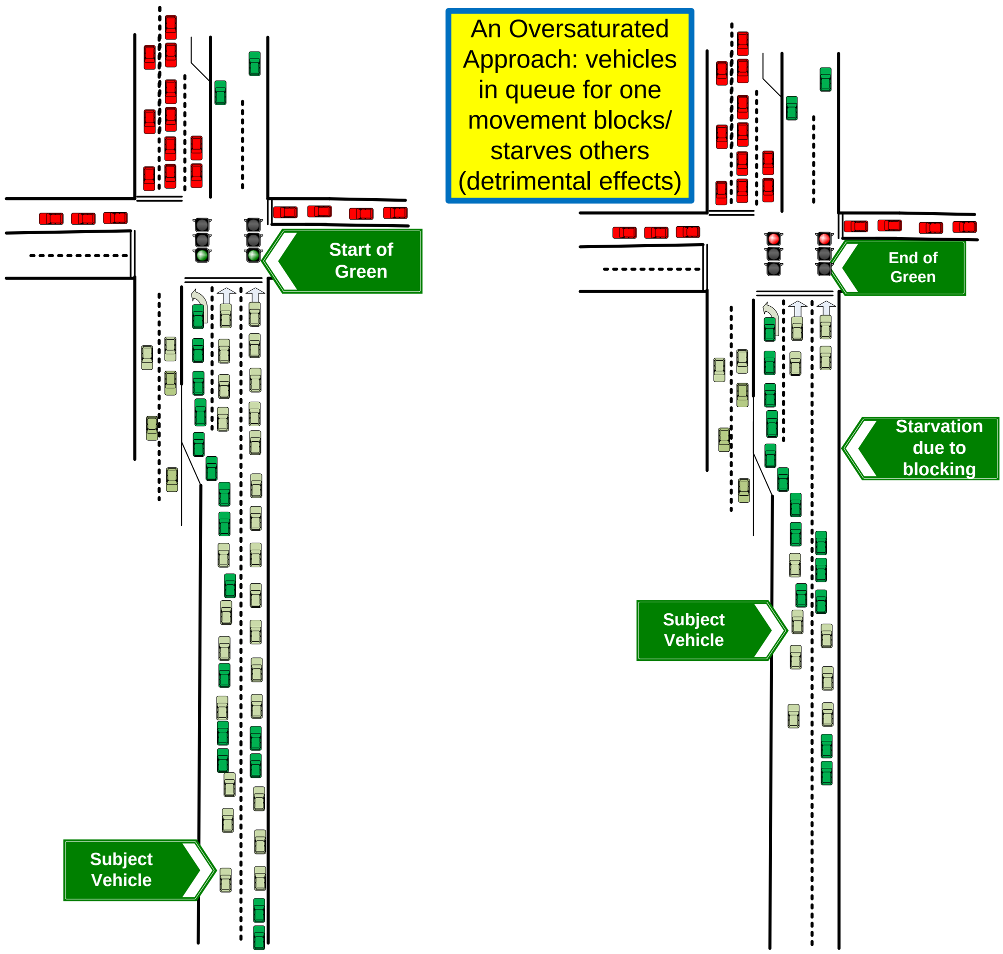


Figure 5. Illustration of oversaturated approach due to starvation

A traffic control ***phase*** would be considered oversaturated if all movements that are served by the phase are oversaturated. Oversaturated conditions that exist on single approaches or single phases can typically be addressed with re-allocation of green time from other under-saturated phases or by changes to the phase sequence. Once the oversaturated condition grows to extend past a single movement or approach, the problem trade-offs become more challenging.

> *An* ***intersection*** *is considered to be* ***oversaturated*** *if two or more incompatible traffic movements at the intersection are oversaturated.*

Blocking and Non-Blocking Conditions

Oversaturation at an intersection can have blocking or non-blocking conditions.

> *A* ***blocking condition*** *exists when the queues on one movement prevent one or more other movements at the intersection from proceeding through the intersection during its associated green time.*

Figure 6 below illustrates an oversaturated intersection where the vehicles in the northbound left turn bay are blocking the movement of the vehicles on the eastbound approach. In a non-blocking situation at an oversaturated intersection, there is no impedance of one approach flow by another incompatible flow.

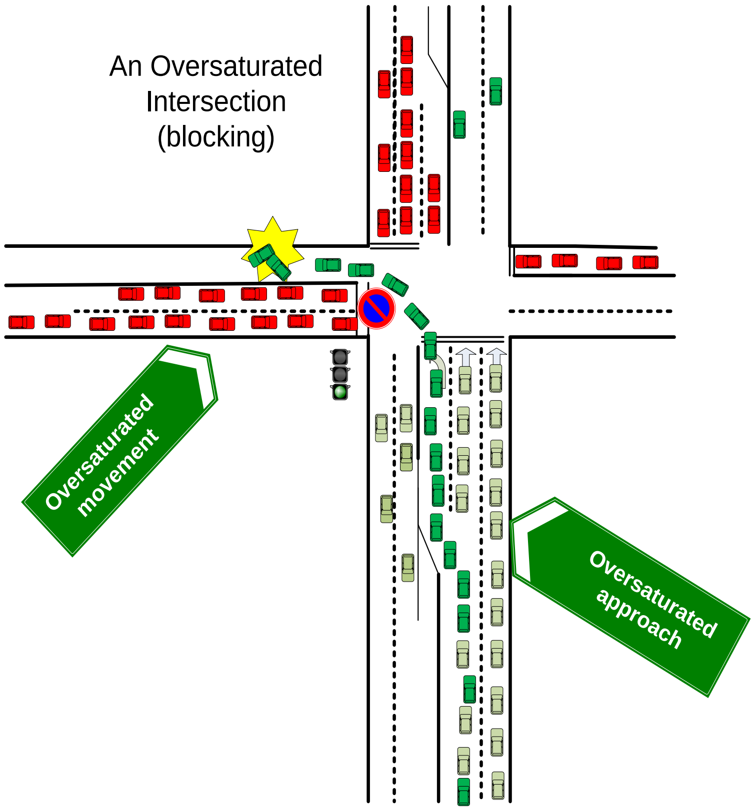


Figure 6. Oversaturation at an intersection caused by blocking

An intersection with a blocking condition is more complex to address than a scenario where blocking does not occur. More careful consideration of the effects of green-time re-allocation must be taken into account before a specific mitigation action can be taken when there is blocking. In most locales, blocking the intersection is illegal and most drivers will comply with these common sense rules.

Oversaturation on a Route

A route is a useful building block definition to identify oversaturation problems that are larger than an individual intersection. The term "route" is not meant to construe an origin-destination pair or any considerable distance from the beginning point to the ending point. Figure 7 illustrates an oversaturated route comprised of a northbound through movement (intersection G→F), a northbound left turn movement (intersection F→B), and then an eastbound through movement (intersection B). These three movements comprise an oversaturated route when they are oversaturated at the same time. Oversaturation on a route can also be a source of blocking conditions at intersections.

> *A **route** is considered to be **oversaturated** if two or more compatible movements on a single travel path through a series of intersections simultaneously have oversaturated conditions.*

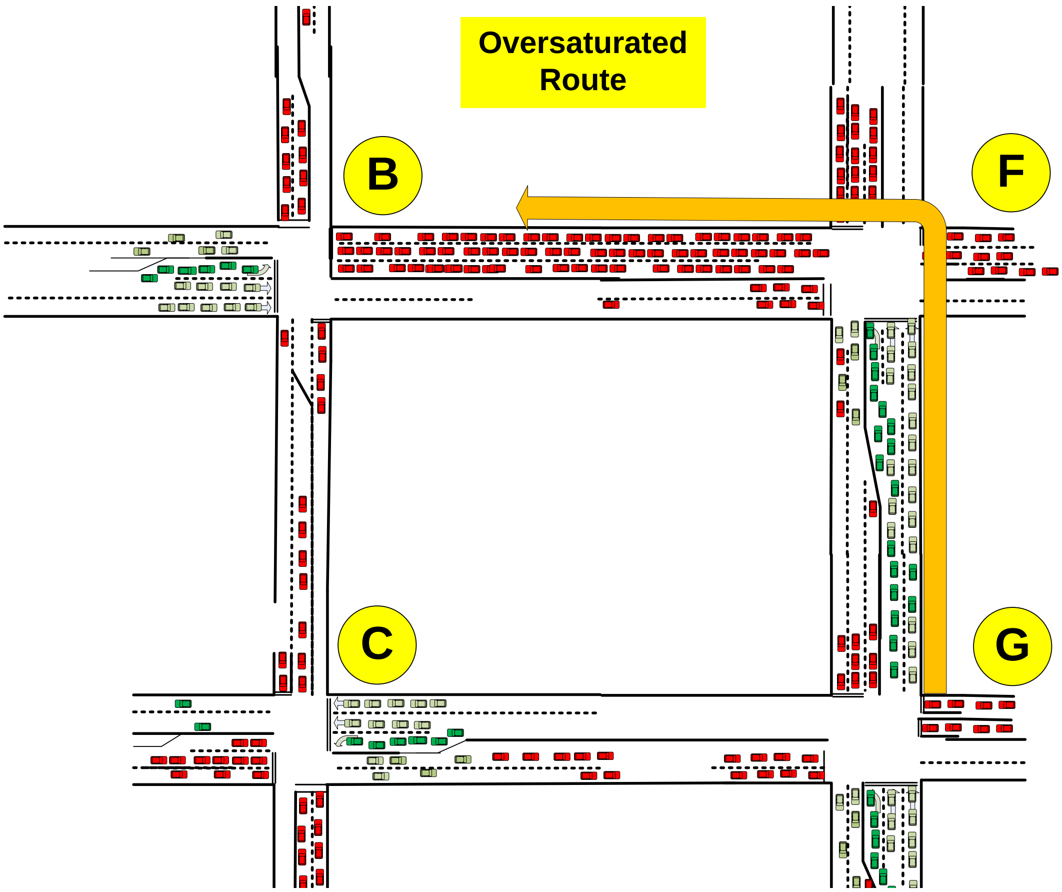


Figure 7. Oversaturated condition on a route

Oversaturation on a Network

Multiple routes that are oversaturated at the same time and that interact with each other define an oversaturated network. An example of this type of situation is illustrated in Figure 8. Figure 8 shows an example of several routes and approaches that are oversaturated at the same time including all of the intersections in the figure except intersections C and D.

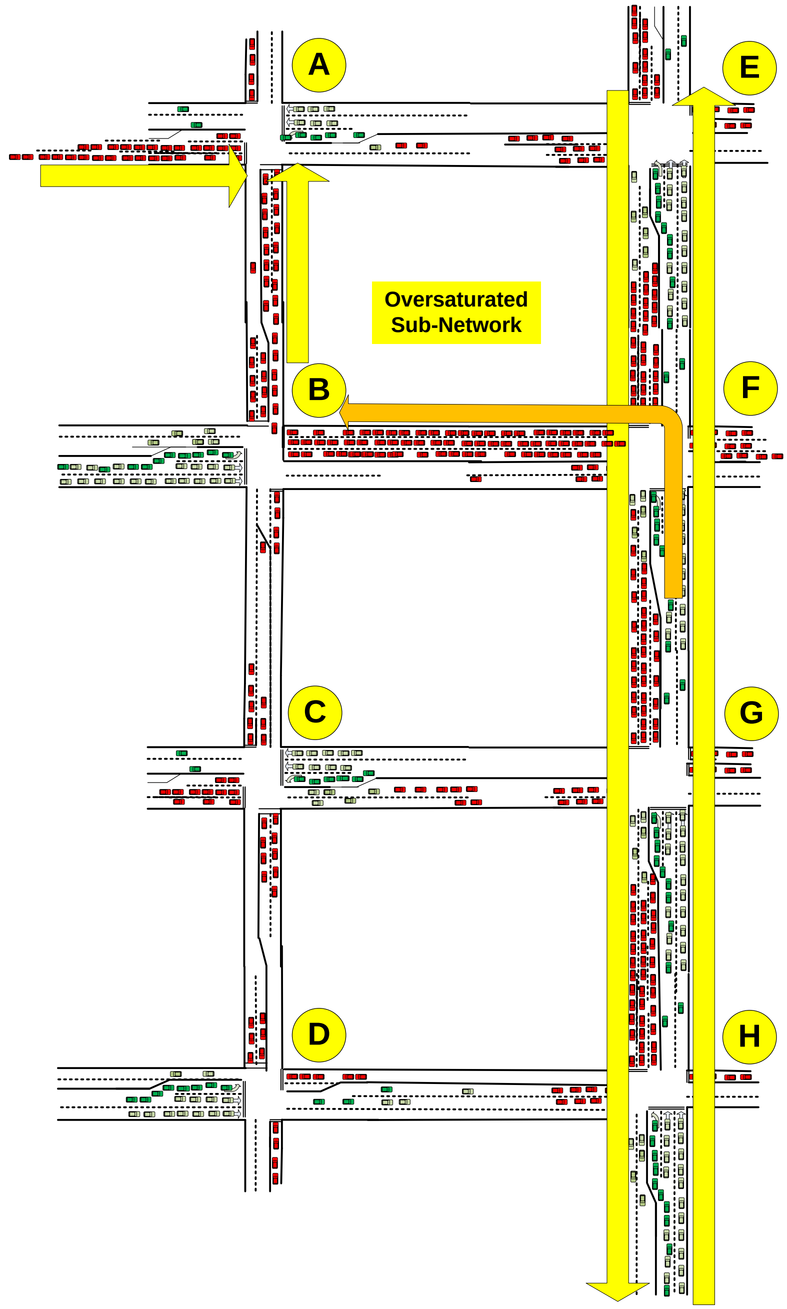


Figure 8. Illustration of an oversaturated network

Special Cases of Network Oversaturation

Several special cases of oversaturated scenarios on networks can be defined. Two examples are freeway-arterial diamond interchanges and arterials with heavy traffic on the arterial and minor flows on side streets. These special cases as listed in Table 2.

Table 2. Special cases of network oversaturation

Special Case	Description
Two-way arterial	Two or more consecutive approaches in both travel directions that are simultaneously oversaturated.
Interchange	Two or more oversaturated routes at the junction of an arterial and a freeway.
Grid	Two or more oversaturated routes in a network of signals that have regular spacing and typically are run together on a common cycle time in central business districts. Coordination in both directions of travel is typically considered.

Figure 9 illustrates the special case of an oversaturated condition on a two-way arterial. In this example, both directions of north and south travel are oversaturated with traffic at the same time along the route from E→F→G→H and from H→G→F→E.

Oversaturated conditions existing on routes and networks are complex problems requiring careful consideration of green-time re-allocation, sequence, offsets, and cycle selection.

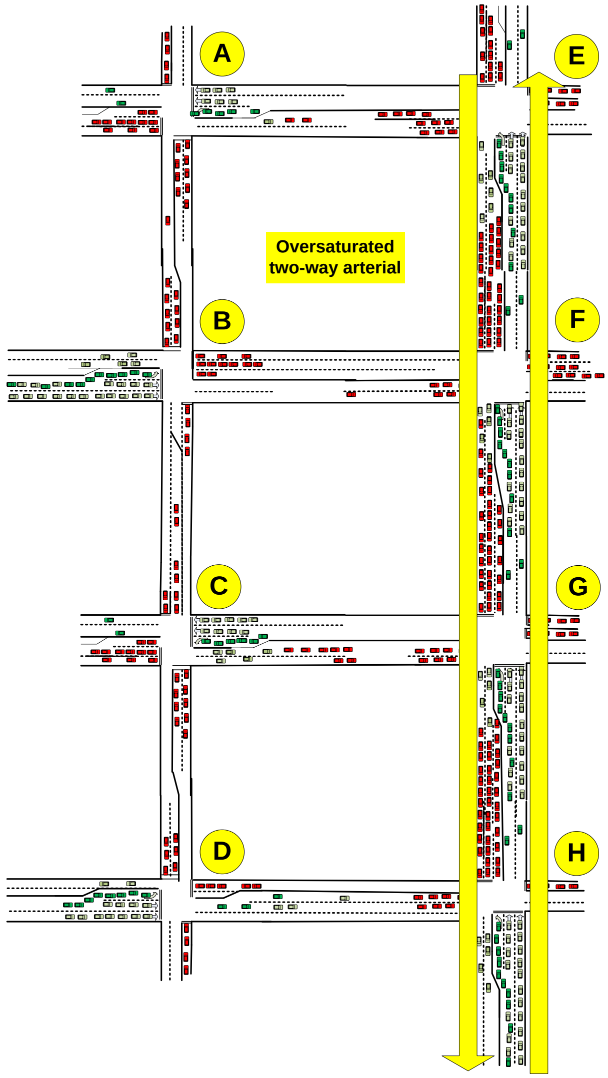


Figure 9. Illustration of oversaturated condition on a two-way arterial

Large-Scale Problems and Gridlock

Wide-spread or regional oversaturated conditions are the most complicated situations to be handled by any mitigation strategy. Situations can arise where a mitigating action in one area of the network exacerbates the congestion in other areas. Multiple interacting areas of oversaturated conditions define a regional oversaturated network as illustrated in Figure 10.

Gridlock is defined as a special case of oversaturated conditions where simultaneous blocking of several movements causes traffic to remain unable to proceed in any direction. During gridlock, the green time is provided to a movement when the vehicles served by that traffic phase are unable to proceed.

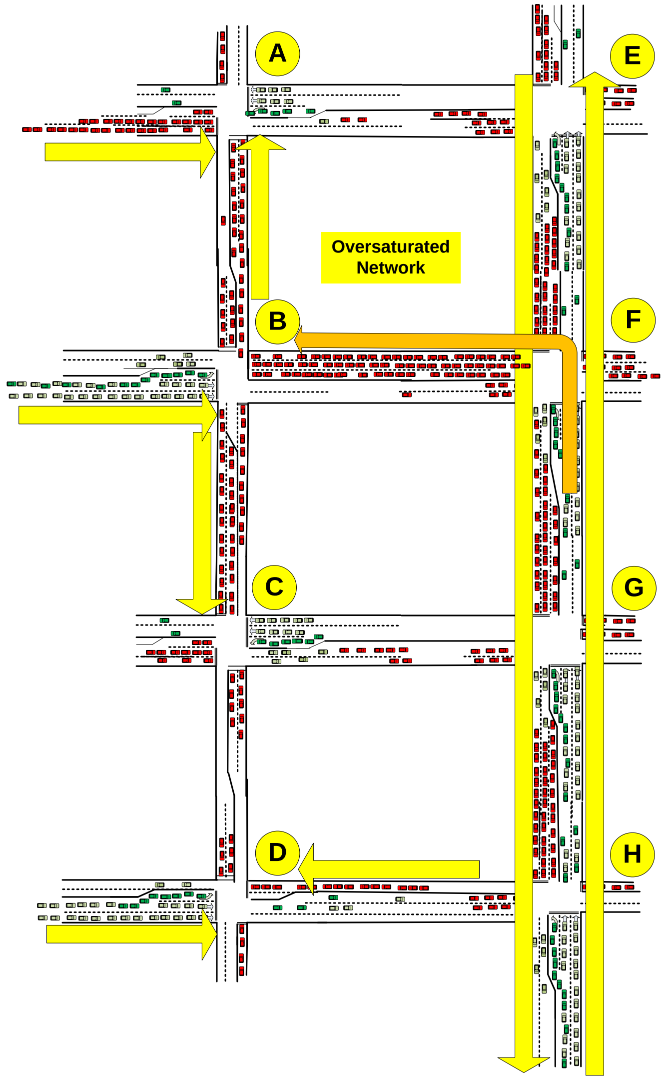


Figure 10. A challenging regional network scenario

Strategic restriction of demand to the network (i.e. "metering") might be expected to be the only reasonable mitigating action that can resolve gridlock situations.

Duration of Oversaturation

The series of definitions in the previous section categorized the ***spatial*** extent of an oversaturated condition. The other dimension defining an oversaturated condition is the ***duration***. Of course these two elements are continually interacting as traffic moves through the system, perhaps creating oversaturation in one area that was previously in a different part of the network. The problem may grow continually larger as flows at one or more critical points create shockwaves that move upstream.

The existence of a persistent (or growing) queue for two or more cycles on a facility (movement, approach, intersection, etc.) defines an oversaturated condition. This is a minimum level of occurrence that provides the definition. As the condition persists for more cycles continuously the condition would be considered to be more severe when combined with the severity level presented by the length of the persistent queue with respect to the storage area for the movement or approach. An oversaturated condition is thus considered to be dissipated when a queue that was persistent from the previous cycle is cleared during the following green time.

However, the nature of traffic is a random process that is influenced by variations in driver behavior and inherent randomness in aggregate arrival rates of traffic due to individual decisions on departure time and route. As such, a condition may be dissipated for one or two cycles only to return in the next cycle due to surge in traffic demand. As shown in Table 2, there are four terms to describe time conditions on the extent of oversaturated conditions:

Table 3. Duration of oversaturation

Duration	Description
Situational	Oversaturated conditions characterized by several consecutive cycles in which the condition persists but is naturally dissipated due to removal of exogenous factors that caused the condition.
Intermittent	Oversaturated conditions characterized by frequent transition between over- and under-saturated conditions.
Persistent	Oversaturated conditions characterized by a considerable number of consecutive cycles in which the oversaturated condition continues. A "considerable number of cycles" might be defined, at a minimum, to be a duration during which it would not be considered typical to modify a signal timing pattern based on a time-of-day pattern schedule. For example, it would not be common to modify signal timing plan parameters more often than once per 30 minutes. At a maximum, the duration of a *persistent* condition might be dissipated within the time where it would be typical to make a pattern change based on a time-of-day schedule. For example, many agencies might be expected to change signal timing parameters in 2 to 3 hours increments.
Prolonged	Extensive duration of oversaturation that extends for time periods that would encompass the duration of more than one pattern in a typical time-of-day schedule. At a minimum, more than one to one and a half hours in duration.

We have tried to stay away from defining "crisp" criteria for definitions of temporal duration such as 15 minutes, 1 hour, or five cycles.

Causal Factors

A wide range of influencing or causal factors can cause oversaturated conditions. As shown in Table 3, the basic categories of causal factors are:

Table 4. Causal factors

Factor	Description
Traffic Demand	Heavier traffic flow than can be processed by the traffic signal system regardless of modifications and enhancements to geometrics, signal timings, or both. Variations in demand level can also lead to intermittent oversaturation.
Geometrics	Physical design characteristics of a traffic facility that exacerbate the ability of the traffic signal system to move traffic efficiently.
Traffic signal operations	Signal timing practices and inefficiencies that contribute to oversaturation due to "sub-optimal" operating policies and principles.
Other travel modes	Service of other travel modes (buses, trains, bikes, pedestrians) by the traffic signal system and modal operations (bus stops, train crossings, etc.) that exacerbate the ability of the traffic signal system to move traffic efficiently.
Anomalous events	Atypical events and conditions including crashes, work zones, weather conditions, and other incidents that exacerbate the ability of the signal system to move traffic efficiently since the saturation flow rates and travel behaviors of drivers are modified significantly.
Planned Special Events	Events that are known to happen at a specific time, such as concerts or athletic events. Ingress and egress to the facility or facilities exacerbates the ability of the traffic system to operate efficiently. Start of oversaturation at ingress is typically difficult to determine, although end time of oversaturation typically occurs shortly after start time of the event. Conversely, start of egress is sometimes less of a certain event (for example, overtime at a sporting event) but the "beginning of the end" might be easier to identify since when the parking lot at the event is empty, the end of the event is known.

A specific condition can, of course, be caused by a combination of these influencing factors. Combinations of factors will increase the intensity of the situation negatively.

Occurrence Frequency

The final component that is necessary for categorizing an oversaturated condition is the frequency in which the oversaturation occurs. Occurrence frequency is divided into two basic categories: recurrent and non-recurrent as described in Table 4. Recurrent situations are easier to study and analyze due to their repeatable nature. Strategies can be applied that are pre-planned and use fixed modifications to signal timings. Non-recurrent conditions can require automated responses based on detector monitoring.

Table 5. Frequency of oversaturation

Frequency	Description
Recurrent	Oversaturated conditions characterized by relatively predictable and repeatable occurrences at certain times of day and days of the week. Geometric physical capacity, peak travel demand rates, traffic signal timing operations, and other modal effects (buses, trains, bicycles, and pedestrians) can be considered as recurrent causes. Situations can include all or any combination of factors.
Non-recurrent	Oversaturated conditions that occur on a traffic facility because of atypical exogenous factors that are not predictable or repeatable. Factors could include crashes and incidents, significant demand pattern shifts, and work zones as well as atypical influence of other modes (buses, trains, bicycles, pedestrians) such as heavy pedestrian crossings due to a special event. Situations can include all or any combination of factors.

Specific Symptoms on Routes and at Intersections

The definitions presented above for spatial and temporal extent represent a high-level view of the extent of queuing in a traffic signal system network. At a more detailed level, from a link-by-link perspective, it is important to quantify the specific type of problem being experienced.

Overflow queues must be considered relative to the amount of storage capacity on the particular movement, phase, or approach. For example, a long persistent queue on a long approach may not be a direct cause for alarm. This situation may actually be the most appropriate mitigation to a particular scenario by storing vehicles where there is the most capacity. However, a relatively short queue that consistently fills a short link might cause a ripple effect. An important step in the process is to consider the extent of overflow queues versus the storage available. In particular, increasing green time at an upstream signal to disperse additional vehicles queued upstream will further exacerbate the downstream situation when there is limited storage available on the downstream link. A straightforward procedure for minimizing the two kinds of overflow queuing problems is presented in Chapter 3.

In addition to overflow queuing, the following symptoms contribute additional complexity to the design and application of mitigation strategies:

- Spillback
- Starvation
- Storage blocking
- Cross blocking

The combination of these effects in oversaturated routes, sub-networks, and networks is what makes the management of oversaturated conditions one of the most challenging problems in traffic signal control.

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 11.

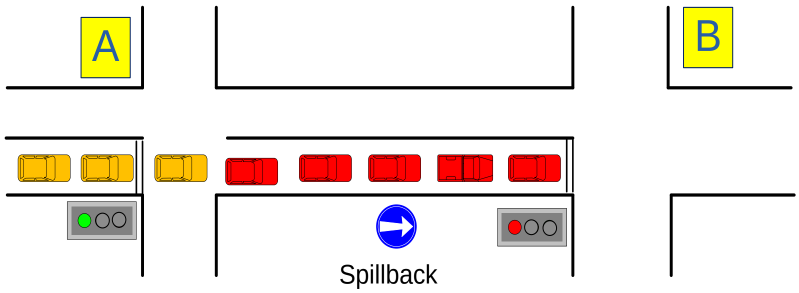


Figure 11. Approach spillback (de facto red)

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. Starvation due to sub-optimal signal timing is illustrated in Figure 12.

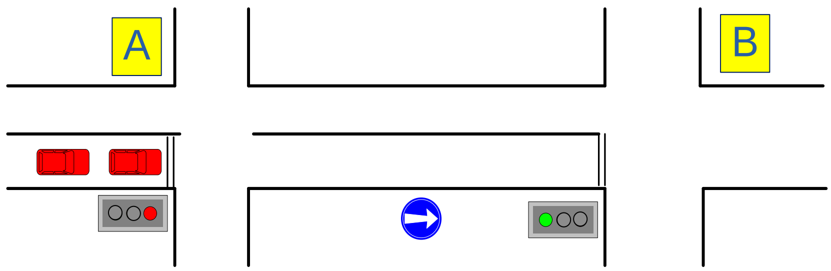


Figure 12. Approach starvation due to signal timing

Storage bay spillback, shown in Figure 13, occurs when turning traffic uses up the entire space of the storage lane and blocks the through traffic. The blocked through movement then experiences starvation.

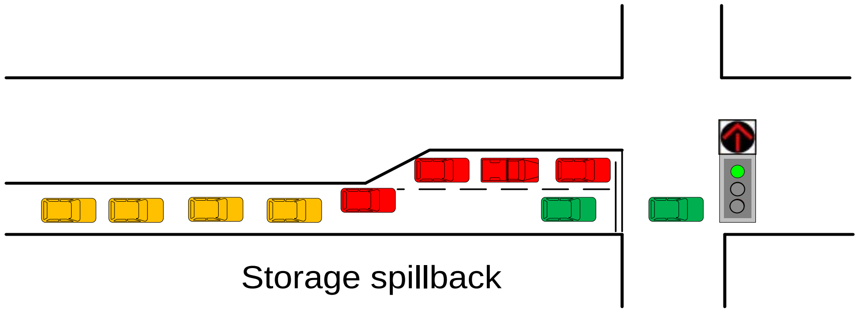


Figure 13. Storage bay spillback

Turning storage blocking, shown in Figure 14, occurs when queues extend beyond the opening of the storage bay. 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. If there are no vehicles in the left turn bay, the left turn can also be skipped completely.

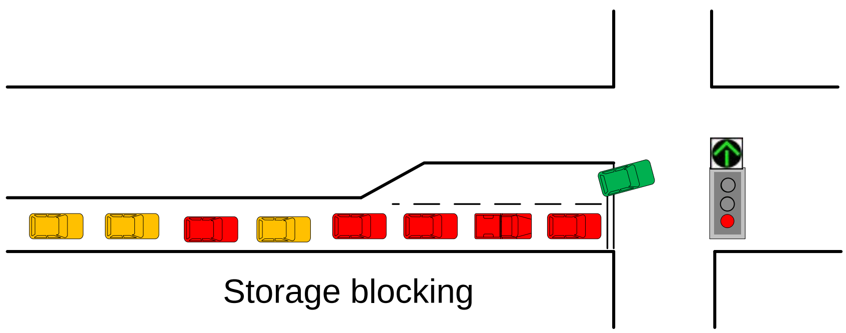


Figure 14. Storage bay blocking

Cross intersection blocking, illustrated in Figure 15, occurs when queues extend into an intersection blocking the progression of crossing vehicles. While most jurisdictions have "don't block the box" laws or policies, these types of situations are not uncommon 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.

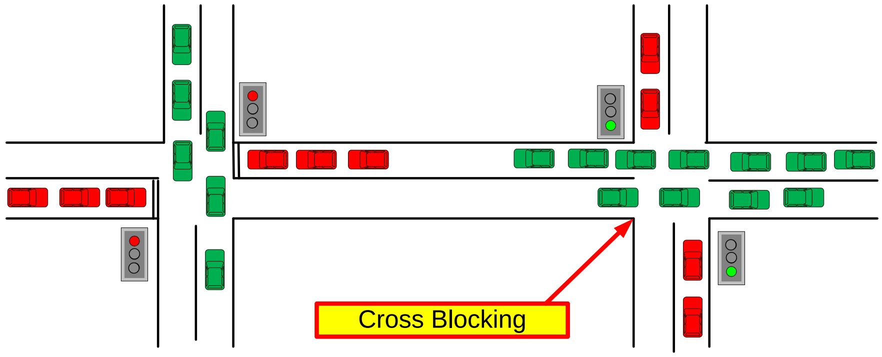


Figure 15. Cross blocking effects

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

Summary of Characteristics that Define an Oversaturated Scenario

A particular oversaturated scenario is defined as a combination of the attributes summarized in Table 6. The purpose of this categorization matrix is to identify what type of problem is occurring in order to identify the appropriate mitigation strategies that are applicable. In the next section, each mitigation strategy will be categorized according to which elements in Table 6 are applicable to that strategy.

Table 6. Summary of characteristics of oversaturated scenario

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		Unplanned Events		
Two-way arterial		Planned Events		
Interchange				
Grid				
Network				

Oversaturation Problem Characterization and System Dynamics

Oversaturated conditions might be characterized as being both easy and hard to identify. A motorist that takes a specific route on a daily basis might easily predict where the oversaturated links will occur on their route and knows almost instinctively when the conditions on certain parts of their route are more heavily congested than normal. Similarly, with extended experience in a particular agency and location, traffic engineers become accustomed to the trouble areas of their jurisdiction and this is not only contained to situations that are recurrent. Special event patterns and intermittent situations (such as those created by bus or train schedules) can certainly be identified.

The first step to characterization is observing and identifying which type of scenario is being experienced. In many situations, with good local knowledge or limited problem extent, identifying the elements in each column of the table is straight forward. In other more complex situations it will be important to collect field data and analyze how the data helps to identify the appropriate element in each column of Table 6.

As part of the scenario definition you will need to define what is "in" the system and what is not "in" the system. This is a subjective decision. A general rule of thumb might be to include intersections that are affected by the oversaturation at some point during the scenario, but no more. Certain mitigation strategies such as gating may cause approaches becoming oversaturated that were not initially oversaturated for the explicit purpose of alleviating downstream conditions.

High-Level System Dynamics

From a high-level perspective, it is well known that daily traffic and recurrent events have repeatable patterns. No traffic system is continually in oversaturated operation and thus any scenario evolves into three regimes of operation:

- Loading
- Oversaturated operation (or "processing")
- Recovery

This concept is illustrated in Figure 16.

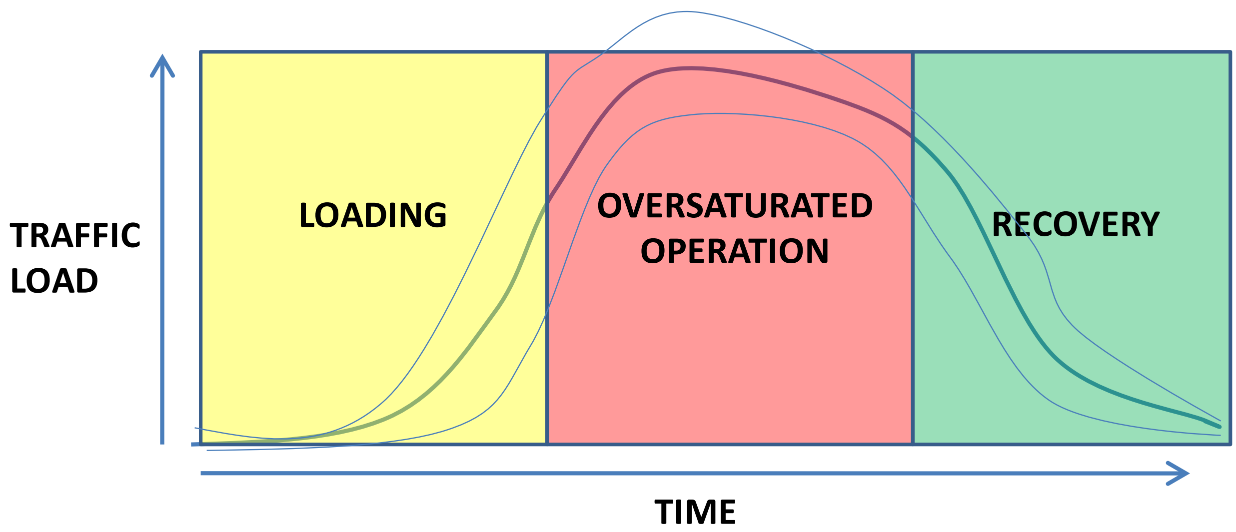


Figure 16. Loading, oversaturation, and recovery regimes of operation

During the ***loading*** regime, the traffic volumes are increasing, route proportions are changing, and in the case of non-recurrent events, the triggering event(s) have started. During loading, overflow queuing and other symptoms such as storage blocking and starvation begin to emerge. Early application of mitigation strategies can *delay the onset* of oversaturated operation. During the loading phase, shifting one's operational objectives from minimizing user delay to maximizing throughput can provide a measurable improvement in performance on the approaches, routes, and networks that will shortly become oversaturated. This principle is confirmed in many of the test cases presented in Chapter 3.

Early application of mitigation strategies is easier to conceptualize when the causal factors are recurrent. If the condition is non-recurrent or if it is difficult to predict when the condition will starts or how long it will last, then it may necessary to use an online tool to identify and then apply a mitigation strategy.

During the ***oversaturated operation*** regime, the traffic volumes and route proportions are such that queues and congestion are not going to be dissipated until either (a) the traffic volumes are reduced, (b) the route proportions are changed (i.e. drivers' avoid the area, adjust their routes, decide to travel later, etc.) or (c) both. This is the operational situation that many practitioners might characterize as "there is nothing that can be done". While we disagree that this is the case for all situations, it is true that it is difficult to discern the difference between different mitigation strategies when the overflow queuing and downstream blockages hinder the ability of traffic to be moved (anywhere). Applying queue management approaches (e.g. decreasing green time or truncating phases when a downstream link is blocked) can provide enhanced service to non-saturated movements and approaches that can increase total system throughput. Mitigation strategies applied during this phase also serve to help the system return to steady-state operation sooner during the recovery phase than continuing to apply the "normal" operational strategies.

During the ***recovery*** regime, traffic volumes, route proportions, or restrictive downstream capacity (e.g. clearance of crash, opening of additional toll booths, removal of construction, reduction in traffic flow) have been adjusted so that the overflow queues begin to dissipate. In this phase of operation, mitigation strategies are especially effective in returning the system to a steady state sooner than continuing to apply the "normal" operational strategies.

Summary

An oversaturated traffic scenario is described by its attributes:

- Extent
- Duration
- Causation
- Recurrence
- Symptoms

The duration and extent attributes of a scenario are further characterized by three regimes of operation:

- Loading (growth of queues)
- Processing (persistence of queues)
- Recovery (dissipation of queues)

In order to identify appropriate mitigation strategies and signal timing plan parameters, the scenario must first be adequately characterized. Direct measurement of queue lengths is the primary method for doing this. Historically, this has been challenging to accomplish in the field due to the need for deployment of extensive sensor networks. In the next section, we describe a novel technique for estimation of queue lengths from typical "advance" detector loops or zones when combined with high-resolution (second-by-second) signal timing phase data. From this basic technique, two derived measures of the type of oversaturation are developed which can directly measure problematic symptoms. Differentiation of the two common symptoms in oversaturation is a key step in determining which mitigation strategies and signal timing approaches are appropriate for a given scenario.

Measuring Length of Queue and Overflow Queuing Effects

In the previous section, we defined the key qualitative characteristics of an oversaturated traffic scenario. To characterize the nature of an oversaturated problem quantitatively, the primary method is to measure queue lengths. Historically, this has been challenging to accomplish in the field due to the need for deployment of extensive sensor networks. This project, we have developed one direct and two derived measurements of oversaturation at the movement or approach level. The direct measurement of oversaturation is the length of the overflow queue relative to the length of the approach. For this project, we have shown that length of overflow queue can be reasonably estimated using second-by-second detector volume and occupancy data from upstream detectors and second-by-second phase timing information. This method uses a fairly simple traffic flow model correlated to the time when the phase that serves the queue is green or red to determine if a queue is growing or shrinking.

The key characteristic of this model is that queue lengths much further upstream from the detection point can be reasonably estimated. This allows the approach to be applied in typical arterials with dilemma zone or extension detection. Typical stop-bar detectors that are 25'+ in length do not do an adequate job of capturing the gaps between traffic to estimate these measures.

Furthermore, this method can capture the overflow queue, or the amount of queue remaining after the light turns red, that was not dispersed during the green time. This is the key distinction between oversaturated operation and operation of the signal during congested, but still undersaturated conditions. If a long queue is generated during the red light but is dissipated fully during the ensuing green phase, this may indicated heavy approach demand, but it is not oversaturated.

Quantitative Characterization of the Severity of Oversaturation

The amount and cause of overflow queuing is the key distinction in determining which mitigation strategies are appropriate for a given traffic scenario. If the overflow queue is created because of not enough green time, then essentially more green time is required. However if the overflow queue is created because of a downstream restriction, then more green time for this phase may only worsen the problem. Two quantitative indicators characterize overflow queuing:

- TOSI – temporal oversaturation severity dimension
- SOSI – spatial oversaturation severity dimension

Detrimental effects in the *temporal* dimension are characterized by an overflow queue at the end of the signal cycle. These overflow vehicles that did not pass through the intersection during the current cycle, cannot be discharged due to insufficient green time. This overflow queue must now be served in the next cycle.

The amount of green time that is now used to service the overflow queue is quantified by TOSI, ranging from 0% to 100%. When TOSI = 100%, all of the green time for the phase is used to disperse the overflow queue. If the arrival rate is still constant (or increasing), the queue will continue to grow so that TOSI could conceptually be 120%, 140%, 200%, and so on. The values of TOSI indicate directly how much additional green is needed to disperse the overflow queue.

Detrimental effects in the *spatial* dimension are characterized by the inability of upstream traffic to proceed due to blockage at the downstream intersection. In this case, vehicles are not discharged from the upstream intersection even though the signal is green. Therefore, some portion of the green time of the upstream intersection becomes unusable. The amount of green time that is unusable is quantified by SOSI, ranging from 0% to 100%. When SOSI = 100%, all of the green time for the phase is wasted because the downstream vehicles cannot move.

SOSI = 100% is a distinctly different situation than TOSI = 100%. When TOSI = 100% and there is downstream capacity to receive vehicles, this phase can benefit greatly from increasing the green time. When SOSI is 100%, there is no need to allocate more green time to the phase, unless the downstream blockage is dissipated. The values of SOSI essentially indicate how much additional green time is needed at the downstream intersection to disperse traffic so that upstream traffic can move. SOSI is also affected by the offset relationship between the upstream and downstream intersections. TOSI is related to how the upstream intersection and the downstream intersection are related. Poor offsets can create TOSI effects if the upstream platoon is released too early. Therefore, in most situations it is important to consider more than one intersection at a time when considering mitigation strategies.

In the next section we provide a brief summary of the motivation for the queue estimation technique followed by the theory and examples of TOSI and SOSI computation measures.

Motivation for the Measurement of Queue Length and Oversaturation Severity

An oversaturated traffic facility, is generally defined as demand exceeds the capacity. The degree of saturation, *i.e.* the volume/capacity ratio, is defined as:

$$X_i = \frac{v_i}{c_i} \qquad \text{Equation (Eq.) 1}$$

where X_i is the degree of saturation for lane group i; v_i and c_i are demand flow rate and capacity for lane group i, respectively. A lane group or approach is oversaturated when $X_i > 1$.

For a single intersection with two competing Eq. 2 streams, Gazis (1964) expanded this oversaturation concept by proposing the following inequality:

$$q_a/s_a + q_b/s_b > 1 - (L/C) \qquad \text{Eq. 2}$$

where q_a and q_b are arrival rates for two conflicting directions; s_a and s_b are saturation flow rates for two directions; L is the total lost time; and C is the cycle length.

Direct application of these definitions to detect the onset and quantify the duration and extent of oversaturation 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 for oversaturated situations, the detectors typically cannot distinguish the true arrival rate because a queue grows past their fixed-point locations. Traffic demand is simply not measurable when a fixed-location detector is occupied with vehicular queue.

Alternatively, oversaturation has also been characterized as "a stopped queue cannot be completely dissipated during a green cycle" (Gazis, 1964), or "traffic queues persist from cycle to cycle either due to insufficient green splits or because of blockage" (Abu-Lebdeh & Benekohal, 2003). These definitions were discussed in the previous section of this report. The key concept in the definitions is not only the estimation of the queue length, but also the portion of the queue that is residual or overflow from the previous cycle.

Many other alternative measures have been defined (refer to Appendix A of this report for a comprehensive literature review) for degree of saturation, but they do not provide enough information on the oversaturated condition for accurate decision making. Methods such as green utilization can indicate the need for more green, but not quantify how much.

To the best of our knowledge, previous research studies using traffic data from signal systems to diagnose and identify oversaturation are mostly qualitative and incomplete. Conceptual definitions discussed in the literature review are either not applicable in the real world or have other deficiencies. Since detection of the onset of oversaturation as well as quantifying the severity of oversaturation is a critical step before applying appropriate mitigation strategies, it becomes imperative to have an implementable and quantifiable measure of oversaturation and a coherent methodology to identify the situation. This component of the research project is intended to fill that gap.

This section is organized as follows. First we discuss overflow queue length as a quantifiable measure of oversaturation. Overflow queue is measurable from typical vehicle-actuated traffic signal systems with advanced detection on the subject approach and access to second-by-second detector and phase timing data. Next, the methodology for identification and quantification of oversaturation is then described. Finally, the section concludes with the results from field testing of the methodology in the real world and a summary of the implications for the other components of the research project. Conclusions and directions for future work are summarized in Chapter 4.

A Quantifiable Measure of Oversaturation

Since the general definition of oversaturation, *i.e.* traffic demand exceeding the capacity of a facility, cannot be applied directly to detect the occurrence of oversaturation, we propose a measure of oversaturation by quantifying its *detrimental effects*. The detrimental effect of oversaturation can be described in temporal and/or spatial dimensions, both of which lead to the reduction of usable green time in a cycle for a signalized approach.

Detrimental effects of oversaturation in the temporal dimension are characterized by an overflow queue at the end of a cycle. These overflow vehicles were intended to pass through the intersection during the current cycle. Because of insufficient green splits, these overflow vehicles were not able to be discharged and must be serviced in the next cycle. The portion of next green time phase that is used to discharge the overflow queue becomes “unusable” for the traffic arrivals during that cycle. If the arrival rate remains the same and the green time remains inadequate, eventually all of the green time is spent serving the overflow queue and the queue continues to grow longer.

Detrimental effects of oversaturation in the spatial dimension can be characterized by a spillover from downstream traffic. When spillover happens on an approach, the downstream link is blocked and vehicles cannot be discharged from the intersection during the green phase. Therefore a portion of the green time becomes unusable, because while the light is green the vehicles at the stop-bar cannot proceed. The most common cause of spillover is that the downstream link is fully occupied by a queue when the light turns green.

Therefore the condition of oversaturation is characterized by an overflow queue at the end of a cycle creating detrimental effects on the following cycle, or by a downstream spillover within a cycle creating detrimental effects on the upstream approach. To quantify the detrimental effects in either the temporal or spatial dimensions, we introduce the ***oversaturation severity index (OSI)*** by using the ratio between unusable green time and total available green time in a cycle. OSI will be a non-negative percentage value between 0 and 100, with 0 indicating no detrimental effect for signal operation, and 100 indicating that all available green time becomes unusable.

We further differentiate OSI into ***TOSI*** and ***SOSI***. TOSI describes the detrimental effects created by overflow queue, *i.e.* the detrimental effect in the temporal dimension. SOSI describes the detrimental effects caused by spillover, *i.e.* the detrimental effect in the spatial dimension. Although both TOSI and SOSI can be calculated using the ratio between unusable green time and total available green time, the meanings of “unusable” are distinctly different. For TOSI, the “unusable” green time is the equivalent green time to discharge the overflow queue in the following cycle while for SOSI, the “unusable” green time is the time period during which a downstream link is blocked and the upstream discharge rate is zero.

Since TOSI quantifies the detrimental effect of oversaturation on the following cycle, the duration and frequency of TOSI greater than zero becomes a fundamental indicator of traffic congestion at the intersection level. On the other hand, SOSI describes the detrimental effect of oversaturation caused by a downstream queue spillover, indicating a route-level problem. From a practical perspective, the presence of TOSI >0 and SOSI>0 in a series of compatible approaches is the primary method by which to judge a critical route in an oversaturated system.

The differentiation of TOSI and SOSI identifies the causal relationship of arterial traffic congestion. Positive TOSI indicates that the available green time is insufficient for queue discharge and an overflow queue is generated at the end of a cycle. Subsequently, in the following cycles, the queue may grow and spillover to an upstream intersection creating a positive SOSI at the upstream intersection. Clearly in this case a positive S-OSI at the upstream intersection is caused by the downstream bottleneck. Note that a downstream bottleneck may lead to the situation that both SOSI and TOSI of the upstream intersection are greater than zero because a portion of the green time is wasted due to downstream blockage (*i.e.* SOSI> 0) and an overflow queue may be generated (*i.e.* TOSI> 0) due to the inability to fully discharge the queue.

With the TOSI and SOSI oversaturation severity indices, the focus of the identification algorithm shifts from *measuring travel demand* to *quantifying detrimental effects*. The classification and quantification of detrimental effects created by oversaturation are very important for traffic management, because different oversaturated situations call for different mitigation strategies For example, an isolated intersection with positive TOSI values on one or more approaches can be mitigated by extending green times. An arterial corridor with multiple intersections having positive TOSI and SOSI values requires the adjustment of green times as well as offsets to prevent further deterioration of oversaturation. A heuristic methodology for directly using TOSI and SOSI values to adjust green times on an oversaturated route is presented later in this Chapter.

In the following section, we describe two algorithms for the identification and quantification of oversaturated conditions. The first algorithm estimates the overflow queue length (and correspondingly, TOSI), and the second algorithm detects spillover conditions and estimates SOSI.

Algorithms for Identification of Oversaturation

The identification algorithms discussed in the following section will work with typical detector (6’x6’) configurations for a vehicle-actuated signalized intersection, *i.e.*, with either short stop-line detectors for vehicle presence detection or advance detectors that are a few hundred feet upstream from the stop line for green extension. Throughout this section, we will assume advance detectors are available and will note the necessary changes if the only available detector is located at the stop-bar. We also assume that high-resolution (*i.e.* second-by-second or event-based) traffic signal interval (i.e. green, yellow, red) data can be collected.

The availability of high-resolution traffic signal data has been increased in recent years. For example, second-by-second detector data has been used by ACS-Lite (Luyanda et al., 2003). Continuous event-based signal data, including both vehicle-detector actuation events and signal phase change events, has been collected and archived by the SMART-Signal system (Systematic

Monitoring of Arterial Road Traffic and Signals) developed at the University of Minnesota (Liu & Ma, 2009; Liu et al., 2009a, 2009b). Phase timing and detector data is also stored on the ASC3 controller and can be retrieved via FTP (Smaglik, et al, 2007). Although the algorithms presented in this section are demonstrated by using event-based data from the SMART-Signal system, they are also applicable to second-by-second signal and detector data coming from any other traffic signal management system.

Algorithm for Overflow Queue Length Estimation

An overflow queue at an intersection refers to those vehicles that are part of the discharging platoon that cannot pass through the intersection during the green time. An overflow queue also represents the minimum queue length at the end of a cycle as additional vehicles may join the queue from side street turning movements. Vehicles in the overflow queue then occupy a portion of green time in the next cycle. The ratio between the overflow queue discharge time and the total available green time is then denoted as TOSI.

Estimation of overflow queue length requires the reconstruction of queue length profiles within a cycle. As the traditional input-output approach for queue length estimation can only handle queues that are shorter than the distance between the vehicle detector and the intersection stop line, in this section we adopt the queue length estimation method developed by Liu et al. (2009a). Both maximum and minimum queue lengths can be calculated using this method. In the following discussion we provide a brief discussion of the queue length estimation algorithm, focusing on the application of this method to the estimation of both oversaturated conditions measured by TOSI and SOSI.

The queue estimation method is based on the identification of traffic state changes and the associated shockwaves that occur during each cycle. The information in Figure 17. Shockwave profile within a cycle

, assumes there is no overflow queue from past cycles, so at the beginning of the red signal, the vehicles that arrive at the intersection are forced to stop which creates a queuing shockwave (v_1) spreading backward from the stop line. At the beginning of the effective green time, vehicles begin to discharge at the saturation flow rate (assuming there is no blockage downstream) forming the discharge shockwave (v_2), again spreading upstream from the stop-line. The discharge shockwave (v_2) usually has higher speed than (v_1), so the two waves will meet some time after the start of the green time, which is the time that the maximum queue length is reached. As soon as the two shockwaves meet, a departure shockwave (v_3 in Figure 17. Shockwave profile within a cycle

) is generated, spreading toward the stop line. Here the front of the departure shockwave reflects the discontinuity between the saturated discharging traffic flow and the new traffic arrivals after the maximum queue length is reached. An overflow queue is formed sometime after the start of the red signal of the next cycle when the departure shockwave (v_3) meets the compression shockwave (v_4). The compression shockwave (v_4) is similar to the queuing shockwave (v_1), as both shockwaves form a stationary queue. The difference between the two shockwaves is that the compression shockwave represents traffic discontinuity from saturated traffic flow to the "jammed" traffic condition, while the queuing shockwave represents a change to the jammed traffic condition from arriving traffic, which is not necessary in a saturated state. The shockwave motions described above will repeat from cycle to cycle.

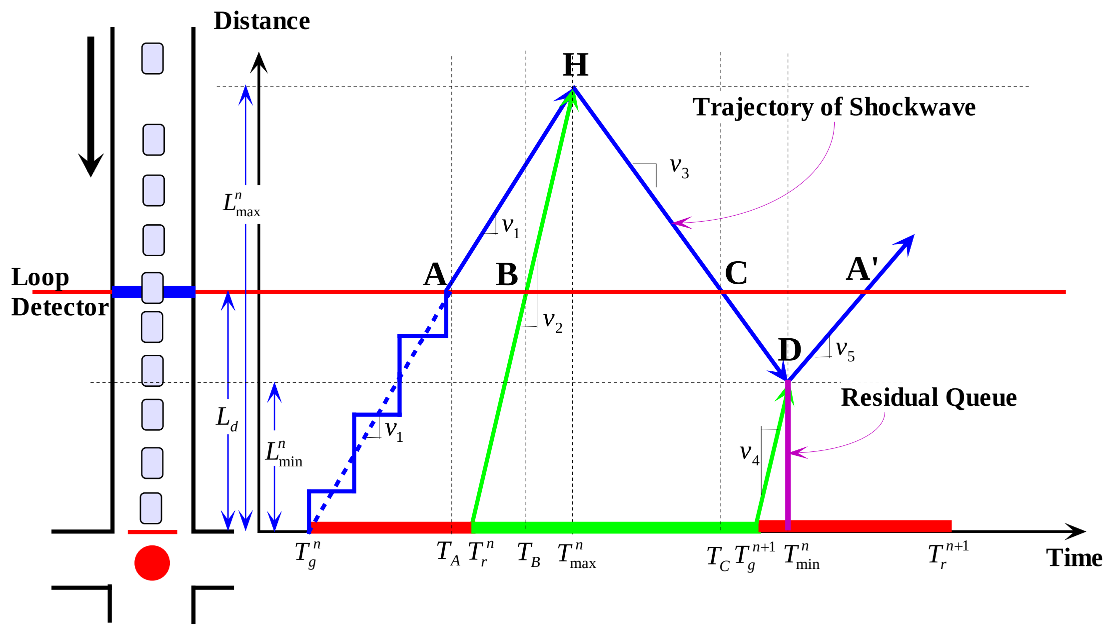


Figure 17. Shockwave profile within a cycle

Intuitively, the queuing profile including the maximum queue length (L^n_{max}, *i.e.*, the maximum queue length for the cycle *n*) and the minimum queue length (L^n_{min}, *i.e.*, the overflow queue length for the cycle *n*) can be easily derived if the times when the shockwaves cross the detector location can be identified (indicated as "break points" *A*, *B*, and *C* in Figure 17. Shockwave profile within a cycle). High-resolution detector data, which provides at least second-by-second detector occupancy time and gaps between consecutive vehicles, implies the changes of traffic state and can therefore be utilized to identify these break points. Figure 18 presents an example of detector occupancy time (see Figure 18a) and vehicle gaps (see Figure 18b) within a cycle. This sample data was collected from an advance detector at one intersection at the SMART-Signal test-bed on TH55 in Minneapolis, MN. As shown in the figure, a sudden increase of occupancy time indicates that queue (v_1) spills back to the advance detector. The time of break point *A* (T_A) can then be identified as the time when the occupancy is significantly increased. Similarly, the time of break point *B* (T_B) can be identified as when the occupancy drops to the normal value. This change indicates that the discharge shockwave (v_2) spreads back to the advanced detector and the vehicles begin to move.

Break point C (T_C)represents the time when the departure shockwave (v_3) crosses the detector line. Before break point C appears, vehicles discharge at the saturation flow rate, *i.e.*, the saturation traffic state (q_m, k_m); and after (T_C), the traffic condition changes back to the arrival traffic state (q_a, k_a). The change of these two traffic states (from saturation to free-flow arrival) is indicated by the variation of the time gaps. As indicated in Figure 18b, before T_C, the time gaps between vehicles are consistently small (less than 2.5 seconds), meaning that most of the vehicles are discharged at the saturation flow rate. But after T_C, the vehicle gaps become much larger and their variances are significantly increased. Therefore a threshold value of a time gap can be used to identify the break point C time T_C.

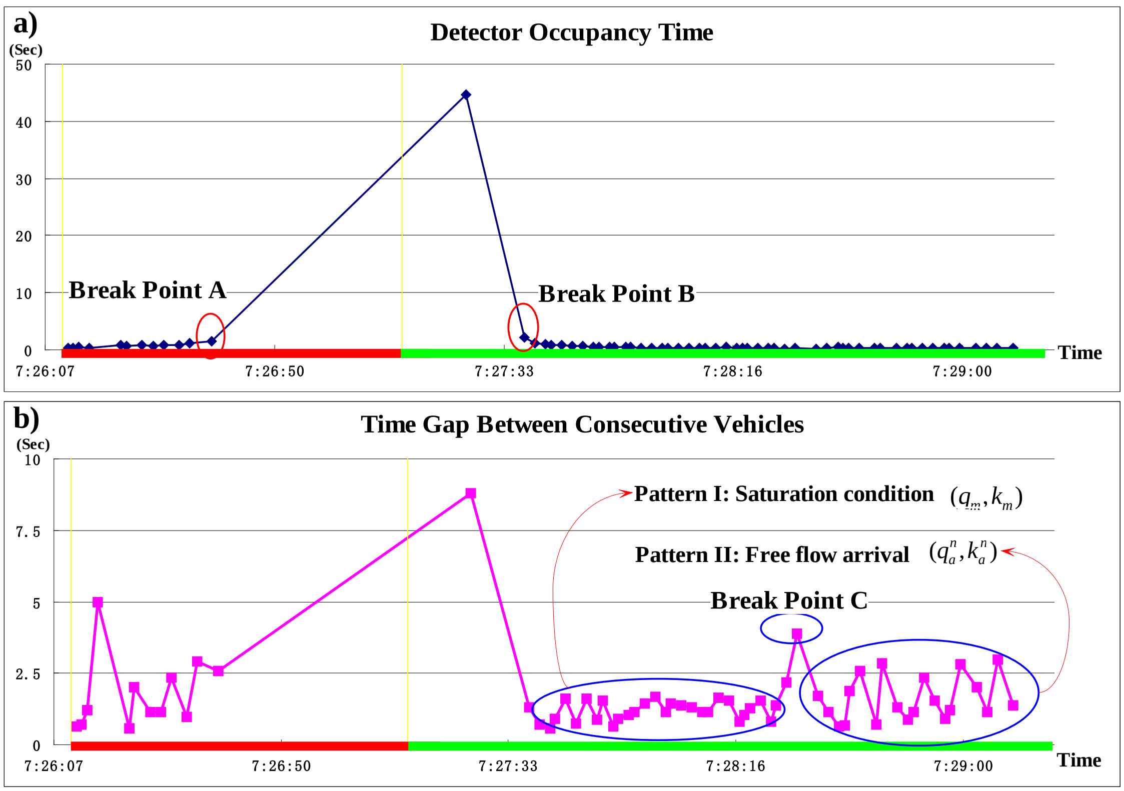


Figure 18. a) detector occupancy profile in a cycle; b) time gap between consecutive vehicles in a cycle.

Once break points (A, B, and C) have been identified, the flow and density of each traffic state (*i.e.* the arrival traffic state (q_a, k_a), saturation traffic state (q_m, k_m), and jammed traffic state (0, k_j)) can be calculated based on detector occupancy times and time gaps between vehicles. Then the wave speeds of v_1, v_2, and v_3 can be estimated using following equation:

$$v = \frac{\Delta q}{\Delta k} = \frac{q_2 - q_1}{k_2 - k_1} \qquad \text{Eq. 3}$$

where q_1, k_1 are the flow and density of the upstream traffic and q_2, k_2 are the flow and density of the downstream traffic.

Using the estimated shockwave speeds from the above equation, the maximum queue (both length $L_{\max}^{n}$ and time $T_{\max}^{n}$) and the minimum queue (both length $L_{\min}^{n}$ and time $T_{\min}^{n}$) during the n^{th} cycle can be calculated based on the shockwave profile (see Figure 17) by the following equations:

$$\begin{cases} L_{\max}^{n} = L_d + (T_C - T_B) \Big/ \left(\dfrac{1}{v_2} + \dfrac{1}{v_3}\right) \\ T_{\max}^{n} = T_B + (L_{\max}^{n} - L_d) \Big/ v_2 \end{cases} \qquad \text{Eq. 4}$$

$$\begin{cases} L_{\min}^{n} = \left(\dfrac{L_{\max}^{n}}{v_3} + T_{\max}^{n} - T_g^{n+1}\right) \Big/ \left(\dfrac{1}{v_3} + \dfrac{1}{v_4}\right) \\ T_{\min}^{n} = T_g^{n+1} + L_{\min}^{n} \Big/ v_4 \end{cases} \qquad \text{Eq. 5}$$

where L_d is the distance from the stop line to the loop detector; and T_g^{n+1} is the end of green of the $(n+1)^{th}$ cycle.

It is necessary to identify whether an overflow queue is present at the end of the cycle before the calculation of the minimum queue length:

$$\begin{cases} \dfrac{L_{\max}^{n}}{v_3} + T_{\max}^{n} < T_g^{n+1} & \text{Without Residual Queue} \\ \dfrac{L_{\max}^{n}}{v_3} + T_{\max}^{n} \geq T_g^{n+1} & \text{With Residual Queue} \end{cases} \qquad \text{Eq. 6}$$

For severely congested traffic conditions, the break point C may not be able to be found during the green phase. In addition, large trucks may make it difficult to identify break point C. In such cases, the traffic pattern does not change during the green phase and vehicles keep discharging at the saturation flow rate (see Figure 19). Then an overflow queue must exist, at least, between the detector location and the stop line. Eq. 4 cannot be applied to calculate the maximum queue length since the shockwave speed (v_3) cannot be calculated.

Under such conditions, the complete queue profile cannot be recovered from the detector data. However, since the entire green time has been used for queue discharge the number of vehicles passing the detector location during the green time can be counted (between T_B and T_g^{n+1}), so that a minimum of the maximum queue length, *i.e.* $\min(L_{\max}^n)$, can be estimated by simply taking the end of cycle (T_g^{n+1}) as T_C (see Figure 20). Since the space headway (l_{jam}) and the velocity of the discharge wave (v_2) can be assumed constant at jammed traffic conditions, the following equation can be used to identify the minimum overflow queue:

If Point C cannot be identified:
$$\begin{cases} \min(L_{\max}^n) = l_{jam} \cdot N + L_d \\ \min(T_{\max}^n) = T_r^n + \dfrac{\min\left(L_{\max}^n\right)}{v_2} \end{cases} \qquad \text{Eq. 7}$$

where N is the traffic count between T_B and T_g^{n+1}; l_{jam} is the space headway at jammed traffic conditions (assumed as a known constant); and T_r^n is the end of the red phase of the n^{th} cycle.

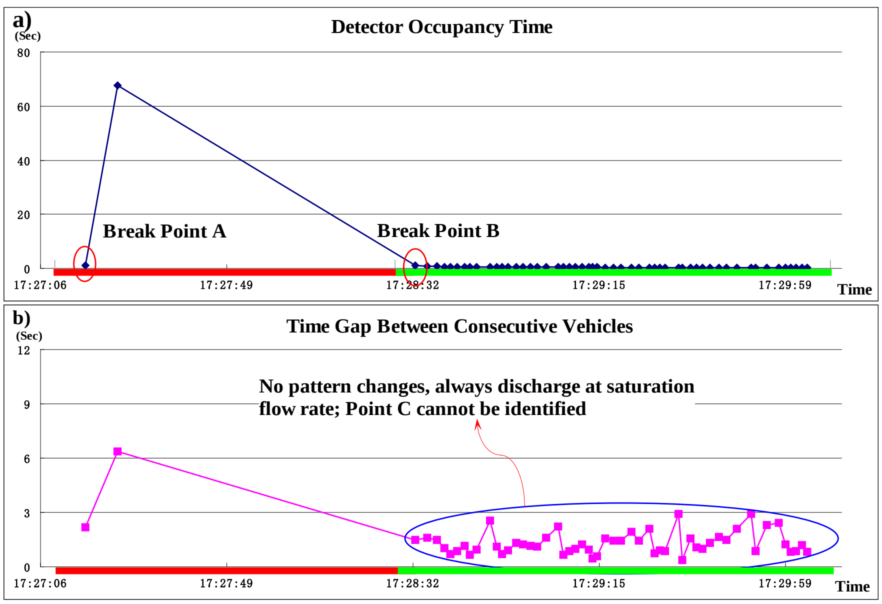


Figure 19. Break points identification (Point C cannot be identified)

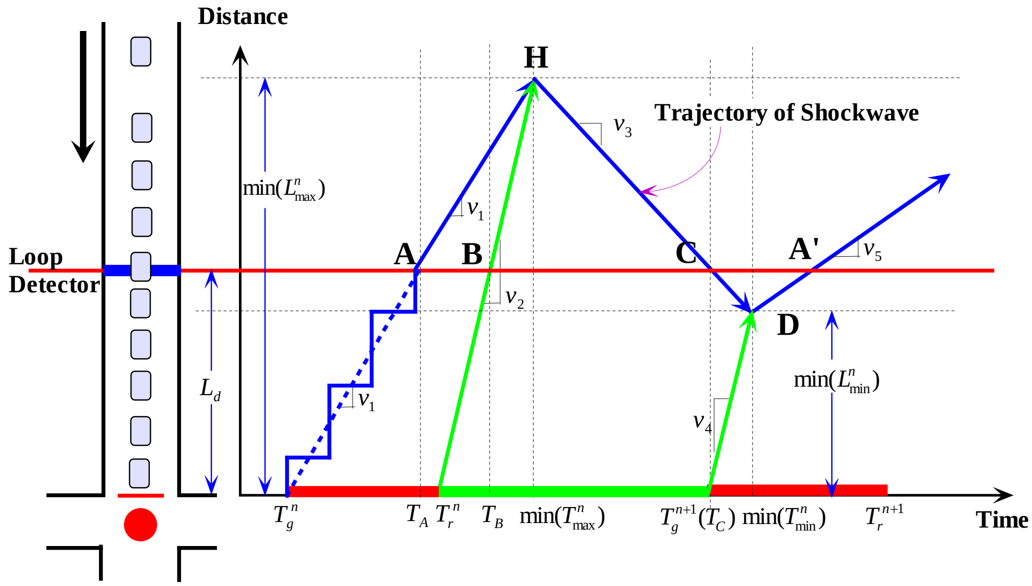


Figure 20. Calculation of overflow queue length when Point C cannot be found

Then v_3 can be calculated by Eq. 8:

If point C cannot be identified: $$v_3 = \frac{\min\left(L^n_{\max}\right) - L_d}{T^{n+1}_g - \min\left(T^n_{\max}\right)} \qquad \text{Eq. 8}$$

The coordinate of the minimum of the overflow queue length, *i.e.* $\min(L^n_{\min})$ and $\min(T^n_{\min})$ for Point *D*, can then be estimated using Eq. 5.

If an overflow queue exists at the end of a signal cycle, some portion of the green time in the following cycle will be utilized to discharge the overflow vehicles and therefore becomes the unusable green time for that cycle. The unusable green time can be determined by calculating the number of vehicles in the overflow queue multiplied by the saturation discharge headway (~2 seconds). The detrimental effect caused by the overflow queue therefore can be quantified by the oversaturation severity index in the temporal dimension (T-OSI):

$$\text{T-OSI} = \frac{\textit{unusable green time}}{\textit{total available green time}} \times 100\% = \frac{L^n_{\min} / l_{jam} \cdot h}{G} \times 100\% \qquad \text{Eq. 9}$$

where *G* is the effective green time, and *h* is the saturation discharge headway.

We should note that, when only stop-line detection is available, an overflow queue can be detected if break point C cannot be identified within the green time. Although the length of the overflow queue cannot be measured with stop-line detection only, it is sufficient to say that oversaturation may have occurred at this intersection for this cycle, *i.e.* TOSI > 0. This assumes the stop-line detector is sufficiently short enough to capture the gaps between successive vehicles at discharge speeds.

Algorithm for Identification of Spillover

Spillover creates detrimental effects for the operation of upstream traffic signals. Identification of spillover is particularly important because it indicates that traffic congestion has started to spread out in the network involving multiple intersections. To identify spillover using traffic signal phase and detector data, we first need to illustrate the concept of Queue-Over-Detector (QOD), *i.e.* the complete occupation of a detector for a relatively long time due to a vehicular queue.

Generally, there are two types of QOD. One is caused by red signal phases. Due to the normal cyclical signal timing, vehicles slow down and stop due to the red light and then resume traveling as the light turns green. If a vehicle in the queue stays on the detector because of a red light, detector occupancy time increases continuously, creating the first type of QOD. The second type of QOD is caused by spillover. When a queue spills back from a downstream intersection to an upstream intersection, the upstream intersection may be blocked and vehicles cannot be discharged even when the signal is green. Some vehicles will stay on the detector for a while creating a prolonged detector occupancy time after the traffic light turns green. Conceptually, the duration of the second type of QOD is equivalent to “the time period within a signal cycle in which the vehicles would be moving at the location of the detector during the green time in the absence of disturbances”, which was briefly discussed in Mueck (2002). Therefore, the detrimental effect of a spillover can be quantified by measuring the duration of the second type of QOD.

It is not difficult to identify QOD using high-resolution phase and detector data since it is simply indicated by a relatively large occupancy time (or percentage occupancy value keeping at 100% for an extended period). In our implementation, a threshold value of 3 seconds (which is roughly equivalent to 5mph of speed assuming a 22ft. effective vehicle length) is used for the identification of QOD.

We now need to differentiate between the two types of QOD. Figure 21 demonstrates both types of QOD by drawing each vehicle trajectory starting from upstream to downstream. Since the first type of QOD is caused by the red signal, the maximum occupancy time is the red interval. Considering the overflow queue from the last cycle and queue propagation at the green start, the first type of QOD can only happen within the range between the compression shockwave)v_4) and the discharge shockwave (v_2), which have the same velocity (see Figure 21). Therefore if QOD occurs between $(T_g^n + L_d / v_4)$ and $(T_r^n + L_d / v_2)$, it is the first type of “normal” QOD.

The second type of QOD, which occurs outside of the time interval $[T_g^n + L_d / v_4,\ T_r^n + L_d / v_2]$, indicates that a spillover has happened at a downstream location. This creates unusable green time, meaning that vehicles cannot be discharged during the green time because of the downstream queue. Therefore when a QOD event is identified, spillover occurs when the QOD starting time T_{start}^{QOD} or ending time T_{end}^{QOD} falls outside of the time interval $[T_g^n + L_d / v_4,\ T_r^n + L_d / v_2]$. An example case of the second type of QOD is demonstrated in Figure 21.

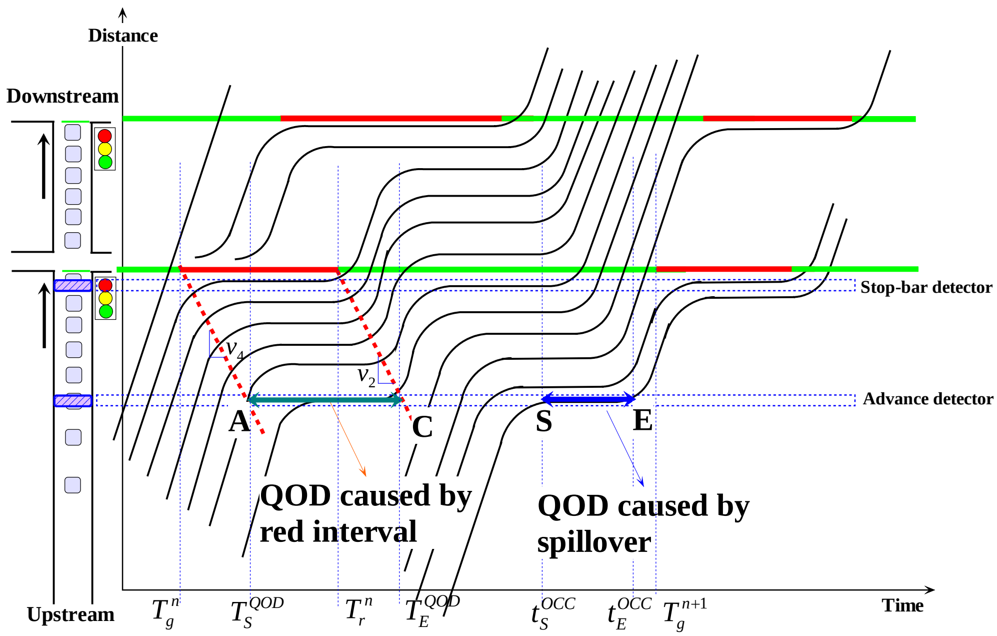


Figure 21. Queue-over-detector phenomena

The oversaturation severity index in the spatial dimension (S-OSI) can then be calculated as:

$$\text{S-OSI} = \frac{\textit{unusable green time}}{\textit{total available green time}} \times 100\% = \frac{\sum \left(T_{end,i}^{QOD} - T_{start,i}^{QOD} \right)}{G} \times 100\% \qquad \text{Eq. 10}$$

where $T_{start,i}^{QOD}$ and $T_{end,i}^{QOD}$ are the starting and ending times of the i^{th} occurrence of the second type of QOD.

In order to improve the robustness of the identification of the spillover condition, the maximum queue length of the downstream intersection should also be estimated using the method discussed previously. If the estimated maximum queue length at the downstream approach is longer than or equal to the link length, then oversaturation on that link is confirmed. This can be used to avoid some diagnosis errors caused by incidents (for example, the detector is occupied by a disabled vehicle for a relatively long period of time or the detector is placed at or near a bus stop and a transit vehicle stays on the detector for some time). These "incidents" may generate the second type of QOD, but it does not necessarily indicate an oversaturated condition on the downstream link.

Here we should note that the queue length estimation method discussed previously cannot be applied directly to an intersection with queue spillover from a downstream link. With spillover, queued vehicles can only be discharged when the downstream blockage is cleared and the signal light remains green. For the example shown in Figure 22, when the traffic light turns green at intersection *i* (T_r^n), queued vehicles begin to discharge, but the discharging process is disturbed because the queue at the downstream intersection *i+1* grows and eventually spills over to the upstream intersection *i*. A second type of QOD will be identified by the advance detector, starting at time T_{start}^{QOD}. QOD ends when the spillover is cleared, *i.e.* at the time T_{end}^{QOD}. Under this condition, the queue estimation method needs to be modified because break points A, B, and C need to be updated as A', B', and C', as shown in the Figure 22. Eq. 4 through Eq. 8 remain valid after updating the definitions of A, B, and C.

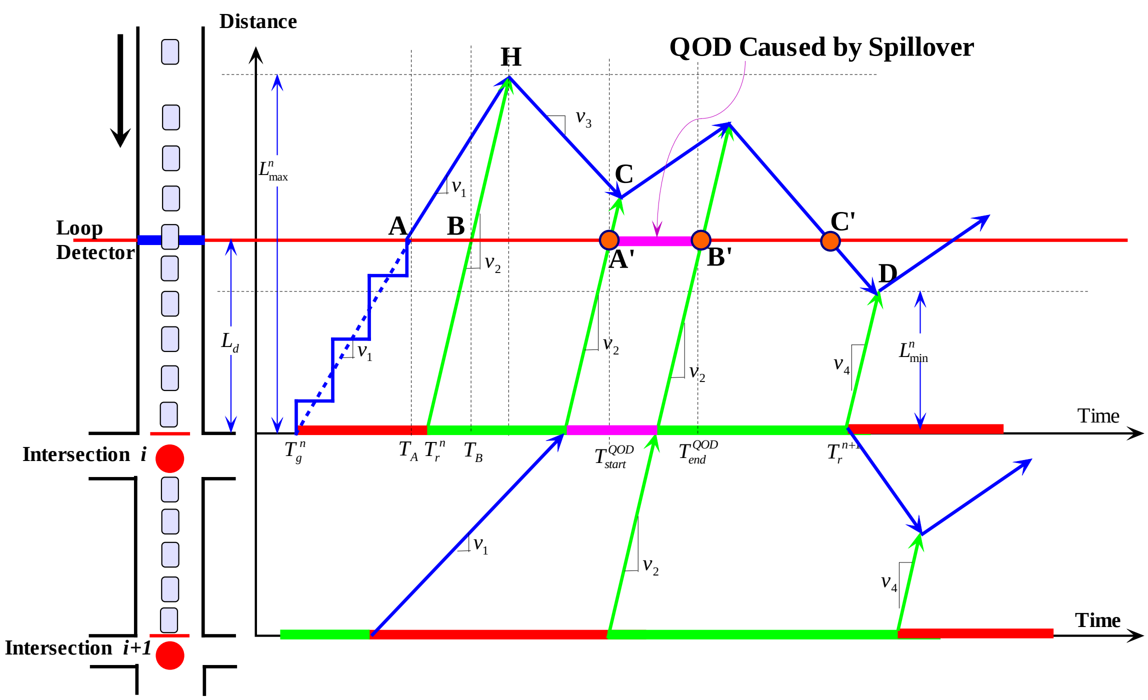


Figure 22. Updated breakpoints A', B', and C'

Example: Field-test Results

Trunk Highway 55 (TH55), a major arterial in Twin Cities, Minnesota, was used as the test site for validating the measurement of TOSI and SOSI values. Figure 23a illustrates the six coordinated intersections at the test location. Figure 23b shows the detector layout of four intersections along the arterial. Advance detectors are located on the major approaches and stop-bar detectors on the minor approaches. Stop-bar detectors are used to detect the presence of vehicles and advance detectors are located about 400 feet upstream from the stop line to detect vehicles for green extension on the coordinated phases. To verify the estimated queue length, we also installed (6’x6’) stop-bar and link entry detectors along TH55 at these six intersections (see Figure 23 for the detector configurations at the four major intersections). These additional detectors are not used for regular traffic signal operations; rather, the data collected from these detectors are used for algorithm verification. The detectors are monitored on a lane-by-lane basis.

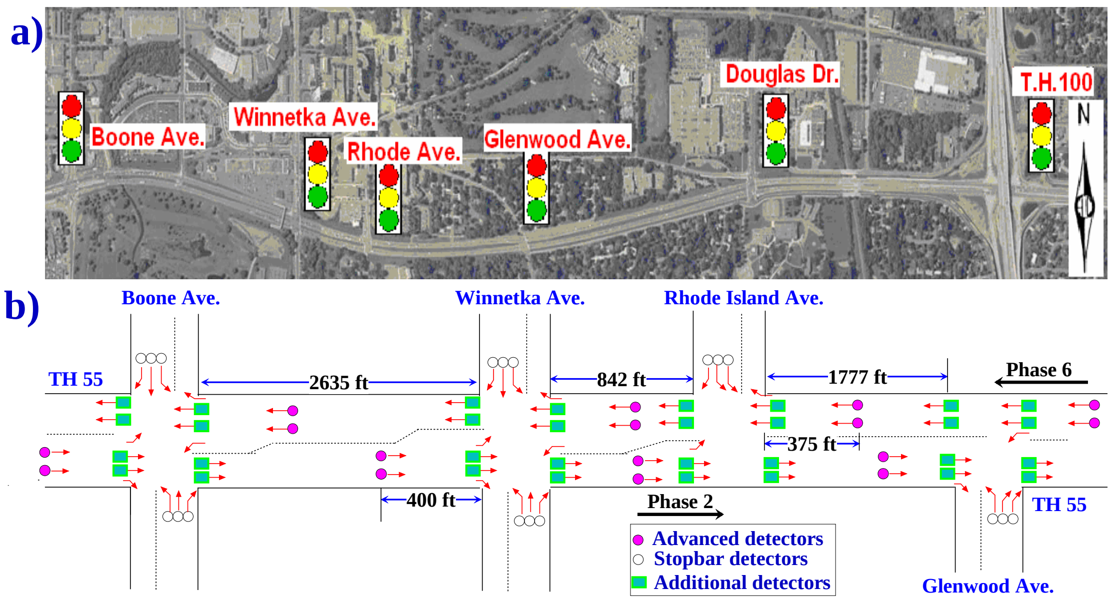


Figure 23. a) TH55 data collection site; b) detector layout

High-resolution event data including signal phase changes and vehicle-detector actuations are continuously collected from the six intersections and are archived by the SMART-Signal system and transmitted in real time back to the University of Minnesota’s ITS Laboratory. Figure 24 shows sample data collected at the study site including the start and end times of each vehicle-detector actuation event and every signal phase change event. The signal phase duration can be calculated from the time difference between the start and end of a signal event. The time interval between the start and end of a vehicle actuation event is the detector occupancy time. The time interval between the end of a vehicle actuation event and the start of next vehicle actuation event (from the same detector) is the time gap between two consecutive vehicles crossing the detector.

08:09:15.012, D8 on, 7.902s → Detector #8 on at 08:09:15.012; Vacant time is 7.902s
08:09:15.481, D8 off, 0.468s
08:09:16.761, G3 off, 29.389s → Green Phase #3 off at 08:09:16.761; Green duration time is 29.389s
08:09:16.761, Y3 on, 179.021s
08:09:17.620, D9 on, 2.686s
08:09:18.151, D10 on, 2.593s
08:09:18.307, D9 off, 0.687s → Detector #9 off at 08:09:18.307; Occupy time is 0.687s
08:09:18.823, D10 off, 0.671s
08:09:20.244, Y3 off, 3.482s → Yellow Phase #3 off at 08:09:20.244; Yellow duration time is 3.482s
08:09:21.649, D22 on, 80.953s
08:09:22.008, D22 off, 0.359s
08:09:23.242, G1 on, 172.806s → Green Phase #1 on at 08:09:23.242; Red duration time is 172.806s

Figure 24. Sample data collected at the test site

Estimation Results of Overflow Queue Length

Using the event-based data from the SMART-Signal system, the queue length estimation method discussed above was applied for the estimation of overflow queue lengths. We note that field evaluations of the queue length estimation method have been conducted by a Minneapolis-based transportation consulting firm and the evaluation results have been reported in Liu et al. (2009a). The average of the Mean Absolute Percentage Error (MAPE) of the queue length estimation is within 15%.

Figure 25 presents an oversaturation case based on the data collected by an advance detector on the eastbound approach at the intersection of Glenwood Avenue on Feb. 28, 2008. As indicated in Figure 25, overflow queues appeared at the end of the first two cycles. In this particular case, the reason for the occurrence of oversaturation was signal preemption on the side street which created a shorter cycle length in the second cycle (the cycle length was 132 seconds during the preemption which was 48 seconds less than the normal cycle length). Due to the insufficient green time, some queued vehicles could not be discharged until the next cycle.

Using the estimated overflow queue length, the oversaturation severity indices for these two cycles are estimated at 7.5% and 7.0%, meaning that at least 7.5% and 7.0% of green time in these cycles will be used for the discharge of the overflow queue. In the calculation of these severity indices, we have assumed that the space headway for the jammed condition is 25 feet and the saturation headway is two seconds.

It should be noted that the estimated maximum queue lengths (500 - 600ft) during these three cycles are not long when compared with the link length (1777ft from Glenwood to Rhode Island). However, overflow queuing occurs at the end of the first two cycles indicating that the volume of traffic joining the discharging platoon after the last stopped vehicle has started to move is rather high. A portion of those newly arriving vehicles joins the discharging platoon but could not pass the intersection during the green phase, which forms the overflow queue.

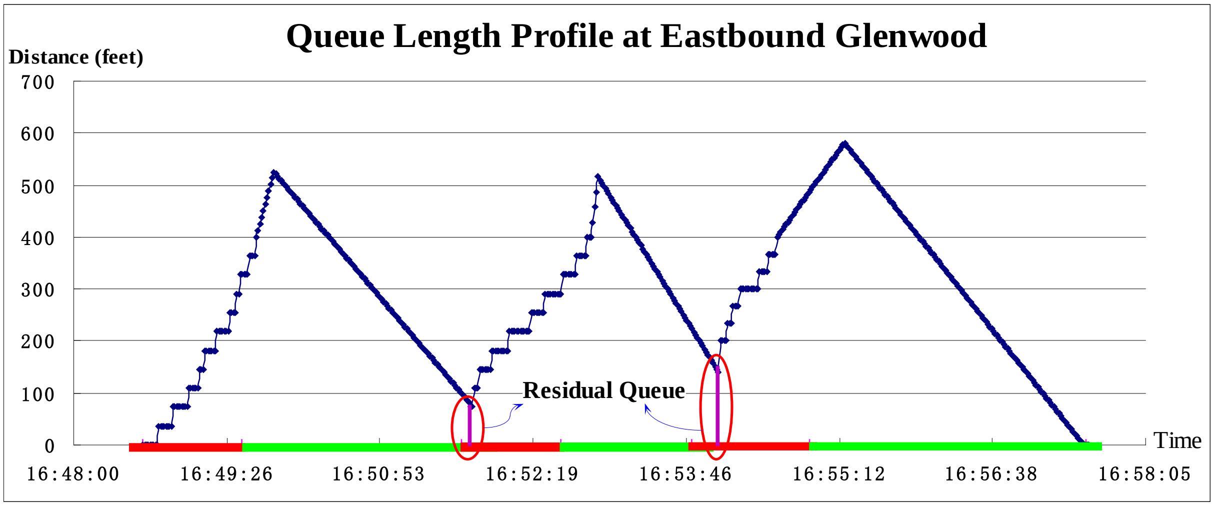


Figure 25. Estimation results of overflow queue for eastbound approach at Glenwood

As we discussed previously, if the entire green time is used for queue discharge, the departure shockwave cannot be identified. In this case we can only measure the minimum of the maximum queue length within a cycle. We present such cases in Figure 26 below. The data was collected from an advance detector on the eastbound approach at the Boone Avenue intersection on Feb. 28, 2008. During the first five cycles, the departure shockwaves could not be identified, so we can only estimate the minimum values of the maximum queue length. The queue lengths in this case are quite long, averaging around 1500 feet in the first five cycles. The minimum values of the overflow queue length are also estimated, as shown in Figure 26. The minimum oversaturation severity indices (TOSI) are estimated at 9.8%, 19.4%, 10.5%, 11.3%, and 10.3% for these five cycles. The oversaturated condition persists until the sixth cycle.

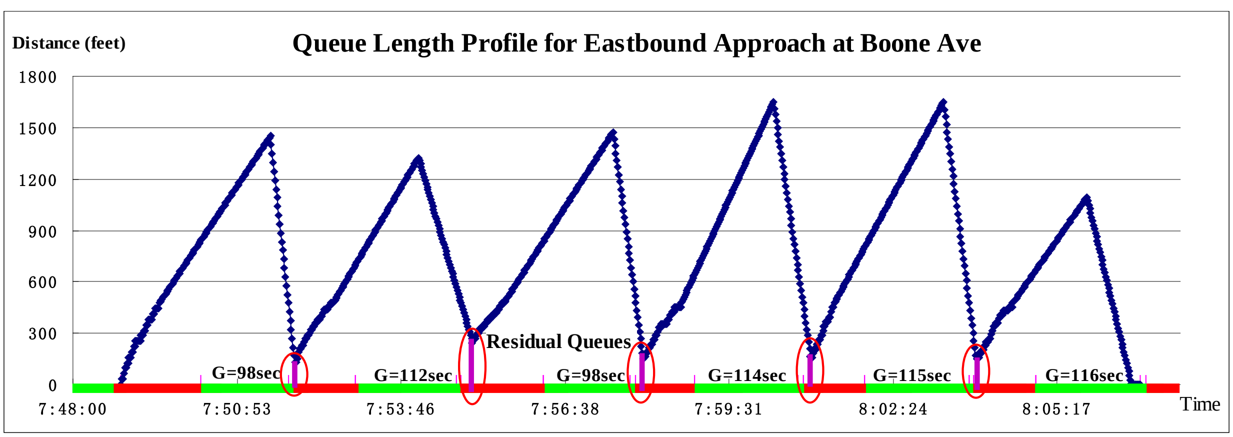


Figure 26. Estimated overflow queue length for eastbound approach at Boone Avenue

Example Results for Detection of Spillover

The spatial detrimental effect caused by oversaturation is characterized by spillover, which can be diagnosed by identifying the second type of QOD. In Figure 27, we present the detector occupancy time within an afternoon peak hour cycle on Nov. 17, 2008 for westbound TH55 at Rhode Island Avenue. As shown in the figure, QOD caused by spillover is identified. This means that vehicles cannot be discharged from the intersection although the traffic light is green. Oversaturation is therefore identified at this intersection for this cycle.

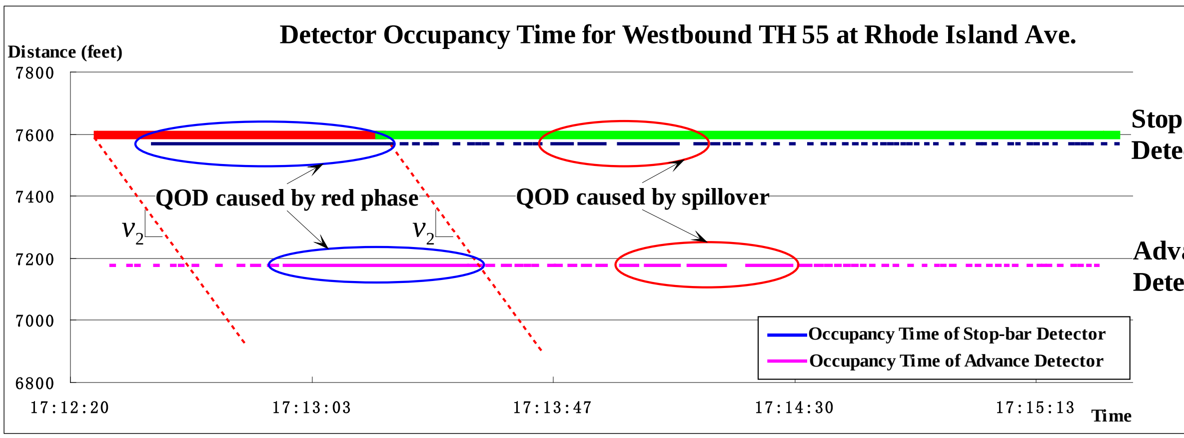


Figure 27. Identification of QOD caused by downstream spillover

To further verify that there is a spillover happening in the downstream link, vehicle trajectories are derived based on the vehicle events collected by the advance detector at the intersection of Rhode Island Avenue. A simplified car-following model, which is similar to the one developed by Newell (2001), is used to generate the vehicle trajectories for the purpose of illustration. Vehicles are assumed to accelerate if their speeds are less than free-flow speed (55mph on this particular arterial) and decelerate if higher than the free-flow speed, approaching a red light or the back of a queue. The distance between any two vehicles satisfies a safety distance constraint which is determined by the speeds of the two consecutive vehicles. The driver's reaction time is set at 1.0 second and space headway between two vehicles in a stationary queue is assumed as 25 feet. The maneuver decision of a vehicle to pass the intersection or not during the yellow time is determined according to its speed, remaining yellow time, and distance from the front vehicle. Lane-changing behaviors are not taken into account in this model.

The estimated vehicle trajectories starting from the advance detector line at the intersection of Rhode Island Avenue and ending at 500 feet downstream from the intersection of Winnetka Avenue are presented in Figure 28. As clearly indicated in the figure, the downstream queue spills back from Winnetka to Rhode Island and blocks the Rhode Island intersection during the green time, resulting in QOD.

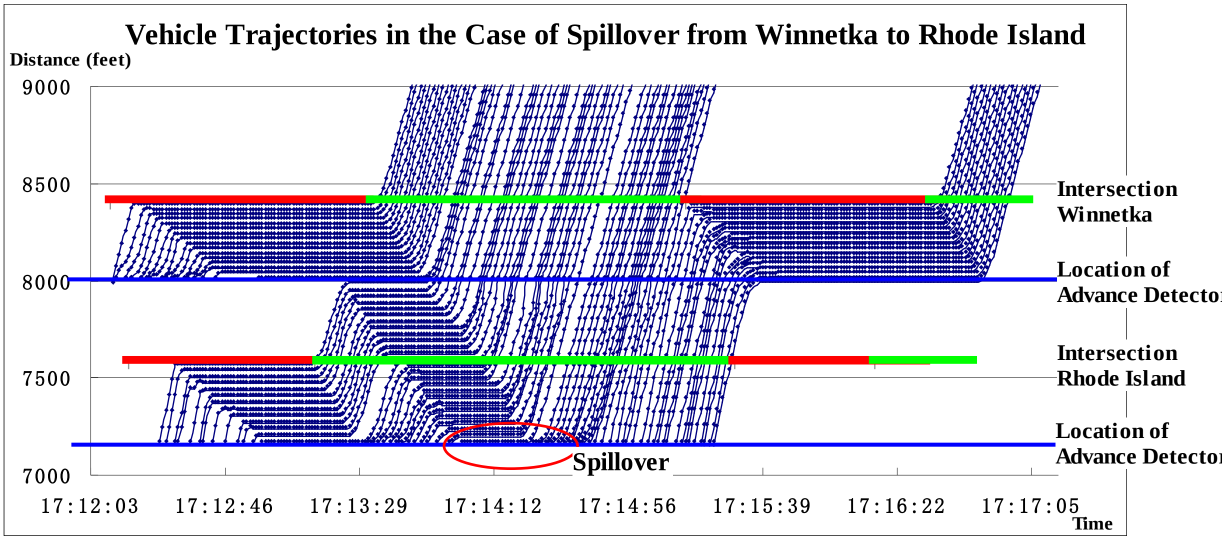


Figure 28. Vehicle trajectories in the case of spillover from Winnetka to Rhode Island

Further investigation indicated that this spillover started at 17:06:31, continued for approximately 30 minutes and ended at 17:36:31. The QOD due to downstream spillover was found in nine consecutive cycles. This is also confirmed by observing the queue length profile of the downstream intersection at Winnetka Avenue during these cycles. As shown in Figure 29, the maximum queue lengths for these nine cycles are around 1200 – 1500ft, which is significantly longer than the link length (842feet). This indicates that, during these cycles, the Rhode Island intersection must be blocked for a portion of green time (S-OSI > 0). Interestingly, because of the reduction in usable green time, overflow queues were also generated at the Rhode Island intersection for some cycles (i.e. T-OSI > 0).

This demonstrates that the oversaturated traffic condition at Winnetka (T-OSI > 0 and S-OSI = 0) has spread upstream, leading to insufficient green time to discharge the queue (T-OSI > 0 and S-OSI > 0) at Rhode Island. Please see Table 7and Table 8 for oversaturation severity index estimates at Winnetka and Rhode Island. Since there was no downstream blockage at Winnetka, S-OSI was always zero during that time period.

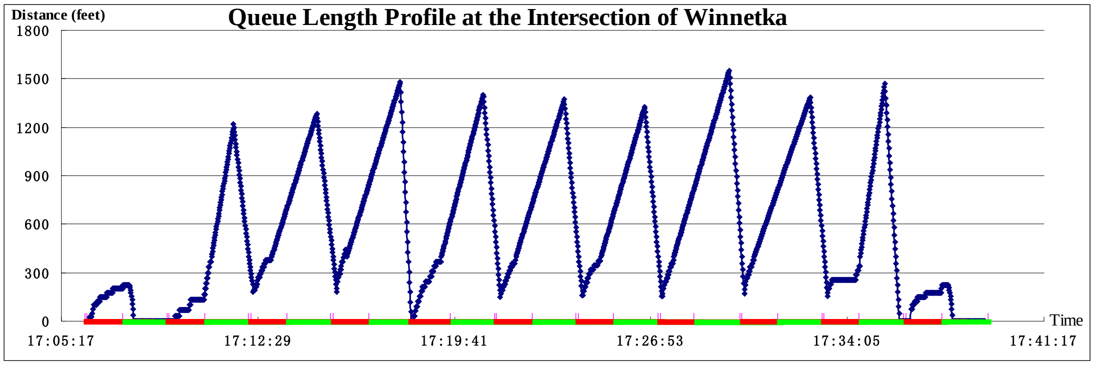


Figure 29. Queue length profile at the intersection of Winnetka when Rhode Island intersection is oversaturated

Table 7. Oversaturation Severity Indices (OSI) for Winnetka Avenue Intersection

Winnetka Avenue						
Cycle Start	Available Green (sec)	OSI: Created by Overflow Queue			OSI: Created by Spillover	
		Overflow Queue (ft)	Unusable Green (sec)	T-OSI (%)	Unusable Green (sec)	S-OSI (%)
17:06:14	101	0.0	0.0	0.0	0.0	0.0
17:09:14	101	180.3	0.0	0.0	0.0	0.0
17:12:14	101	178.8	14.4	14.28	0.0	0.0
17:15:14	101	0.0	14.3	14.16	0.0	0.0
17:18:14	101	149.1	0.0	0.00	0.0	0.0
17:21:14	101	157.6	11.9	11.81	0.0	0.0
17:24:14	102	156.4	12.6	12.36	0.0	0.0
17:27:14	106	130.1	12.5	11.81	0.0	0.0
17:30:14	101	153.4	10.4	10.31	0.0	0.0
17:33:14	105	0.0	12.3	11.69	0.0	0.0
17:36:14	102	0.0	0.0	0.00	0.0	0.0

Table 8. OSI for Rhode Island intersection

Rhode Island						
Cycle Start	Available Green (sec)	OSI: Created by Overflow Queue			OSI: Created by Spillover	
		Overflow Queue (ft)	Unusable Green (sec)	T-OSI (%)	Unusable Green (sec)	S-OSI (%)
17:06:31	136	0.0	0.0	0.0	0.0	0.0
17:09:31	136	0.0	0.0	0.0	3.0	2.2
17:12:31	136	89.6	0.0	0.0	28.0	20.6
17:15:31	136	164.3	7.2	5.3	28.8	21.2
17:18:31	136	0.0	13.1	9.7	15.0	11.1
17:21:31	136	180.4	0.0	0.0	41.7	30.6
17:24:31	135	165.3	14.4	10.7	34.1	25.2
17:27:31	139	138.2	13.2	9.5	25.2	18.1
17:30:31	120	125.3	11.1	9.2	16.3	13.6
17:33:31	141	0.0	10.0	7.1	8.6	6.1
17:36:31	135	0.0	0.0	0.0	0.0	0.0

Summary of Diagnostics for Severity of Oversaturation

In this section, we presented a methodology for measurement of oversaturation (overflow queue lengths) from conventional detectors and high-resolution phase timing data and two derived metrics for the severity of oversaturation. The temporal detrimental effect of oversaturated conditions is characterized by the occurrence of overflow queues and the spatial detrimental effect of oversaturated conditions is characterized by the effect of downstream spillover on upstream queue discharge. We developed two algorithms to identify oversaturated approaches. The first method estimates the overflow queue length using a simple traffic model based on the estimation of forward and backward traveling shockwaves as vehicles arrive and depart at the approach. The second method estimates the effects of downstream spillover by identifying long detector occupancy times during the green phase.

Two OSI were defined for quantitative characterization of the two types of oversaturation. The TOSI measure indicates the percentage of green time that is spent during the next cycle clearing the queue from the previous cycle. The SOSI measure indicates the percentage of green time that is wasted because of downstream blockage. Our field-test results on the TH55 arterial in the Twin Cities area demonstrate that the developed algorithms are very effective in identifying and quantifying oversaturated conditions. The approach to collecting and processing this information was implemented using the SMART-signal hardware harness and field processors developed in other previous research, but the models are simple enough to be implemented by others with access to second-by-second signal timing and lane-by-lane detector data.

The presence of TOSI > 0 indicates the need for additional green time for a phase. For an individual intersection, the values of TOSI provide a direct indicator of how much additional green time is necessary for that phase to clear the overflow queue. Similarly, SOSI > 0 indicates the need for additional green time at a downstream location, or might indicate that the upstream green time should be shortened. In simple situations with just one affected approach, mitigation strategies are relatively simple to identify. However, when presented with oversaturation on more than just a single phase at an intersection or with TOSI and SOSI >0 for many intersections on a route it is not obvious what a rational approach to solving the problem might be. In addition, TOSI and SOSI do not directly measure storage blocking and starvation symptoms that are also problematic issues in oversaturated conditions.

Our approach to the remainder of the research based on this initial development of real-time diagnostics is three-fold. First, we look at the problem from a "top-down" perspective (whereas, considering diagnostic metrics would be considered a "bottom-up" approach) by developing a methodology for design and evaluation of signal timing plans for oversaturated conditions. This methodology considers design principles for cycle, splits, and offsets to minimize the occurrence of TOSI and SOSI effects before they occur at all. The methodology considers both the design of timing plans and the schedule for implementing them by time-of-day to address the three primary regimes of oversaturated conditions (loading, processing, and recovery). This methodology is described in the next section of the report. Two simulation examples of applying the methodology on real-world test networks are described in Chapter 4.

Secondly, we describe the development of a tool for online application of pre-configured strategies (that could be designed with the methodology presented in the preceding section) using real-time detector data to determine when to switch from a "normal" operation to a plan designed for "loading" or from a plan designed for "loading" to one designed for "processing" and so on. Each strategy can be triggered using detection of TOSI and SOSI above configured thresholds. One simulation example of applying the approach to a real-world network is detailed in Chapter 4.

Finally, we developed a heuristic approach using measured TOSI and SOSI values to directly modify green times on an oversaturated route. The heuristic approach adjusts splits and offsets on the route to drive TOSI and SOSI as close to zero as possible in order to maximize throughput on the critical route. This methodology and its application to two real-world networks are described in Chapter 4.

A Multi-Objective Methodology for Designing and Evaluating Signal Timing Plans Under Oversaturated Conditions

In the previous section we described algorithms to calculate the intensity of oversaturation on an individual approach. This detailed information is necessary from a bottom-up perspective to quantify the presence and effects of oversaturated conditions. In this section, we describe a top-down methodology which will be used to explore the performance differences between signal timing strategies in oversaturated conditions. There are four essential parts to the design of timing plans for handling oversaturated conditions:

(1) The identification of the critical routes and the flows along those routes,
(2) The determination of ranges of feasible timing plan parameters that meet criteria to address oversaturation on those routes,
(3) The selection of the objective function to solve for optimal timings, and
(4) The scheduling of a sequence of timing plans that are tailored to each regime of oversaturated operation.

Based on the lack of significant research in this area, for this part of the project, we explore several questions in this area and develop a comprehensive process for implementing this analysis:

1. Assuming certain design principles for timing plan design during oversaturated conditions, does optimization for different objectives result in substantive differences in performance?
2. As of traffic demand and resulting congestion changes over time, what is a best schedule for implementation of multiple timing plans?
3. Are there sets of timing plan parameters that out-perform other timing plan parameters for more than one optimization objective?
4. What design components of a complicated mathematical process can be distilled into principles that practitioners might be able to apply?

This process is then applied in several simulation test cases which illustrate typical results for real-world networks. Comparisons of one set of timing parameters and the schedule for implementing multiple timing plans is then made using the comprehensive output measures available from the simulation model.

It is first 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. Designing optimal control strategies in these cases becomes challenging because of the wide range of potential approaches that might be taken to mitigate any specific scenario. The systematic mathematical optimization procedure described in this section was developed in an attempt to provide a rational approach for arriving at specific timing plan parameters that take into account both oversaturated conditions and critical route flows. This is an experimental approach that builds upon previous research but still requires additional development to arrive at a process that a practitioner can directly apply.

Overview of the Methodology

The proposed methodology shown in Figure 30 starts by identifying the critical routes in the network. Using traditional origin-destination (O-D) estimation methods during congested periods can be challenging because congestion causes low volume counts on links and long travel times. Nevertheless, the analysis can still be carried out with reasonable accuracy. Critical routes can be determined using judgment calls by the local staff after on-site observation of vehicle flows. In most cases, experienced local staff would have a reasonable idea of the critical routes and peak periods. They can build from this knowledge to determine critical route design scenarios. This part of the analysis is critical in the overall timing strategy framework.

Once the network's critical routes have been identified and mapped into the network configuration, the problematic traffic symptoms of oversaturation can be directly determined (i.e., spillback, starvation, storage blocking, and cross blocking). A wide array of control strategies can be considered to address the detrimental effects of these problematic traffic symptoms. We provide a systematic procedure for determining a set of cycle, splits, and offsets for a particular objective function. Phase sequence, phase reservice, and other tactical treatments such as green time extension or truncation are not directly included in the procedure. Strategic measures, such as gating, are applied in an ad-hoc manner by determining the size of the network (traffic is gated at the most upstream location(s) considered in the system).

Evaluation of potential control strategies is then carried out by extracting performance measures from the simulation output. Optimal control strategies can then be determined with regard to each performance objective using multi-objective (Pareto front) analysis. The details of the design principles for control strategies that address oversaturated conditions are presented in the next sections. These timing plan design principles are based on known traffic engineering concepts from the literature.

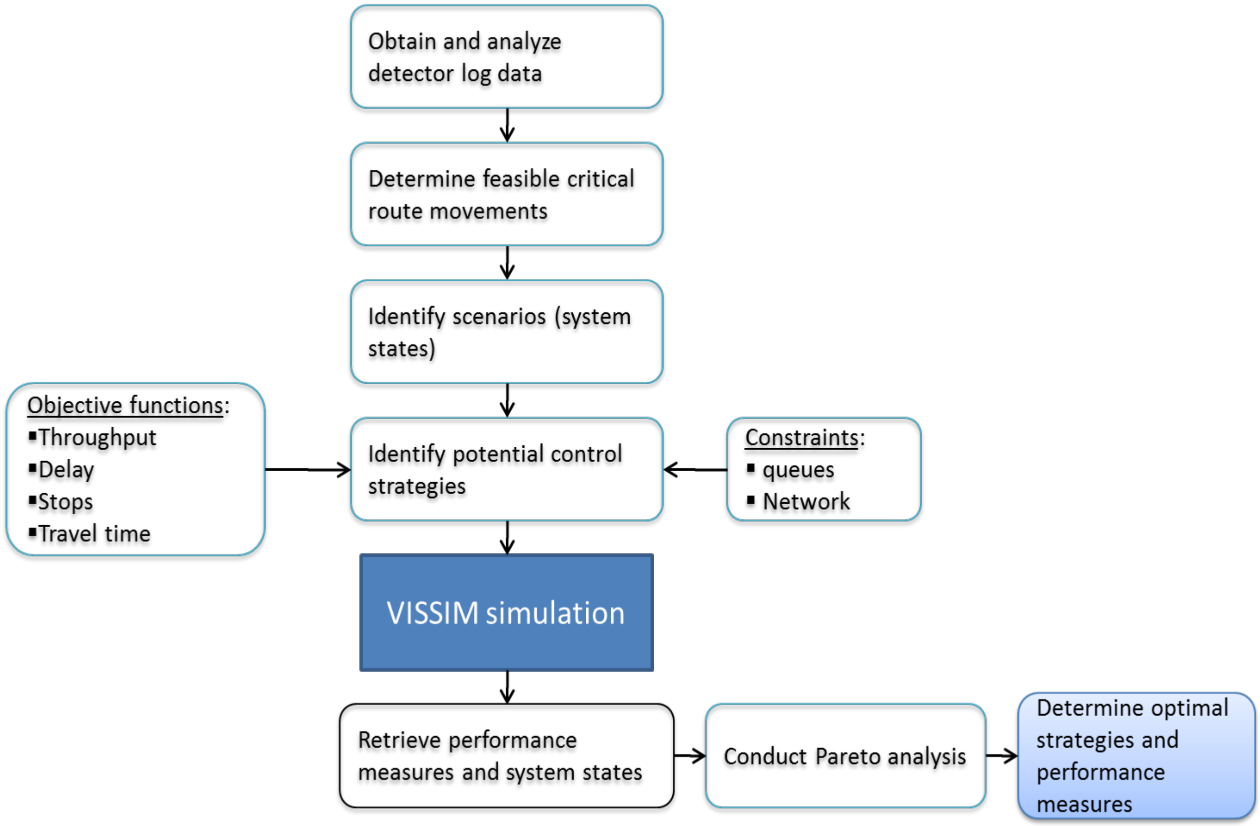


Figure 30. Framework for determination of control strategies used in this research

Framework for Determination of the Traffic Flows on Critical Routes

Volumes on critical routes can be determined using detailed theoretical analysis, including synthetic O-D estimation, volume correlation analysis, and clustering of detector data. It can also be determined using relatively new technological methods, such as Bluetooth data or data from the upcoming Connected Vehicles technologies. Alternatively, it could also be reasonably estimated using field observation and local knowledge of traffic patterns on the congested system.

The methodology used in this research is illustrated conceptually in Figure 31. This step is indispensable in designing appropriate signal control strategies in the next step. There are two main parts to the identification of critical route flows. The first is to use correlation analysis between detector counts at different locations to determine which detector movements change simultaneously. For example, if all eastbound through movement detector volumes are highly correlated (mostly increase and decrease together) and their volume is high, then one can consider that as an indication that the eastbound movements on a corridor constitute a critical route. The same rationale can be applied to a whole network to trace all of the critical routes. The magnitude of traffic flow on each route determines the importance of each route and its contribution to the design of the control strategy. The second key principle is the concept of volume spillover. If the volume on the critical route cannot be accommodated in the current time period, one must consider that additional flow (spillover) is present during the next analysis time period and one must consider timing plan parameters (cycle, splits, and offsets).

In order to design a set of control strategies that can account for the detrimental effects of congestion, a set of problematic scenarios should be designed to encompass the extremes of the possible traffic patterns. That is, when the same system detector counts on a network can correspond to two possible critical route scenarios, both possible scenarios should be taken in consideration when designing control strategies. This is illustrated in Figure 31 in the step where the maximum feasible volume is determined on each critical route.

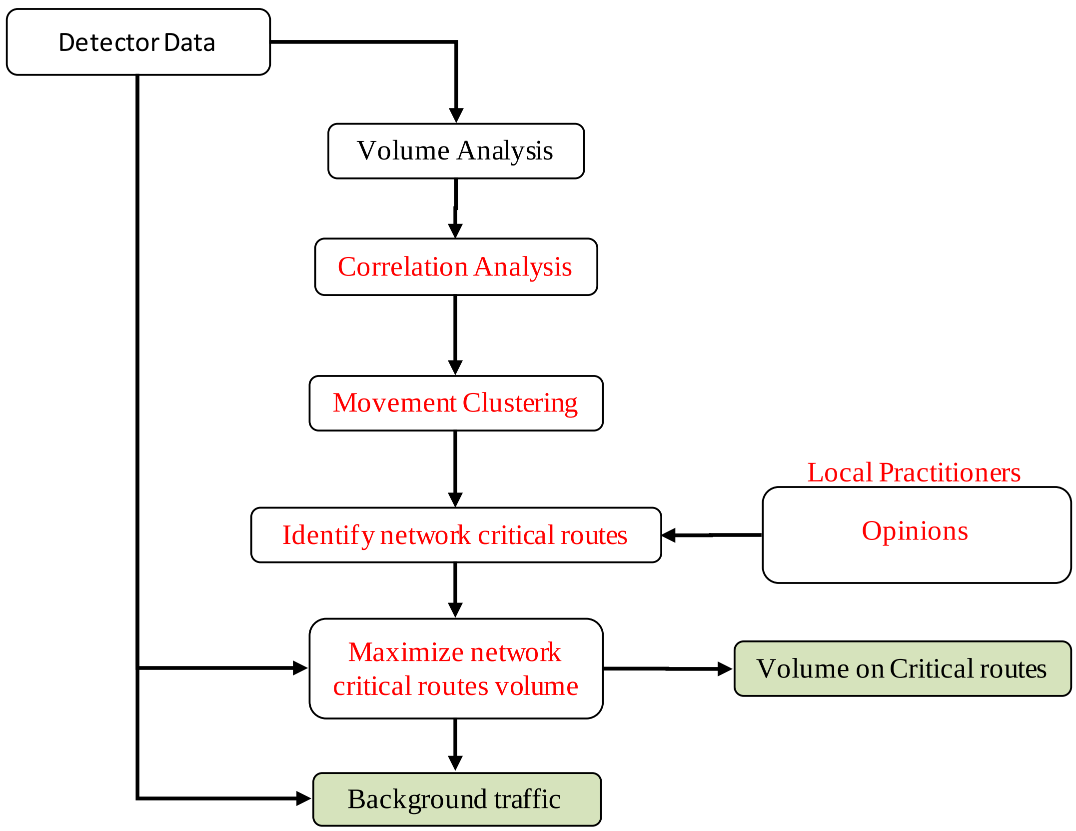


Figure 31. Framework for volume estimation on critical routes

Figure 32 illustrates the research framework for designing timing plans for critical routes under oversaturated conditions. The framework utilizes network configuration (i.e. link lengths, storage bay lengths, etc.) and critical route volumes to produce the basic signal timing parameters of cycle time, splits, and offsets. Cycle length in oversaturated conditions is determined based on the network's link geometry and volume levels. Cycle length cannot exceed certain values on certain links to prevent queue spillback. Splits and offsets are then determined for each intersection through an iterative procedure. This iterative procedure protects the capacity on critical routes and attempts to prevent spillback and starvation which results in several sets of timing parameters. Since we now have a range of possible timing values, simulation is then used to evaluate the performance of the generated timing plans for the three principle optimization objectives: minimize delay, maximize throughput, and manage queues. Next, multi-objective analysis is conducted to eliminate the dominated timing plans and to identify the best non-dominated sets of timing parameters (cycle, splits, and offsets). This framework is illustrated in Figure 32and described in more detail in the following section.

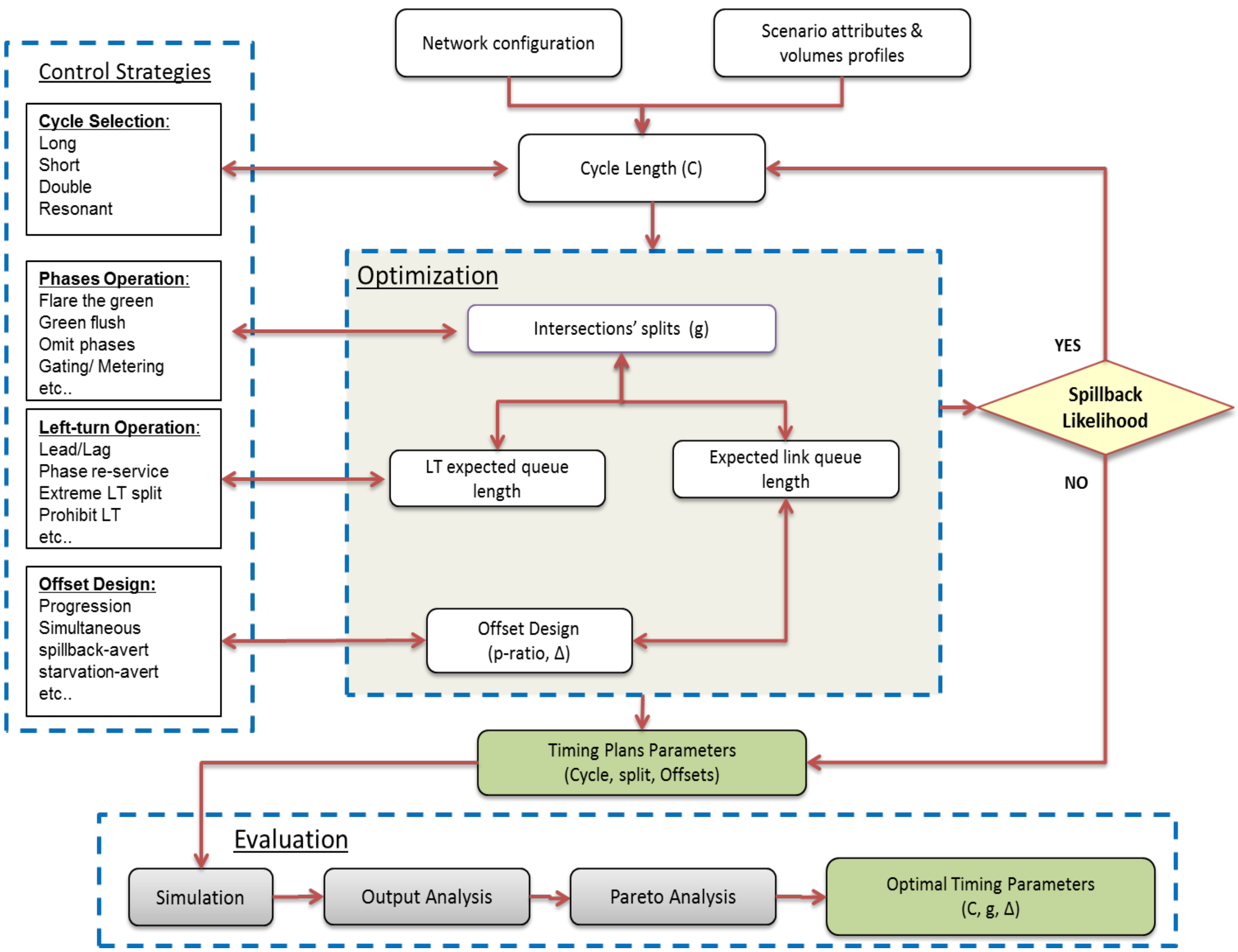
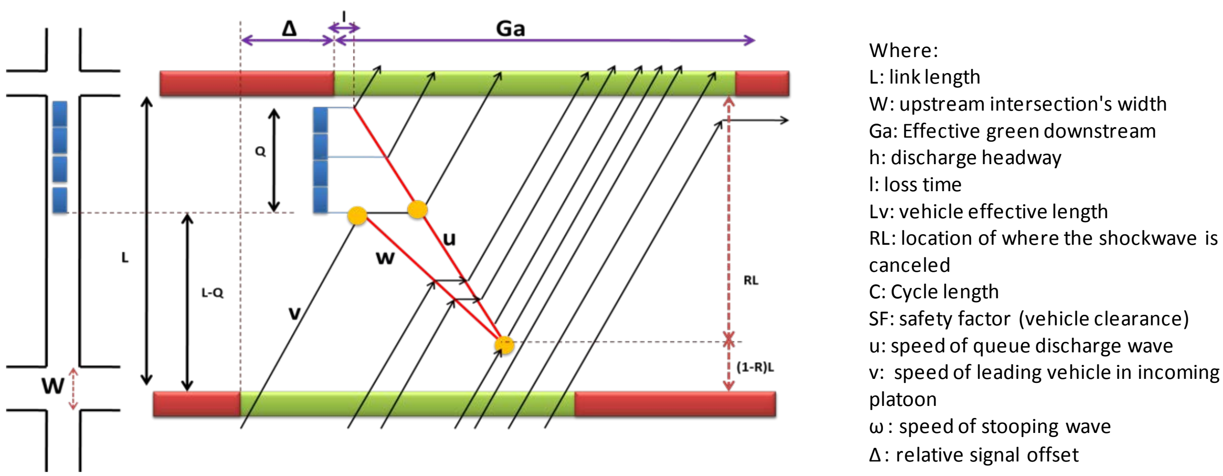


Figure 32. Timing framework for oversaturated conditions

Cycle Length Determination

Determining cycle length for the network during oversaturated conditions is a critical task. Failure to determine the optimal cycle will increase the possibilities of spillback and intersection blockage. In undersaturated conditions, cycle length is directly a function of the volumes and capacities of the approaches in the system. In oversaturated conditions, however, cycle length is a function of the storage capacity of the links, the arrival rate during red intervals along the arterial, and the green split ratio at each intersection. For example, a network including a short link with a high arrival rate during red on that link would require a shorter cycle length to prevent spillback into the upstream intersection. Networks with longer approach distances can tolerate longer cycle times before problematic symptoms arise.

The Internal Metering Policy (IMP) developed by Lieberman et. al. (2000), provides an upper bound of intersection background cycle length to avoid spillback at upstream intersections. This is the maximum cycle that ensures the queue formation shockwave dissipates before reaching the upstream intersection. Figure 33 illustrates the calculation of the maximum cycle length that prevents spillback (Lieberman, et al, 2000).



$$C \leq \frac{Lh}{L_v} \frac{\left(1 - \frac{W}{L} - \frac{(SF)L_v}{L}\right)}{\left[\frac{G_a - l}{C}\right]}$$

Figure 33. Shockwave at signalized intersection

Some typical upper bounds for cycle lengths using this equation are shown in Figure 34. A family of curves is presented that illustrates the highest feasible cycle length for specific link lengths and split ratios for a specific approach demand. Each of the blue lines indicates that cycle lengths which are below the line will not create spillback to the upstream intersection and generate overflow queues. For example, assuming the 800vph approach demand, a split ratio of 0.5 for the downstream through phase, and a 700ft link length, the maximum cycle time that could be implemented before spillback would occur is 150 seconds.

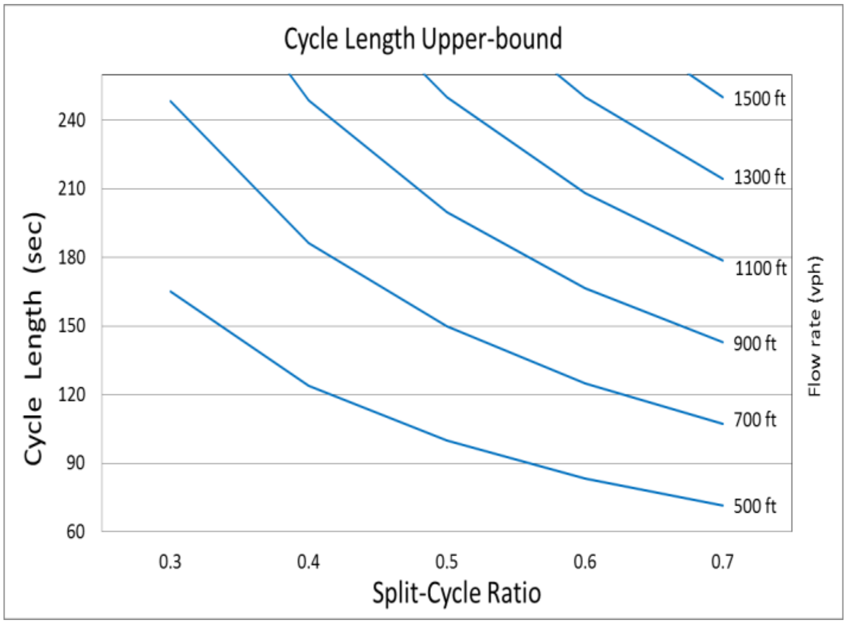


Figure 34. Upper-bound of cycle length that prevents spillback

Roess et. al., (2004) presented another formulation to calculate the maximum cycle length in oversaturated conditions. In this formulation, maximum cycle length is a function of the downstream critical lane discharge rate (v) and the link length (L).

$$C \leq \frac{L}{D} \times \frac{3600}{v_i} \qquad \text{Eq. 11}$$

Some typical upper bounds for cycle lengths using this equation are shown in Figure 35. A family of curves is presented that relate the highest feasible cycle length for a specific link length (800ft) and red split ratios for varying saturation flow rate. Each of the blue lines indicates that cycle lengths which are below the line will not create spillback to the upstream intersection and generate overflow queues. For example, assuming an 1800 vph saturation flow rate and a 0.5 red split ratio, the maximum cycle calculated with this equation is 130 seconds.

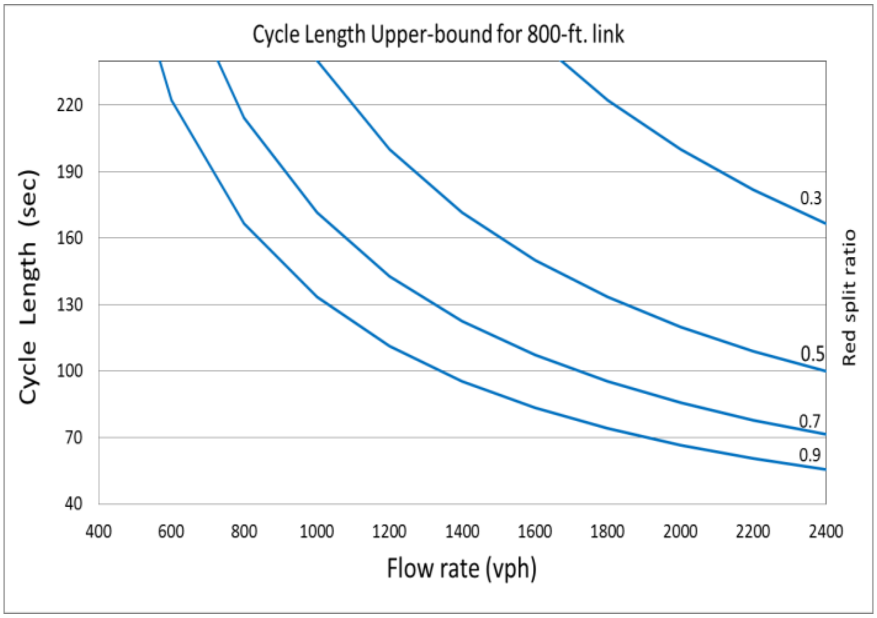


Figure 35. Upper-bound of cycle length as a function of saturation flow rate

As illustrated in the figures, short link distances create the most severe limits on cycle length during high demand periods. While this theoretical cycle time calculation can be used as a guide, many other factors must be considered in the selection of the cycle including pedestrian requirements and minimum green times. In the optimization methodology presented here, we use the Lieberman equation for determination of cycle time as it results in more restrictive values. The Roess equation tends to permit or suggest higher cycles than are typically feasible.

Design of the three components (cycle, splits, and offsets) together is critical to arriving at a feasible control strategy with fixed timing parameter values. For example, short links can be protected from spillback not only by shortening the cycle length, but also by reducing the arrival rates during red by adjusting the offset values for the critical routes that pass through this link (i.e., by using offsets that avoid spillback). Another approach to avoid spillback is to provide additional green time at the downstream intersection (i.e. flaring the green) creating a metering effect at the upstream intersection. The following sections describe a methodology for determining splits and offsets that mitigate oversaturated conditions. Following the description of the methods for separately determining splits and offsets for a given cycle time, we describe a methodology for combining the two computations.

Determination of Splits

The optimum splits of an intersection depend on the relative demand on the conflicting approaches, discharge rates, and minimum green times that satisfy pedestrian and safety policy requirements. The maximum and minimum green times are typically regulated by agencies (e.g., minimum green time that accounts for pedestrian crossing time). In our design framework, the green splits are initially allocated based on the v/c ratios given a specific cycle time. In the case of phase failure (i.e., v/c> 1), the 95th percentile queue of the critical approach is checked against the storage capacity. If the storage capacity of an approach is violated, splits are re-allocated to avoid the spillback. Other methods can be used to protect critical movements of an intersection to minimize overflow queues. One method is to impede the flow on critical routes upstream at pre-specified gating links with considerable storage capacity. Another method is to double cycle the subject intersection creating a phase reservice effect. Double-cycling is most effective for heavy arrivals during the red phase. At this stage of timing plan design, these methods are applied based on engineering judgment. We do not yet have mathematical methods developed for a comprehensive selection of the various combinations of treatments.

Design of Offsets

Maintaining progression during oversaturated conditions is a difficult task. Queues that form as a result of insufficient downstream capacity and from heavy turning from upstream side streets inhibit the movement of platoons through the arterial, thereby reducing the overall system effectiveness. There are two competing objectives when it comes to designing offsets in oversaturated conditions: (1) preventing spillback at the upstream intersection and (2) maximizing the green time at the subject intersection.

The first objective is achieved by an offset that prevents the stopping shockwave from reaching the upstream intersection. The stopping shockwave progresses towards the upstream intersection during the red interval of the through phase, reducing the speed of the discharge wave at the upstream intersection. This may even cause blockage as illustrated in Figure 36. The second objective is to prevent starvation at the subject intersection. This is achieved if the first vehicles released from the upstream intersection join the discharging queue before crossing the downstream intersection. Further delay in releasing vehicles from the upstream location will cause downstream starvation where some of the green time is wasted between the discharge of the overflow queue and the arrival of the next platoon. This is illustrated in Figure 37. Fundamentally, the range of offsets that satisfies both conditions of starvation and spillback avoidance is computed from the length of the link, the vehicle travel speed, the overflow queue lengths, and the queue discharge rate. The fundamentals of spillback and starvation avoidance are described in the next two sections.

Offsets to Avoid Spillback

Spillback-avoidance offsets prevent spillback at the upstream intersection by causing the stopping shockwave to dissipate before reaching the upstream intersection. The ideal offset can be calculated using several methods that consider (a) vehicle dynamics, (b) queue perception impact that influences the approaching vehicle's speed, and (c) the platoon dispersion effect. The following equation is derived based on Newton's 2nd law of motion. For a given length (L) of a link, queue-link ratio (ρ), headway (h), vehicle length (L_v), and discharge rate (U_s). The value of (p) is the design value used for computing maximum and minimum offset values, such that

$$\Delta \geq \left(\frac{L}{u_s}\right) - \frac{L(1-\rho)}{L_v}.h \qquad \text{Eq. 12}$$

where Δ is the difference between the starting times of the upstream and downstream green phases. Note that Δ is not the offset that one would key into a traffic controller. ρ is a constraint or assumption on the length of the average overflow queue.

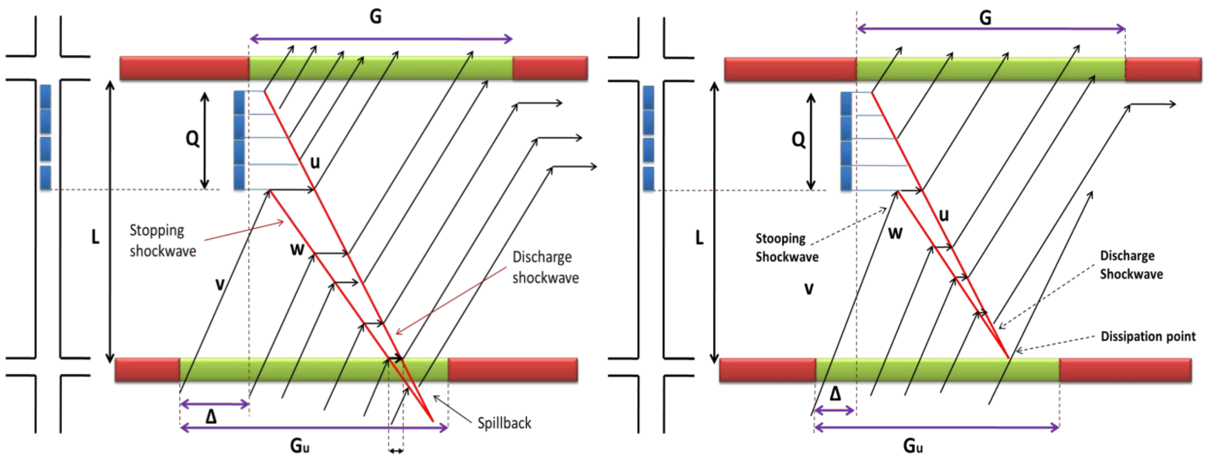


Figure 36. Spillback avoidance offset

Offsets to Avoid Starvation

Starvation occurs on the subject approach when vehicles discharging at the upstream intersection arrive later than the time that the standing (overflow) queue has been discharged. Starvation results in loss of capacity by the wasting of valuable green time. A starvation avoidance offset ensures that the first released vehicle joins the discharging queue at the downstream intersection just as the back of the queue begins to move. The maximum value of a starvation-avoidance offset can be computed as follows:

$$\Delta \leq \frac{L.\rho.h}{L_v} - \frac{L}{v} \qquad \text{Eq. 13}$$

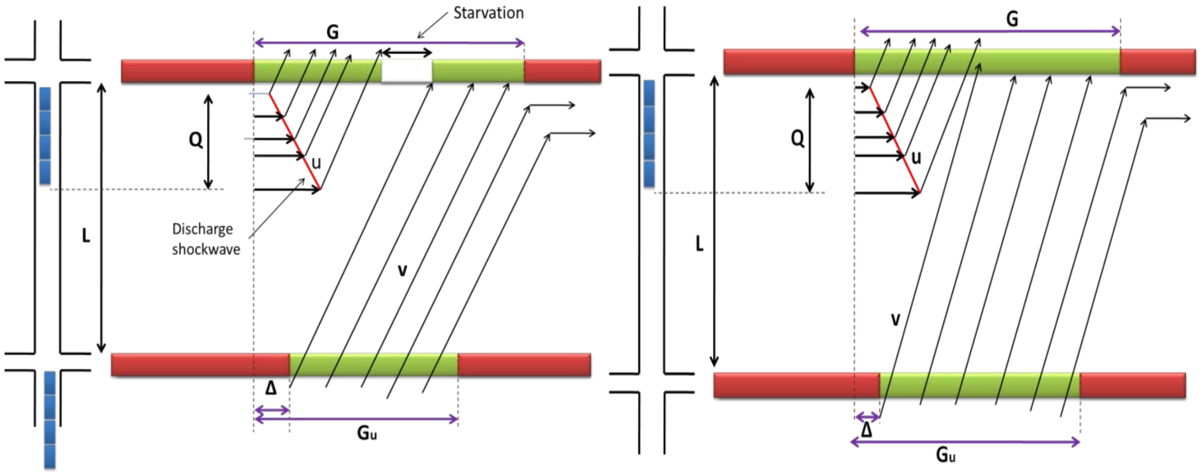


Figure 37. Starvation avoidance offset

Similar to the computation of the spillback avoidance bound on offsets, a design value for ρ must be assumed. By calculating the bounds on the offset values (max and min), a feasible zone for offset values can be established. This is illustrated in Figure 38. This feasible zone determines the offset range that can meet objectives of efficiently utilizing the capacity of both the upstream and subject intersections.

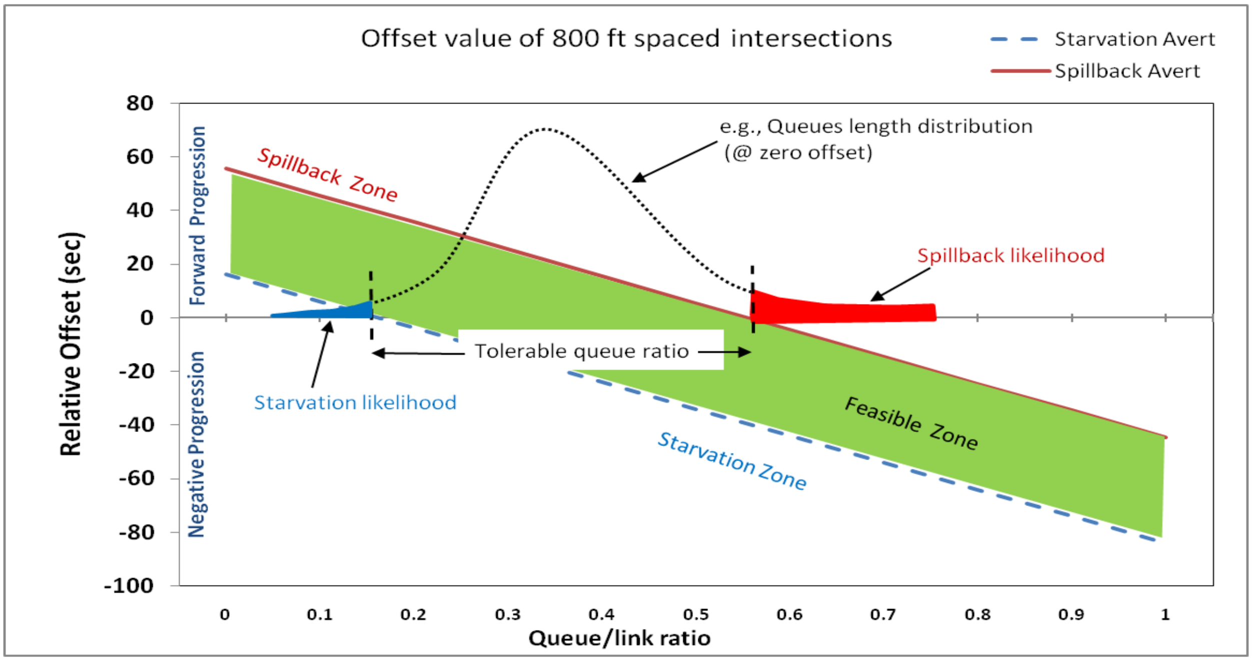


Figure 38. Offset values feasible region

As shown in Figure 38, the feasible zone of offsets that satisfies both objectives is shown as a function of the queue ratio (ρ). This feasible zone is highlighted in green. For example, if the queue ratio is 0.5 (half of the 800 ft link length is filled with a queue) then the relative offset is constrained in the region of offsets between (-30s, 10s). If the relative offset is less than -30s, the downstream green phase will be starved. The light will remain green after the overflow queue is discharged and before the oncoming platoon arrives. If the relative offset is greater than 10s, the oncoming platoon will arrive at the back of the queue too early and the resulting shockwave will spillback into the upstream intersection. As would be expected, the higher the queue length ratio becomes, the more negative the relative offset needs to be. If the offset value exceeds the cycle time, the modulus operation is used to determine the value.

The value of (ρ) is selected based on the expected overflow queue, link length, and vehicle travel speed. However, due to the stochastic flow rates of through and side street turning movements, (ρ) can be expressed by a probabilistic function that identifies the likelihood of both spillback and starvation. As illustrated in Figure 38, there is a range where (ρ) can be contained without violating the control objectives. However, if (ρ) 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 a waste of downstream green time. Therefore, it is essential to design the offsets to control the value of (ρ) within a range. This can be achieved by (1) reducing cross-street green time, and (2) prohibiting cross-street turnings during red to regulate the growth of the queue length on the receiving link. This procedure for determining offsets and splits is applied to the critical route. Creating queues on side streets and minor phases is sometimes unavoidable in order to address the problematic symptoms of oversaturation.

Combining the Design of Splits and Offsets

Figure 39 illustrates the split-offset calculation procedure for an oversaturated network. The initial splits based on calculation of v/c ratios are used as a starting point in an iterative procedure. Next, the expected queue lengths are estimated for approaches with v/c > 1. The ρ values for approaches that are part of critical routes are then constrained by pre-specified thresholds to inhibit the buildup of further queues. The resulting queue-to-link ratios are then used to determine the bounds of offsets that can avoid spillback avoidance and starvation. The following section lists the equations used in the calculation framework.

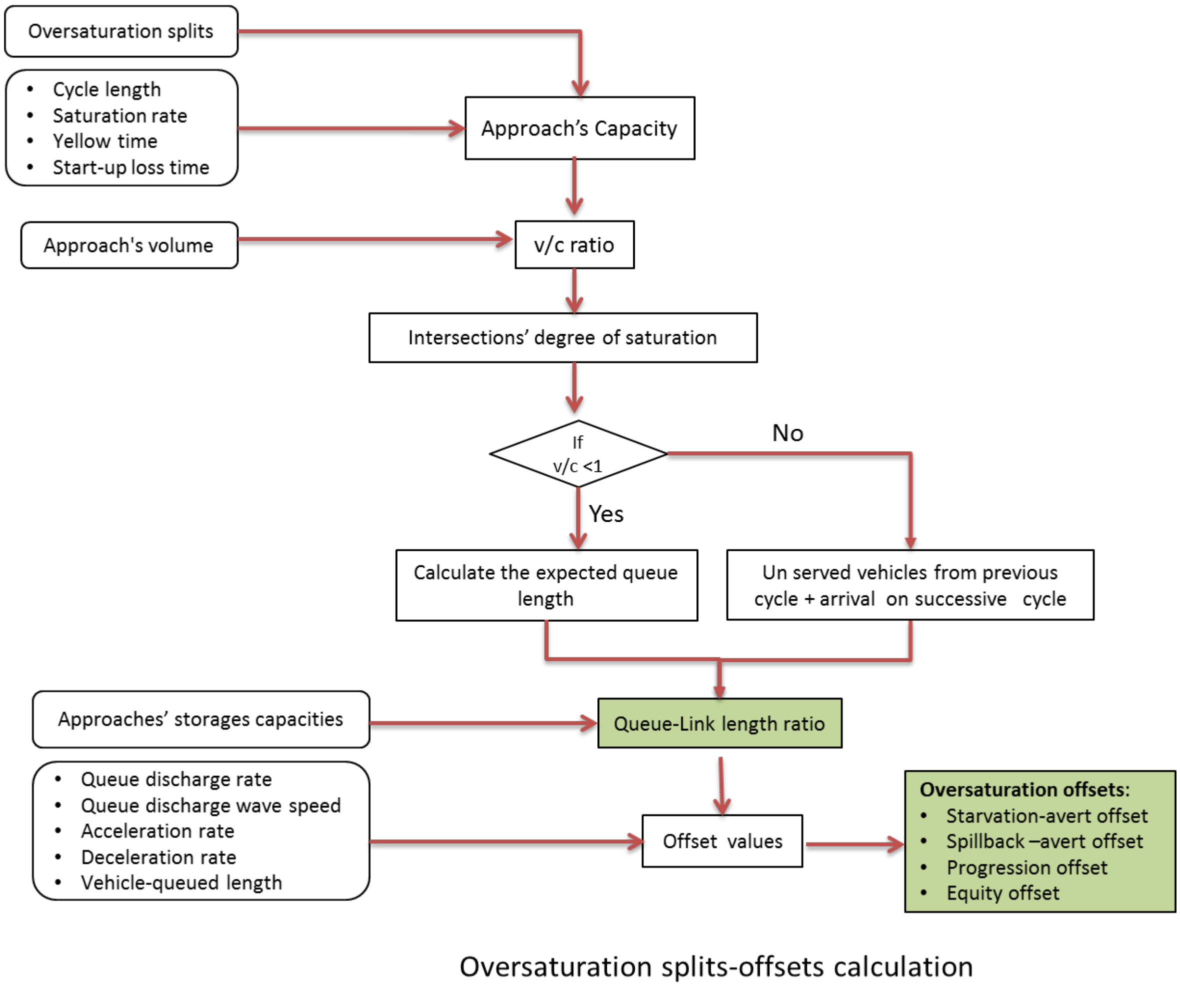


Figure 39. Split-offset calculation procedure

1-Volume per cycle (VPC): Compute the average arrival volume on an approach during a cycle length:

$$VPC = \frac{V \times C}{3600} \qquad \text{Eq. 14}$$

2- Compute the 95th percentile volumes on each link based on the expected volume and the peak hour factor (PHF).

$$V_{\alpha} = V \times PHF \times (1 + \sigma_{95th} \times \frac{\sqrt{VPC}}{VPC}) \qquad \text{Eq. 15}$$

3- Calculate the initial splits for each intersection based on the v/c ratio for each approach j of intersection i $(v/c)_{ij}$. The capacity of each approach is iteratively changed after the queue control offsets are computed in Step 6, such that

$$(v/c)_{ij} = \frac{V_{ij}}{s \times {}^{g_{ij}}/_{C}} \qquad \text{Eq. 16}$$

4- Calculate the expected queue length on each approach- (Q_{ij}) for v/c greater or less than 1:

v/c< 1

$$Q_{ij} = \frac{V_{ij}}{3600} \times (C - g_{ij}) \times (1 + \frac{1}{{}^{s}/V_{ij}}) \times \frac{L}{n_{ij}} \qquad \text{Eq. 17}$$

v/c> 1

$$Q_{ij} = \left[(V_{ij} - s \times \frac{g_{ij}}{C}) \times \frac{C}{3600} + V_{ij} \times \frac{C}{3600}\right] \times \frac{L}{n_{ij}} \qquad \text{Eq. 18}$$

5- Calculate the queue-link length ratio (p_{ij}) based on the expected queue length (Q_{ij}) from Step 4 and the available storage capacity for each approach (L_{ij}), where

$$p_{ij} = \frac{Q_{ij}}{L_{ij}} \qquad \text{Eq. 19}$$

6- Calculate the bounds for offset values that avoid starvation and spillback according to the (p_{ij}) ratio from Step 5.

For the main approach

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

$$\Delta_{min} > \left(\frac{L_{ij}}{u_s}\right) - \frac{L_{ij}\left(1 - p_{ij}\right)}{L_v}.h \qquad \text{Eq. 20}$$

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

The splits and offsets for the network are then calculated using an optimization procedure that protects critical routes from both spillback and starvation by minimizing the degree of saturation on all approaches on the critical route for the given cycle time. This procedure's inputs and outputs is illustrated in Figure 40.

Optimization Problem Formulation

Minimize the degree of saturation on all approaches on the critical routes

$$Min \quad \sum_{r \in R} \sum_{i \in I} \sum_{j \in J} \left({v_{ij}^r}/{c_{ij}^r} \right) \qquad \text{Eq. 21}$$

Subject to the following constraints:

Cycle length constraint

$$\sum_j (g_{ij} + Y_{ij}) = C \qquad \text{Eq. 22}$$

Queue lengths constraints (storage capacity):

$$Q_{ij} < L_{ij} \qquad \text{Eq. 23}$$

Degree of saturation constraints on the critical routes, such that:

$$v_{ij}^r / c_{ij}^r \leq 1 \qquad \text{Eq. 24}$$

Queue-link ratio constraints (queue management):

$$p_{ij}^r < \gamma \qquad \text{Eq. 25}$$

Minimum greens constraints:

$$g_{ij} \geq g_{min} \qquad \text{Eq. 26}$$

Maximum greens constraints:

$$g_{ij} \leq g_{max} \qquad \text{Eq. 27}$$

For all critical routes in the network:

$$\forall\, r \in R, \forall\, i \in I, \ \ \forall j \in J \qquad \text{Eq. 28}$$

Where:

C: cycle length (s)
c: capacity (veh/hr)
v: volume (veh/hr)
g: effective green (s)
Q_b: size of initial queue (ft)
u: delay parameter
s_i : adjusted saturation flow rate per lane of approach (i), veh/s

$d_{(k)I}$: mean departure rate from approach (i) during cycle k, veh/s
$q_{(k)I}$: length of queue on approach (i) at the beginning of cycle k, veh
$g_{(k)I}$: effective green time for arterial approach (i) during cycle k, s
$gc_{(k)I}$: effective green time for cross approach (i) during cycle k, s
$AV_{(k)j}$: arriving vehicles into approach (i) during cycle k, veh
V_{ij}: Approach (j) volume at intersection (i) (vehicle per hour)
g_{ij}: Approach (j) green time at intersection (i) (sec)
g_{min}: Minimum green split (sec)
$(v/c)_{ij}$: v/c ratio of approach (j) at intersection (i)
Q_{ij}: Queue length of approach (j) at intersection (i) (ft)
p_{ij}^{r}: Queue-link ration for approach (j) at intersection (i) for the critical route (r)
γ_{ij}^{r}: Maximum value for the queue-link ratio for approach (j) at intersection (i) for the critical route (r)
L_{ij}: Storage capacity of approach (j) at intersection (i) (ft)
$\forall\, r \in R$: For each critical route (r) in the network's set of routes (R)
$\forall\, i \in I$: For each intersection (i) in the network's set of intersections (I)
$\forall j \in J$: For each approach (j) in the intersection's set of approaches (J)

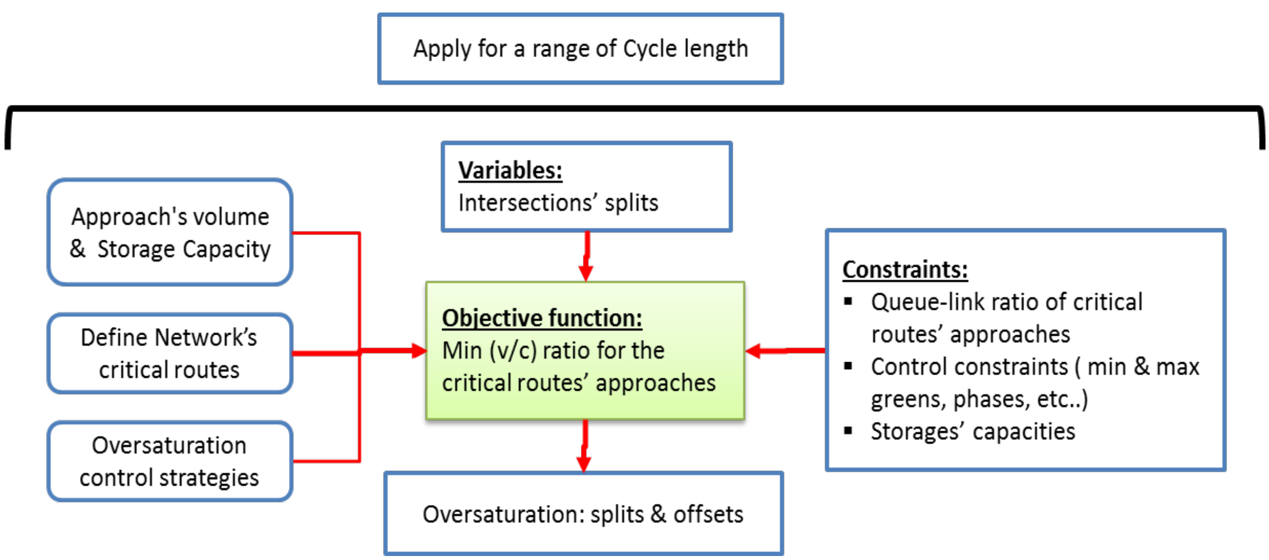


Figure 40. Split-offset optimization framework

Construction of the Pareto Front

Multi-objective evolutionary algorithms are used to examine the effectiveness of alternate strategies for oversaturated intersections. A multi-objective approach allows for the optimization of several objectives simultaneously. Unlike traditional methods of assigning pre-defined weights to each objective function, multi-objective evolutionary algorithms produce Pareto fronts. A Pareto front is the combination of compromise solutions of all of the objectives being considered in the problem. Figure 41shows two objectives (for the purpose of illustration, any number of objectives can be used): throughput and delay. In this example, each point on the Pareto front would correspond to a complete set of timing plan parameters (e.g., the cycle, splits, and offset).

Consider, for example, that a particular control strategy for oversaturated conditions was successful in reducing both system-wide delay and increasing throughput. A control action “X” (e.g., increasing major phase duration) can further increase throughput, where another control action “Y” (e.g., left-turn reservice) can further reduce the total delay. Both points A and B in Figure 41 correspond to optimal solutions, but with different importance assigned to each objective. Point A for example, corresponds to a weight of 100 percent assigned to the objective of increasing throughput, and 0 percent assigned to the objective of minimizing delay. Point B corresponds to a 0 percent weight assigned to maximizing throughput and a 100 percent weight assigned to minimizing total delay.

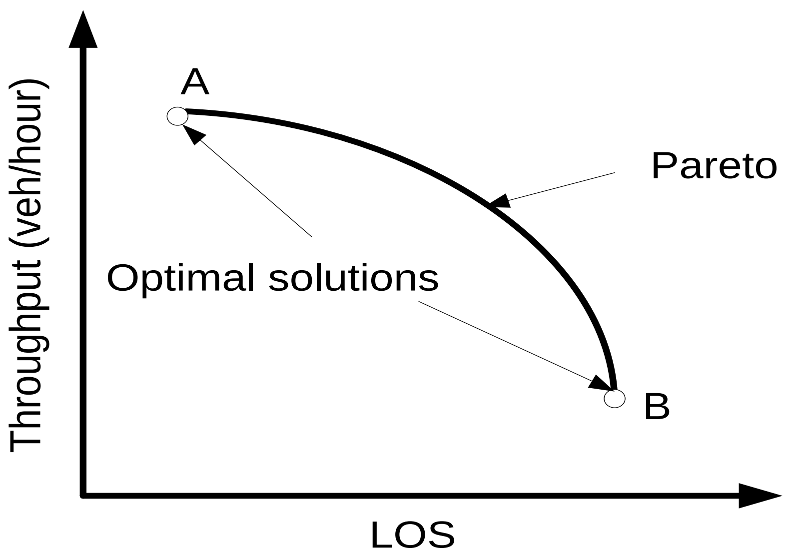
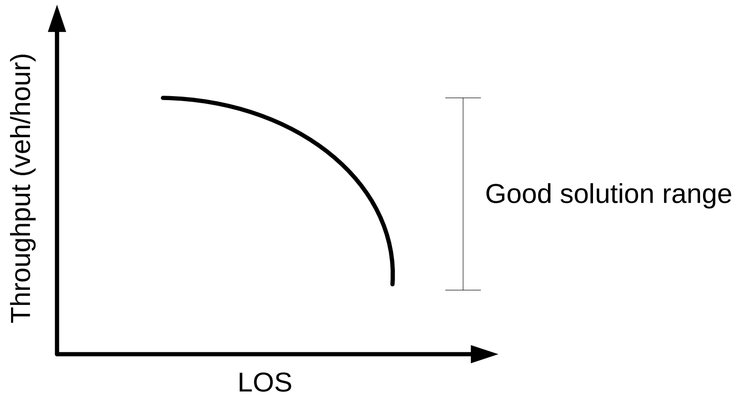
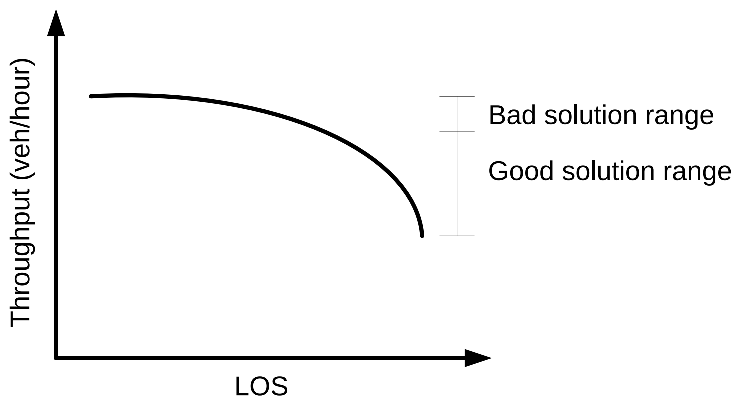


Figure 41. Conceptual illustration of Pareto front in assessing multiple objectives

In addition, the *shape* of the Pareto front itself provides invaluable information to the analyst. One would know, looking at the shape of the Pareto front, how much an objective function would be compromised if another objective function were to be favored. For example, in Figure 42a, the engineer can choose any solution that lies in the Pareto front, where in Figure 42b, the bad solution range is labeled as such because any additional increase in throughput will result in a very large increase in delay.



(a)



(b)

Figure 42. Information provided by the shape of the Pareto Front (a) Pareto front of unrestricted solution range (b) Pareto front of restricted solution range

This approach differs from traditional methods used in signal timing optimization because it allows for the consideration of multiple objectives simultaneously. The assignment of weights to each individual objective can be done at the end of the analysis as opposed to assigning arbitrary weights to different objectives without knowing the consequences of that assignment beforehand. The use of multi-objective optimization is a key element of the research approach since there is no easy way to quantify when a system should be transitioned from an objective of delay minimization to one of queue management or maximizing throughput.

Control Objectives

There are three major management objectives considered in this research: (1) delay minimization, (2) throughput maximization, and (3) queue management. We will briefly discuss each from a qualitative perspective and then present the mathematical formulation used in the optimization methodology for each objective.

Delay Minimization

Most offline optimization tools use some kind of formulation for minimizing delay and stops, perhaps balanced with some consideration for providing progression on an arterial route. In undersaturated conditions, this objective is handled sometimes quite loosely to perhaps include solutions that might only be considered to be effective or only acceptable and not really a true minimum total delay. This includes (a) minimizing the delay at a single intersection and (b) minimizing delay in a network or series of intersections on a travel route (progression).

Intersection delay might be characterized in one of two manners. The first is what might be referred to as the classical Webster's method which is to allocate green time to phases to minimize the total delay at the intersection given known demand for all movements. Since demand fluctuates and is not directly measurable with traditional detection systems, the second and more commonly used interpretation of the objective minimize delay is to minimize the frequency that a traffic signal does not serve all the waiting cars during the green time. In other words, minimize delay means to minimize phase failures in actuated-coordinated signal systems. The former is minimization of system delay. The latter is minimization of user delay.

Minimization of user delay is the traditional basis for most actuated-coordinated signal timing in North America. Green splits are typically allocated to ensure that side streets and left-turn phases have a cushion of additional time in case they need more than they typically need on average to minimize cycle failures. In most cases this additional time is re-allocated to other phases when the phase gaps out due to lack of additional demand. Depending on the use of fixed or floating force offs, the extra time is either returned to the coordination phase or the next phase in the sequence is given additional time.

Avoiding cycle failures is an equitable traffic management policy; one that over-emphasizes the importance of light traffic movements such as left turns and side streets. It does not minimize total delay; it rather minimizes each driver's perception of being delayed at the signal. In undersaturated conditions, this policy can be improved upon by adaptive traffic control strategies, but in most low and medium flow scenarios, a cycle-failure avoidance policy is hard to beat as it meets the objective to provide a consistent user experience for drivers.

However, when green times are not long enough to serve all the vehicles that were waiting at the start of the green time, policies with fixed split times have no way to react to these changes. Even if the arrival rate remains constant but is still higher than the service rate (maximum green time), the overflow queue will continue to grow for that phase. Objectives that continue to consider equity service at the intersection will over-emphasize the minor movements at the detriment to the movements and phases that are oversaturated.

Progression is an objective that blurs the lines between minimizes delay and maximize throughput. Progression is achieved in signal systems by arranging for the green time windows to be consecutively opened (by way of setting offsets) in a desired direction of travel to allow vehicles to continue through a sequence of intersections without stopping. By carefully setting the offset values, the objective of minimizing delay (equity treatment for all users) can still be satisfied at individual intersections while at the same time meeting the system objective of progression and providing the consistent user experience that drivers' tend to expect on arterial roads. This objective is hindered when overflow queues begin to form on the movements that are designated for progression. Offsets that were designed (i.e. forward progression offsets) assuming that no queues were present will further exacerbate the situation by creating situations where the overflow queues grow faster than necessary because the upstream vehicles arrive before the overflow queue has started to move.

Throughput Maximization

Minimizing user delay is simply not appropriate when the situation is oversaturated and it is no longer possible to avoid cycle failures. Thus, maximizing the number of vehicles actually served by the intersection, with respect to the vehicles presented to the intersection (the load), is a more appropriate objective. This keeps as much of the system operational as possible, with the unfortunate effect of delaying movements or phases where the total traffic demand is quite low. From an equity perspective, strategies that maximize throughput might be considered to punish light movements to the benefit of the greater good. This is done by moving much heavier phases for longer amounts of time more frequently than would be expected by the typical actuated control approach.

Strategies that maximize throughput have the following characteristics:

- Make best use of the physical space (e.g., lag heavy left turns, run closely-spaced intersections on single controller).
- Make best use of green time in the cycle (e.g., prevent actuated short greens, separate congested movements from the uncongested ones, phase re-servicing).
- Reduce the negative impact of other influences (e.g., buses, pedestrian movements) on the overall ability of the signal system to process vehicle flows.

Measurement and Assessment of Throughput

There are several ways to measure or assess the throughput of a traffic signal network or system:

- The total number of vehicles input to a system of intersections.
- The total number of vehicles output by a system of intersections.
- The ratio of the total vehicles output from the system to the total vehicles input to the system.

Throughput is a rate such as vehicles per hour or vehicle-miles per hour. The concept of an input rate and an output rate must be considered together with the identification of the spatial extent of the system of intersections of interest. When a control strategy is operating efficiently, the overall output processing rate of the intersections in the system closely matches the input processing rate and overflow storage of vehicles in the system does not occur. In oversaturated conditions, the output rate is less than the input rate and overflow queuing begins to build up within the system at various points. At some point in many oversaturated systems, queues will build outside of the system cordon boundaries when queues inside of the system reduce the ability for those vehicles from entering the area.

When a mitigation strategy increases both the total output and the total input rates, it can truly be determined to have increased the capacity of the traffic control system. If the mitigation strategy increases total system input, but not output, then it uses more of the available spatial capacity of the system. This in itself is a valuable performance improvement and is desirable during the loading and processing regimes. Similarly, if a mitigation strategy increases total system output, but not input, it is reducing the congestion and oversaturation inside the system cordon line. This is preferable during the recovery regime.

Queue Management

The goal of throughput maximization strategies is increasing input, increasing output, or both. At some point, however, no further revision to the signal timing will increase maximum throughput and queues will continue to grow until demand diminishes. The reason that strategies largely have the same performance during the peak time is that the queues are so pervasive that the cause-and-effect relationships between control actions and the traffic situation are masked by hysteresis.

Hysteresis is a delay between an offered input and the system output. For example, when the green time of a downstream intersection can process only one-third or one-quarter of the upstream queue, it is no longer that important if the offset is set for positive or negative progression. When the light turns green at the upstream intersection, there is limited storage for the entering traffic and spillback begins to create SOSI > 0 at the upstream location. The one-to-one dynamic of the offset relationship from one intersection to another is no longer applicable. The shorter the link distances between intersections, the faster the system can quickly degrade from stable operation to pervasive queues, thereby skipping any potential improvement that a throughput maximizing strategy could have achieved.

This usually requires constraining capacity upstream from a bottleneck at locations where queue storage will not cause network gridlock, or increasing green time at downstream signals to increase output flow. As noted by Denney, 2008, if throughput maximization strategies are a curative approach, then queue management strategies can be considered a palliative approach with the objective of treating symptoms rather than seeking a cure.

It is in this context that the commonly held belief that “there is nothing that can be done, there is simply too much traffic” is mostly true. Synchronizing the actions of multiple controllers in a system of intersections for the purpose of queue management is very difficult within the context of actuated-coordinated control by commanding patterns with different parameters. The coordination of actions between intersections for queue management must closely resemble the operation of a diamond interchange with a carefully orchestrated sequence of actions, with rapid feedback between the detection of queue extent and the application of rapid-response mitigation strategies such as phase truncation and green extension. Design of timing plans according to the principles described in Step 4 and adjustment of green times according to the technique described in Appendix B can begin to address these situations.

Formulation for Delay Minimization Objective

To compare the performance of timing plans optimized for delay versus throughput versus queue management, we need to first formulate optimization objectives for all three goals. The delay minimization objective is represented in this process by choosing green splits throughout the network that minimize the total delay based on the HCM methodology (Akcelik, 1988). The HCM control delay formula accounts for random and platooned arrival of vehicles as well as progression quality and delay resulting from pre-existing queues. The minimization process is constrained by a feasible cycle length range, minimum green, maximum green, queue storage limits, and v/c ratios on approaches that are part of critical routes. Unlike other timing programs, the developed timing tool generates an optimal timing plan for each time period in the volume profile (e.g., 15-min). Moreover, for approaches with v/c > 1, un-served volumes are shifted to the next period volume (i.e., volume spillover) to represent the temporal effect of un-served demand.

Optimization formula:

$$\text{Min (Delay} = d_1 + d_2 + d_3) \qquad \text{Eq. 29}$$

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

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

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

Subject to the following constraints:

Cycle length

$$C_{min} < C < C_{max} \qquad \text{Eq. 30}$$

Cycle length constraint

$$\sum_j (g_{ij} + Y_{ij}) = C \qquad \text{Eq. 31}$$

Queues lengths constraints (storage capacity):

$$Q_{ij} < L_{ij} \qquad \text{Eq. 32}$$

Critical routes movements degree of saturation constraints:

$$v_{ij}^{r} \Big/ c_{ij}^{r} \leq 1 \qquad \text{Eq. 33}$$

Minimum greens constraints:

$$g_{ij} \geq g_{min} \qquad \text{Eq. 34}$$

Maximum greens constraints:

$$g_{ij} \leq g_{max} \qquad \text{Eq. 35}$$

For all critical routes in the network:

$$\forall\, r \in R, \forall\, i \in I, \ \ \forall j \in J \qquad \text{Eq. 36}$$

Throughput Maximization

The throughput maximization objective is represented in this approach by optimizing green splits throughout the network in order to maximize the objective developed by Abu-Lebdeh (2001). This objective was designed to obtain signal control so that system throughput is maximized subject to constraints on state and control variables (i.e., green times, and offsets) such that no *de facto* red is formed. Offsets are set to ensure continuity of movement and green times on the critical route(s) are within specified ranges and have v/c ratio less than 1.

Optimization formula:

$$Max\left(\sum_k \sum_i gc_{(k)i} d_{(k)i} + j \times \sum_k Min\left(q_{(k)i} + \frac{AV_{(k)j}}{g_{(k)j-1}} \times \left(g_{(k)j} - \frac{q_{(k)j}}{s_j} \right), AV_{(k)j} + q_{(k)j} \right) \right) \qquad \text{Eq. 37}$$

Subject to the following constraints:

Cycle length

$$C_{min} < C < C_{max} \qquad \text{Eq. 38}$$

Cycle length constraint

$$\sum_j (g_{ij} + Y_{ij}) = C$$ Eq. 39

Queues lengths constraints (storage capacity):

$$Q_{ij} < L_{ij}$$ Eq. 40

Critical routes movements degree of saturation constraints:

$$v^r_{ij} / c^r_{ij} \leq 1$$ Eq. 41

Minimum greens constraints:

$$g_{ij} \geq g_{min}$$ Eq. 42

Maximum greens constraints:

$$g_{ij} \leq g_{max}$$ Eq. 43

For all critical routes in the network:

$$\forall\, r \in R, \forall\, i \in I,\ \ \forall j \in J$$ Eq. 44

Where:

d_1: Uniform delay
d_2: incremental delay
d_3: delay due to pre-existing queues (for under saturation $d_3 = 0$)
C: cycle length
c: capacity
v: volume, and
g: effective green
Q_b: size of initial queue
T: analysis period, hour
t: duration of oversaturation within T, h
u: delay parameter
s_i : adjusted saturation flow rate per lane of approach (i), veh/s
$d_{(k)I}$: mean departure rate from approach (i) during cycle k, veh/s
$q_{(k)I}$: length of queue on approach (i) at the beginning of cycle k, veh
$g_{(k)I}$: effective green time for arterial approach (i) during cycle k, s
$gc_{(k)I}$: effective green time for cross approach (i) during cycle k, s
$AV_{(k)j}$: arriving vehicles into approach (i) during cycle k, veh
V_{ij}: Approach (j) volume at intersection (i) (vehicle per hour)

g_{ij}: Approach (j) green time at intersection (i) (sec)
g_{min}: Minimum green split (sec)
$(v/c)_{ij}$: v/c ratio of approach (j) at intersection (i)
L: Link length (ft)
Q_{ij}: Queue length of approach (j) at intersection (i) (ft)
p_{ij}^{r}: Queue-link ration for approach (j) at intersection (i) for the critical route (r)
γ_{ij}^{r}: Maximum value for the queue-link ratio for approach (j) at intersection (i) for the critical route (r)
L_{ij}: Storage capacity of approach (j) at intersection (i) (ft)
$\forall\, r \in R$: For each critical route (r) in the network's set of routes (R)
$\forall\, i \in I$: For each intersection (i) in the network's set of intersections (I)
$\forall j \in J$: For each approach (j) in the intersection's set of approaches (J)

Queue Management Control

The queue management control objective is achieved primarily by generating green splits that minimize the degree of saturation of pre-determined critical routes in the network. Critical movement's (v/c) ratios and links' p-ratios (i.e., queue-to-link ratio) upper bound are set according to the control strategies that will be applied (e.g., metering, gating, green flaring).

Optimization formula:

Minimize the degree of saturation on approaches on the critical routes

$$Min \quad \sum_{r\in R} \sum_{i\in I} \sum_{j\in J} \left({v_{ij}^{r}}/{c_{ij}^{r}} \right) \qquad \text{Eq. 45}$$

Subject to the following constraints:

Cycle length constraint

$$\sum_{j} (g_{ij} + Y_{ij}) = C \qquad \text{Eq. 46}$$

Queue lengths constraints (storage capacity):

$$Q_{ij} < L_{ij} \qquad \text{Eq. 47}$$

Critical routes movements degree of saturation cosnstraints:

$${v_{ij}^{r}}/{c_{ij}^{r}} \leq 1 \qquad \text{Eq. 48}$$

Queue-link ratio constraints (Queue management):

$$p_{ij}^{r} < \gamma \qquad \text{Eq. 49}$$

For all critical routes in the network:

$$\forall\, r \in R, \forall\, i \in I,\ \ \forall j \in J \qquad \text{Eq. 50}$$

Where:

C : cycle length
c : capacity
v : volume, and
g : effective green
Q_b: size of initial queue
T : analysis period, hour
t : duration of oversaturation within T, h
u : delay parameter
s_i : adjusted saturation flow rate per lane of approach (i), veh/s
$d_{(k)I}$: mean departure rate from approach (i) during cycle k, veh/s
$q_{(k)I}$: length of queue on approach (i) at the beginning of cycle k, veh
$g_{(k)I}$: effective green time for arterial approach (i) during cycle k, s
$gc_{(k)I}$: effective green time for cross approach (i) during cycle k, s
$AV_{(k)j}$: arriving vehicles into approach (i) during cycle k, veh
V_{ij}: Approach (j) volume at intersection (i) (vehicle per hour)
g_{ij} : Approach (j) green time at intersection (i) (sec)
g_{min} : Minimum green split (sec)
$(v/c)_{ij}$: v/c ratio of approach (j) at intersection (i)
Q_{ij} : Queue length of approach (j) at intersection (i) (ft)
p_{ij}^{r} : Queue-link ration for approach (j) at intersection (i) for the critical route (r)
γ_{ij}^{r} : Maximum value for the queue-link ratio for approach (j) at intersection (i) for the critical route (r)
L_{ij} : Storage capacity of approach (j) at intersection (i) (ft)
$\forall\, r \in R$: For each critical route (r) in the network's set of routes (R)
$\forall\, i \in I$: For each intersection (i) in the network's set of intersections (I)
$\forall j \in J$: For each approach (j) in the intersection's set of approaches (J)

Development and Analysis of Timing Plans for Managing Oversaturated Conditions

The procedure described in this chapter is designed to produce a range of potential timing plans that are optimized for different objectives. Those timing plans are then compared with each other based on the performance measures for which they were or were not optimized. This allows the analyst to use Pareto analysis to determine those timing plans that are non-dominated.

In addition, the methodology considers the time-varying nature of the traffic demands. Particularly for oversaturated conditions, the process of generating the timing plans must consider that any un-served demand must be serviced in the next time period. Any un-served demand in that time period will spill over to the next time period, and so on.

Similarly, since the oversaturated problem is dynamic in nature, the process explicitly considers that multiple timing plans will be needed. The methodology allows the results of the analysis to indicate where it is beneficial to switch from one timing plan to another, which also indicates in general where the objectives of the strategy may change from minimizing delay to maximizing throughput, or managing queues. These considerations are discussed further in the following sections.

Using Volume Profiles on Critical Routes in the Design of Signal Timing Plans

Traditionally, a signal timing plan is generated based on the highest observed volumes of the conflicting movements. It is assumed that this condition is the worst possible condition that the timing plan will experience, and thus all other combinations of traffic volumes will have at least as good, if not better, performance using the same timing parameters as designed for the limiting case. However, during oversaturated conditions, generating a timing plan using the highest volumes does not account that approach capacities are limited, and the critical route flows and demand rates will change over time. Observing a change in the volume ratios is a good indication of the need for a new signal timing plan to accommodate the new demand. Periods when a new signal timing plan may be warranted are identified by the changes in either the total traffic demand or demand for a particular phase split. As volume profiles of conflicting movements vary from one time period to another, optimal splits will change accordingly. This is well known, but is not typically considered in optimization for undersaturated conditions because the causal linkages between the two traffic conditions are relatively weak since overflow queues are not generated. Therefore, the methodology described here generates several timing plans for each time period and compares the performance of applying the different timings during time periods where the traffic demand is much different than the time period for which it was designed. This is illustrated conceptually in Figure 43.

The y-axis represents traffic demand for each conflicting movement. In this example, only northbound and westbound traffic flows are shown. The x-axis represents time and is divided into 11, 15-minute time periods. The timing plan that was optimized for time period 5 results in timing parameters where green allocation to the phases serving northbound and westbound traffic are relatively equal. However, the timing plan that was optimized for time period 9 results in the westbound volume being more than double that of the northbound volume. Therefore the timing plan parameters would be very different from the timings optimized for period 5. This illustrates the necessity of designing a set of timing plans that accommodate the fact that the profile of the future traffic is changing dramatically. For example, it may be beneficial over the long run to have started running the timing plan optimized for the conditions in time period 9 during time period 5, instead of allowing the westbound queues to grow until the timings for time period 9 are finally implemented. These kind of trade-offs are analyzed and brought to light by the multi-objective Pareto front analysis process.

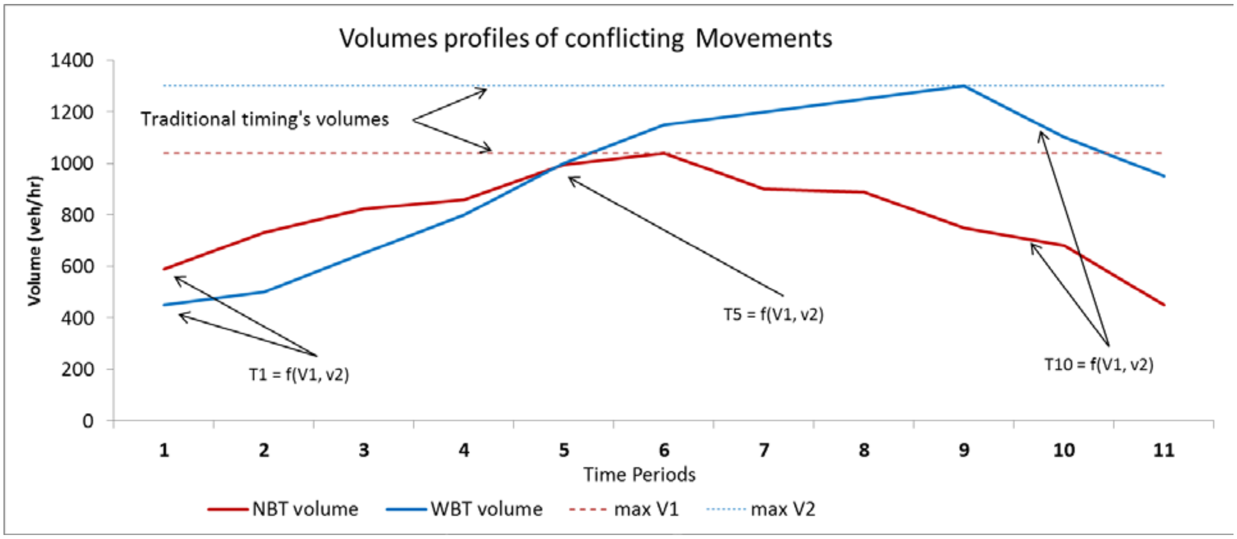


Figure 43. Two conflicting movement volume profiles

Explicit Consideration of Volume Spillover

As discussed above, the temporal qualities of the volume profile on the critical routes are essential to take into consideration. This methodology explicitly takes into consideration the fact that any un-served volume for each time period will be added to the arrival volume in the next time period. The volumes on several approaches during oversaturated conditions would be expected to exceed their capacities (green splits). This concept is illustrated in Figure 44. As the red line (demand) exceeds the green line (capacity), that latent demand is stored outside of the system and must be served at a later time period. The black line indicates how the un-served volume builds up and then dissipates over time. The application of this concept ensures that optimal timing plans are evaluated through the entire volume profile and that the new traffic state created due to using a specific plan is used in the evaluation or update of timing plans.

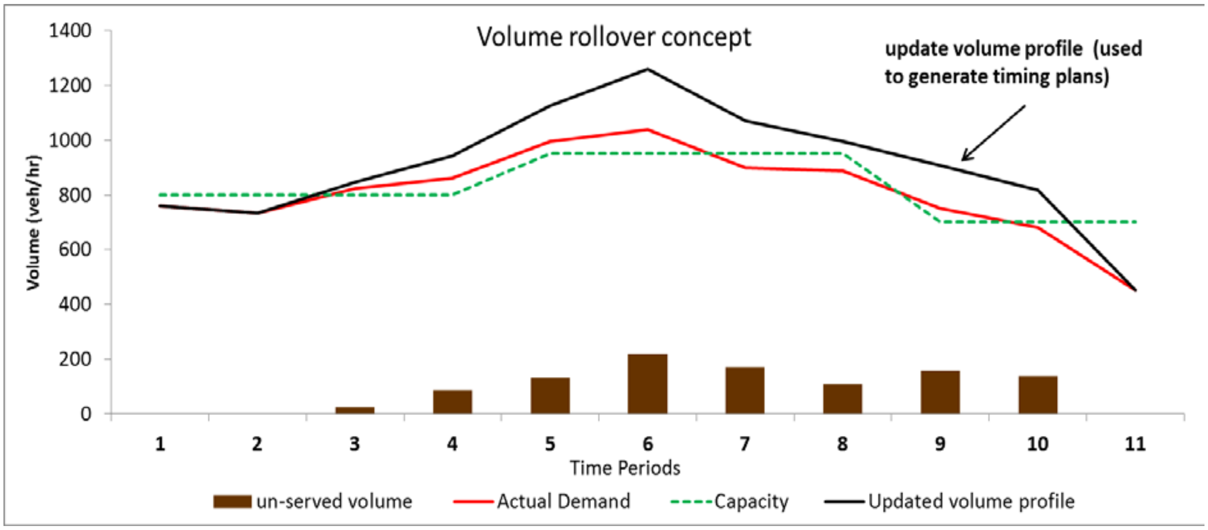


Figure 44. Concept of volume profiles that generate un-served demand

Optimization Procedure

The optimization problems for minimizing delay, maximizing throughput, and managing queues are mixed integer nonlinear problems with nonlinear constraints. The Generalized Reduced Gradient (GRG) algorithm, one of the methods provided by Excel-solver, is used to search the timing plan space and find the optimal timing plans. The GRG method is a robust nonlinear programming method that uses a hill-climbing iteration process. The GRG method has been proven to be one of the most efficient and effective methods for Nonlinear Programming problems with Nonlinear constraints. The basic concept of GRG method entails linearizing the nonlinear objective and constraint functions at a local solution with a Taylor expansion. Then, the concept of reduced gradient method is employed which first divides the variable set into two subsets of basic and non-basic variables. Next the GRG method employs the concept of implicit variable elimination to express the basic variables as functions of the non-basic variables. Finally, the constraints are eliminated and the variable space is reduced to only non-basic variables. This approximated problem is used to search for the optimal solution. The process is fairly efficient and allows reasonably responsive problem resolution using common tools (Microsoft Excel).

Performance Measure Evaluation for the Generated Optimal Timing Plans

The optimization tool generates several optimal timing plans. These timing plans are generated based on the control objective (either total delay or total throughput) and the assumed volume profiles on each movement. The concept of volume spillover as described earlier is used to optimize a sequence of timing plans for each objective function. Response surfaces are then created for each performance measure that shows how each timing plan would perform if that timing plan were applied to the entire volume profile. Figure 45 shows the response surface for delay and Figure 46 for throughput.

The surface for total delay uses the HCM equation to compute the performance of the timing plan for each time period. Similarly, the surface for total throughput uses the Abu-Lebdeh (2000) objective function to compute the performance. The x-axis of the surface represents the 11 time periods of the scenario. The y-axis represents the performance measure. The z-axis lists the timing plans that were generated from the optimization process for each optimization objective. Timing plans P1 through P8 represent plans obtained from optimizing total delay. Timing plans P9 through P17 represent the plans obtained from optimizing for total throughput.

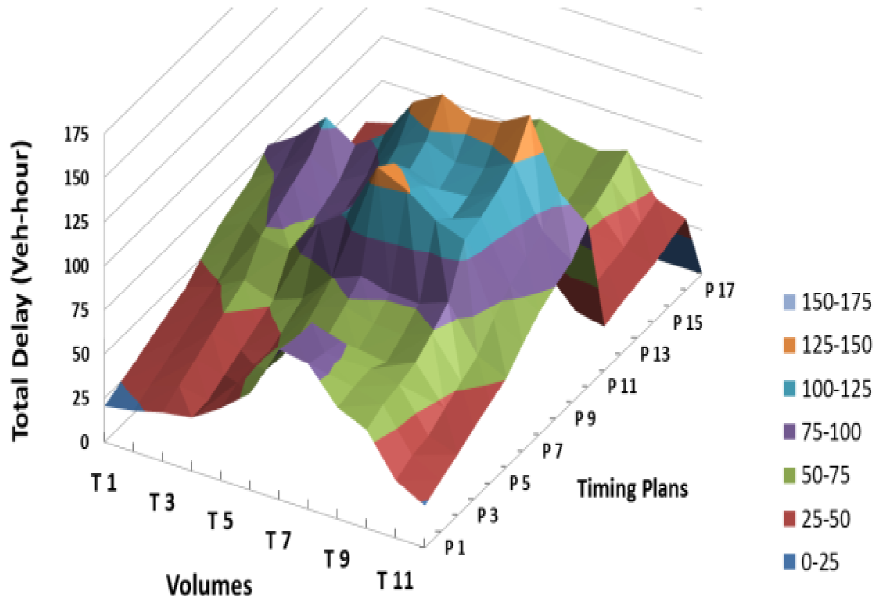


Figure 45. Delay surface representing the performance of each timing plan for each time period

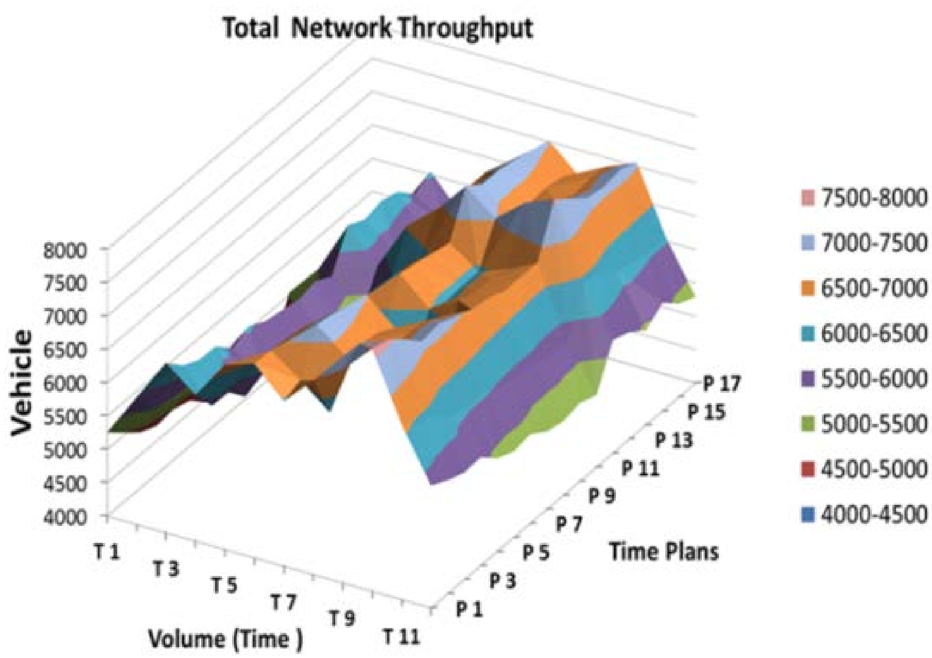


Figure 46. Throughput surface representing the performance of each timing plan for each time period

Switching Between Control Strategies

The decision to switch between control strategies is based on several factors:

1. The sensing/identification that the traffic scenario has switched from one canonical state to another, and

2. The mapping of the identified scenario to the appropriate strategy based on the recommended objective for that particular degree of saturation.

These decisions thus must be based on a robust mechanism to appropriately switch from one objective to another in order to identify which timing plan is more suitable. The mechanism used in this research is based on robust pattern recognition techniques such as the Bayesian pattern recognition approach or discriminant analysis (Abbas and Sharma, 2006).

Thus, a second optimization procedure can be applied that will result in the optimal timing plans and the optimal time that the plans should be switched from one to another to adjust the overall objective of traffic management from one objective to the next. Based on the general concept of an oversaturated scenario having three regimes of operation (loading, processing, and recovery), we defined two canonical sets of objectives for testing and evaluating this approach. Two canonical control strategies were developed and evaluated in this research:

- Delay minimization -Queue management -Throughput maximization

- Throughput maximization -Queue management -Throughput maximization

A timing plan that is optimal for the first objective would be applied during the loading regime. A timing plan that is optimal for the second objective would be applied during the processing regime. A timing plan that is optimal for third regime would be applied during the recovery period. Identifying the times at which it is optimal to switch from one objective to another is the goal of applying the pattern recognition methodology. This methodology requires running many combinations of the timing plans in simulation and observing the results. In addition to the duration constraints for each regime of the peak period, a minimum duration of each strategy is considered. This is because when a new timing plan is selected, the controller must transition to the new plan. During this transition period, systematic operation in the network is interrupted while each controller adjusts to the new parameters. Because of this disruption, transitions are minimized and one must be assured that the benefits of the new plan overcome the dis-benefits of the transition period.

Figure 47 illustrates the total input, total output, and resulting total vehicles in the system for a test network under the throughput-queue management-throughput strategy. The blue line roughly depicts the capacity of the plans that are applied during each regime. Similarly, Figure 48 illustrates the total input, total output, and resulting total vehicles in the system for a test network under the delay-queue management-throughput strategy. Note in Figure 47how the total output is higher using a plan that is designed for throughput maximization than the total output in Figure 48 for a plan that is designed for minimizing total delay. The resulting performance is quite different in the two cases during the processing regime as the total vehicles in the system jumps to over 2,800 vehicles when starting with delay minimization and maxes out to only 2,000 vehicles in the system using throughput maximization in the loading regime. Also, when using a minimizing delay strategy during the loading regime the throughput maximizing strategy during the recovery regime must have much higher capacity to process the additional queues that are stored in the system in a timely manner.

Two examples of the application of this procedure on test networks are reported in Chapter 3 for the Reston Parkway in Northern Virginia and the Post-Oak area of Houston.

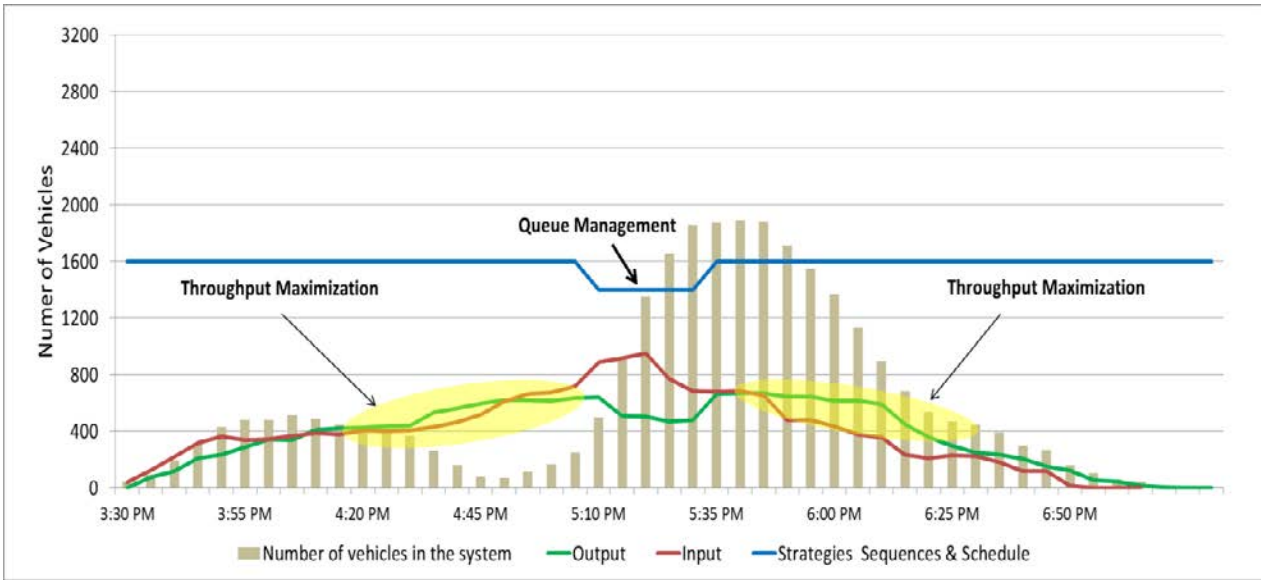


Figure 47. Example of optimal timing plans and scheduling based on the minimum delay objective

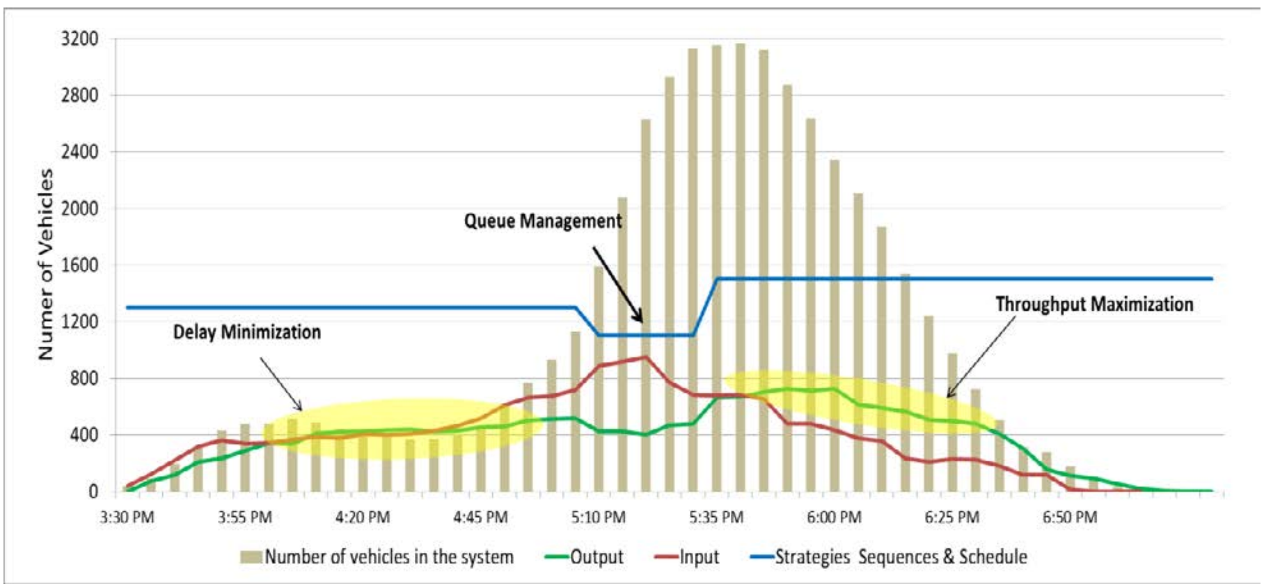


Figure 48. Example of optimal timing plans and scheduling based on the throughput – maximization objective

Summary

There are four essential parts to the design of timing plans for handling oversaturated conditions:

1. the identification of the critical routes and the flows along those routes,
2. the determination of ranges of feasible timing plan parameters that meet criteria to address oversaturation on those routes,

3. the selection of the objective function to solve for optimal timings with respect to a particular objective, and
4. the scheduling of a sequence of timing plans that are tailored to each regime of oversaturated operation.

In this part of the research project we developed a comprehensive process that addresses these four elements. The process is non-trivial and should be considered experimental at this time. Significant additional research and development will be necessary to transfer this type of approach from state-of-the-art to state-of-the-practice. The process of developing the procedures and testing the results provided a wide range of insights into the critical components of handling oversaturated conditions with fixed-value traffic signal timing plans and traditional time-of-day scheduled responses.

The process begins by identifying critical oversaturated routes in a network of interest. The procedure is designed to find timing plans that will minimize problematic symptoms (spillback, starvation, and storage blocking) on the critical route(s). There are wide range of timing (cycle, split, and offset) combinations that can satisfy these goals. Thus a second step is to apply an optimization procedure that searches the range of feasible combinations of settings and addresses a particular objective. Three objectives were tested to evaluate the theory (or generally held assertion) that different objectives are needed in different regimes of operation. The third step is to identify when to switch from one regime of operation to the next. This is done with a simulation-in-the-loop, multi-objective procedure. Test cases for this approach are reported in Chapter 3 and findings from the tests are reported in Chapter 4: Conclusions.

Online Implementation of Mitigation Strategies

In the previous two sections we described quantitative measures for estimating the severity of an oversaturated condition and a process for generating timing plans that are designed to mitigate oversaturated conditions. When the oversaturated scenario is recurrent, either of these methods can be used to generate mitigation timing plans and apply those timing plans on a TOD scheduled basis.

However, when the scenario is non-recurrent, mitigation strategies should be implemented that use the status of field detectors to determine when to implement the mitigation. This section of the report describes an online processing tool that can be used to implement non-recurrent strategies. Once the mitigation timing plans are designed offline and loaded into field controllers, logic rules based on field detection can be employed to react to the oversaturation.

To implement a mitigation strategy online, detection zones must be deployed, those detectors must be connected to field controllers, and there must be communications from the field controllers to a central system. Existing traffic responsive logic in central systems might be configured to select among mitigation strategies and normal operating timing plans. Central systems that use the UTCS/USDOT traffic responsive logic should be configured such that "K" values in V+KO calculations are sufficiently large so that the occupancy data far overshadows the volume inputs. This is critical since the volume measurement is largely unreliable and the detector data is dominated by occupancy during oversaturated conditions. Using existing traffic responsive logic components of central systems does not require high-speed communications between field controllers and central.

While detector occupancy increases during oversaturation, its value is capped at 100%. As long as the queue extends across the detection point, there is no way with occupancy alone to determine the difference between a queue of 200 vehicles from a queue of 50 vehicles. As part of the research project, experimental software was developed that can assess TOSI and SOSI and measure queue lengths. These algorithms are integrated with a logic processor tool that can use these measures in selection of mitigation strategies. This software is part of the deliverables for this project.

The key features of the logic processor is that it uses straight-forward "if…then" threshold rules to determine if a detector status condition is met or not. This can include TOSI, SOSI, queue length, and detector occupancy. These conditions can be combined for several detector stations in "AND" and "OR" logic to make more complicated decisions based on multiple detection inputs. Similarly, thresholds and "if…then" logic can be applied to determine when to *stop* applying a certain strategy.

Queue measurement and TOSI/SOSI measures provide a more accurate determination of oversaturated conditions, but it is not required that these measures be used to trigger online mitigation strategies. Traditional occupancy measurements can also be used with the logic processor application, given that the detectors are located appropriately and aggregation of the data is configured appropriately. Queue measurement and TOSI/SOSI measures requires high-resolution (i.e. second-by-second) data on phase timing and detector occupancy data to be collected and returned to the central system for processing.

Determining Detector Locations

Queue length estimation and identification of saturated occupancy is best applied with detectors that are significantly upstream of the stop bar and as short as possible (standard 6'x6' loops or zones are preferable). Utilization of occupancy data from stop bar detectors is not recommended since those detectors will report 100% occupancy during the red interval of the traffic phase, as well as during the green interval if sufficient demand is present. 100% occupancy on a stop bar detector during a cycle does not immediately indicate that the phase is oversaturated since the entire standing queue may be dissipated during the green phase. It is necessary to use detectors that are upstream of the stop bar to indicate that persistent queues are present. As stated earlier, it is also important that these detection zones are reported back to the controller on a lane-by-lane basis. Considering just one zone for a multiple lane approach will result in significant over-estimation of the level of saturation.

The balance between "too far upstream" and "too close to the stop bar" is a case of engineering judgment. The further upstream of the signal that a detection zone is situated, the longer it will take for the growing queue to occupy this detection zone for a significant portion during the cycle. If the detection zone is too far upstream, situations where it would be helpful to apply a mitigation strategy could be missed. The closer that the detection zone is located to the stop bar, the sooner the occupancy level will be close to 100% during the cycle time. If the detection zone is too close to the stop bar, the occupancy data used for decision making may result in "false positive" indications resulting in application of mitigation strategies to intermittent conditions.

In our test cases, reasonable performance was obtained with detector locations that are approximately located where dilemma zone protection or extension detectors are typically placed; between 150-500ft upstream of the stop bar or at mid-block locations as illustrated in Figure 49. Situating detection zones at approximately mid-block locations seems to be a good compromise between false-positive, false-negative, and reaction time considerations.

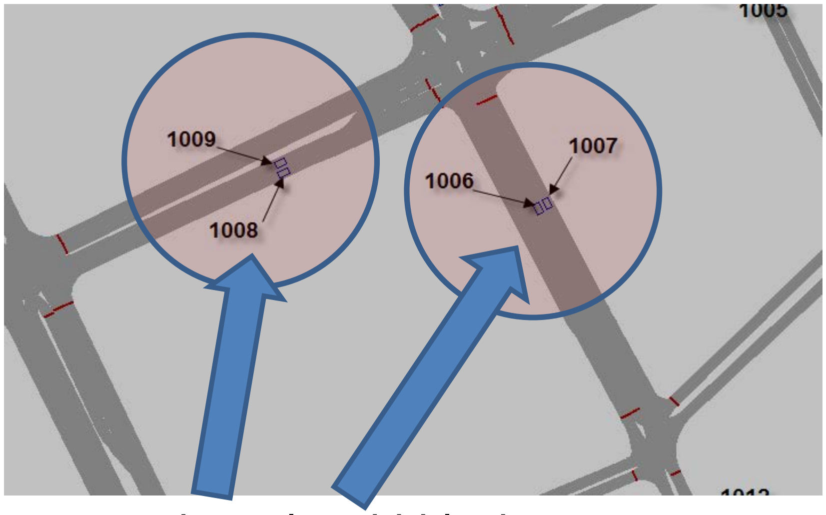


Figure 49. Recommended placement of oversaturation detection zones

Placement of Detectors for Online Recognition of a Scenario

In simple situations involving an individual approach or an individual intersection, it is straight-forward to identify where detection zones are needed since they should be placed on the approaches where the queuing is experienced. Similarly when dealing with a route or network scenario, the detection points should be placed where the queuing will determine when one or another mitigation strategy is necessary. However, when dealing with a route or network situation, it is a more subjective process to determine where the detection zones should be and which zones should be considered in decision making.

While logic conditions could be constructed that are quite complicated, in most situations, simpler logic and fewer detection points will be easier to manage and understand. Figure 50 illustrates the placement of oversaturation detectors for a common scenario. In this test case, there are three critical routes competing to access the same destination (toll booths at a border crossing). At various times of day, each route can become more critical than the others. In addition, the toll booth plaza can also become saturated to the point that no addition in flow of vehicles is possible. Thus, it is important to monitor each of the routes and select an appropriate timing plan depending on which combinations of routes are oversaturated.

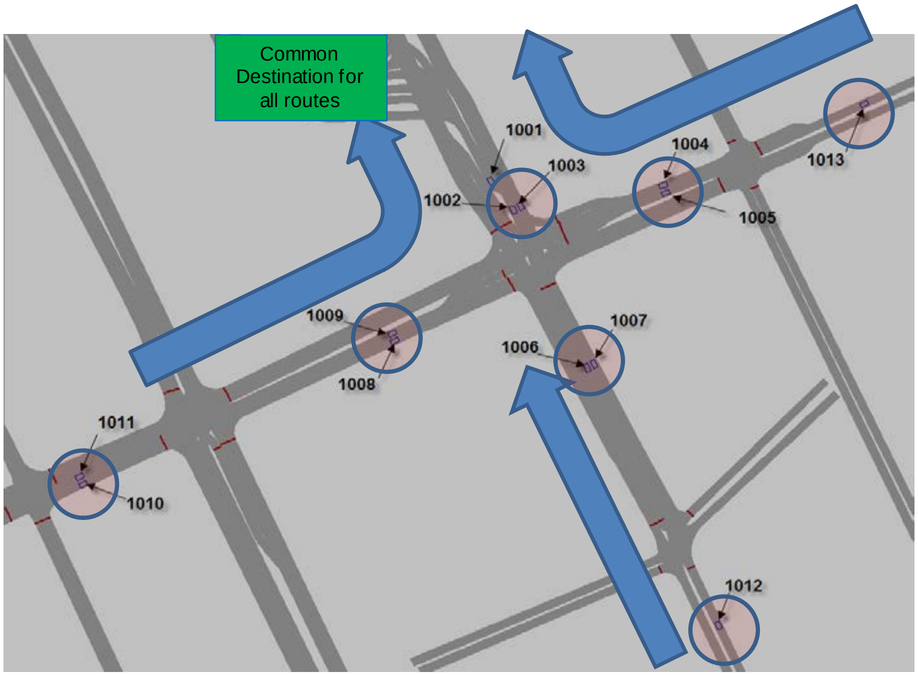


Figure 50. Example placement of oversaturation detection points

Detector Data Aggregation Intervals and Persistence Time

Aggregation of detector occupancy data is also important in balancing the occurrence of false positive and false-negative conditions. There are two important considerations here: (a) the interval over which the detector occupancy is reported and (b) the number of intervals for this condition to be true before an oversaturated condition is considered to be true. Most controller firmware allows aggregation of detector data on a minute-by-minute basis. We recommend using the lowest possible aggregation level of one minute. Persistence times of three to five minutes (larger than most typical normal cycle times) are recommended to provide responsive reaction to oversaturated conditions.

Selection of an occupancy threshold (or queue length, or TOSI/SOSI measures, if used) is also an important consideration in determining reaction time and balancing false positives and false negatives. In order to minimize the number of false-negative indications, it is important that the threshold be set relatively high. However, the threshold should also not be set at 100%. Occupancy thresholds in the 80%-90% range seem effective in providing a compromise between reacting to oversaturated conditions in a timely fashion, but not resulting in false positives.

Logic Configuration Example

Considering the case shown in Figure 51, a number of mitigation timing plans can be applied based on the status of the oversaturation detectors. There are three components to setting up the congestion response plan in the logic tool:

- Configuration of the detection points
- Configuration of the logic clauses
- Configuration of the mitigation timing plans

This setup is illustrated in Figure 51. In this example, not all of the oversaturation detectors are used as inputs.

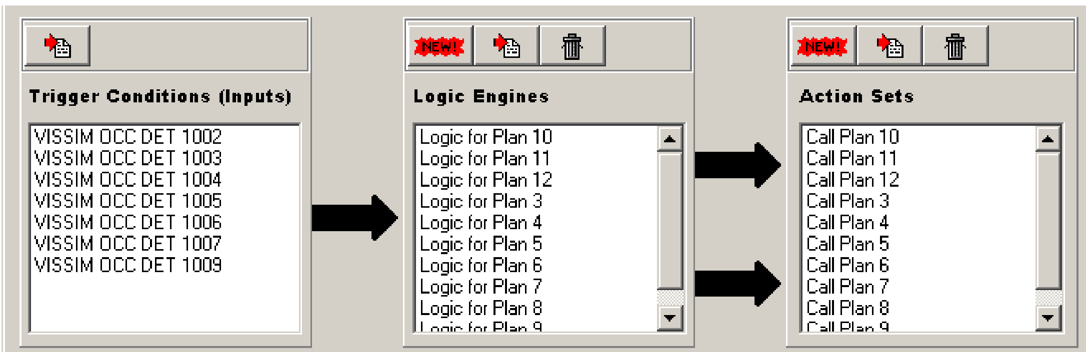


Figure 51. Set up for inputs, logic, and actions

The key component of the setup is the configuration of the logic engines. The logical clauses should consist of mutually exclusively true clauses that result in only one of the timing plans ("action sets") selected for implementation. In this test case, we constructed a logic table to determine what types of actions would be appropriate under each combination of detector conditions. This is illustrated in Table 9.

The top part of the table is the truth table. This should include all combinations of detection stations either detecting a queue or not. If detection of a queue at that point is not important, a (-) line is shown. For example, in the first column if queues are detected on the first three links, the westbound right-turn-on-red (RTOR) indication would be disallowed with a blank-out sign. The timing plan change includes the actions listed in the bottom part of the table. For example, the sixth column action indicates that the green time for the northbound and westbound phases will be increased. Thus, a new timing plan will be commanded to the field with the same cycle time but with green re-allocated from the eastbound left-turn phase to the westbound and northbound phases.

In this example, the action set consists of a timing plan change only at the key intersection. More comprehensive actions could be constructed for timing plan modifications at the adjacent intersections as well and included in the same action sets. Table 10 provides cross-reference information for the link names in Table 9 with Figure 50.

Table 9. Example logic conditions and actions

Condition	Queue detected on NB Plaza Entry Link	Y	Y	Y	Y	N	N	N	N	N	N	N	N
	Queue detected on NB Goyeau Link	Y	Y	N	N	Y	Y	N	N	Y	Y	N	N
	Queue detected on EB Left-Turn at Goyeau	Y	N	Y	N	Y	N	Y	N	Y	N	Y	N
	Queue detected on WB Right-Turn at Goyeau	-	-	-	-	Y	Y	Y	Y	N	N	N	N
Action	Increase EB Left Turn Phase							√		√		√	
	Increase NB Through Phase						√				√		
	Increase WB Through Phase						√	√	√	√			
	Eliminate RTOR for WB Right Turn	√	√	√	√								
	Omit NB Through Phase				√								
	Omit EB Left Turn Phase		√	√	√								

Table 10. Cross reference of link names in Table 9 with Detectors in Figure 50.

Link	Detectors
Northbound Plaza entry link	1002,1003
Northbound Goyeau	1006,1007
Eastbound Left turn	1009
Westbound Right turn	1004,1005

One logical clause (the third column in Table 9) is illustrated in Figure 50. In this case, high occupancy (oversaturation) is detected at the entrance to the toll plaza, and on the northbound route, but the eastbound left turn route is not oversaturated. Under this condition, the eastbound left turn is omitted for a short time and the RTOR of the westbound approach is also disallowed, allowing the traffic on the northbound route to get as much preference as possible. After the northbound queue is dissipated (below 50% occupancy in the logic shown in Figure 52) a different timing plan can then be selected.

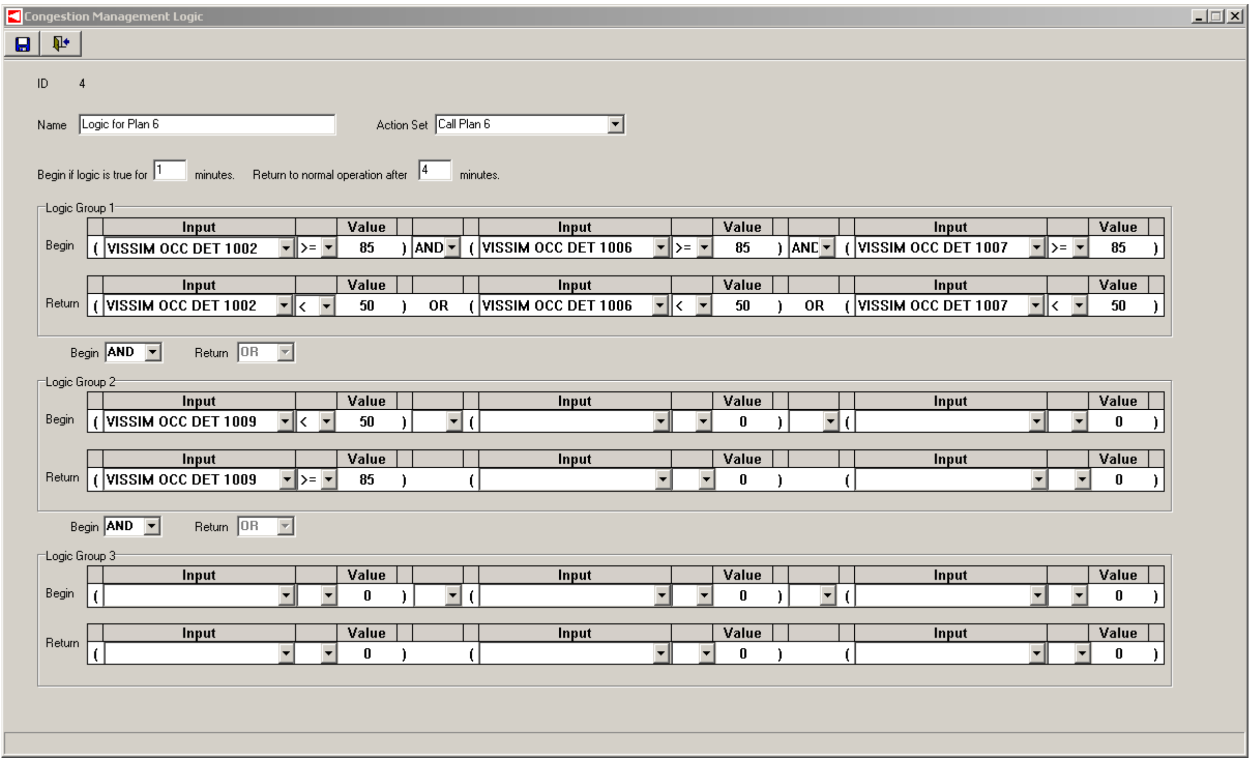


Figure 52. Logic engine example

As shown in this example, thresholds for each detector input can be selected independently and AND…OR logic can be applied between the inputs for a certain logic clause. The example shows three inputs per logic clause, but this could conceptually be increased to any number. We developed this initial tool with three inputs per clause to represent a three-lane approach or segment of roadway. Thus, the congestion identification logic could look at up to three lanes (AND) or any of the three lanes (OR) to exceed the congestion thresholds to have the clause evaluate as true.

After a single logic clause is evaluated as true or false, logic for up to two other locations can be combined with this clause (although this again can easily be extended to more detection stations). Each of these locations can be evaluated using AND…OR logic in combinations (A and B and C) or (A or B or C). The current design does not allow for more complex logic such as “A and (B or C)” or “(A and B) or (C and D)”, although these type of extensions would not be difficult if cases can be identified that require more complex combinations.

Begin and Return Conditions

Each clause has a begin and a return condition. This will reduce waffling of the logic from on and off conditions. This is an identical concept to the way that traffic responsive thresholds key the currently selected plan running for 15-30 minutes. In this example, the AND condition in the first row indicates that the occupancy of the detectors 2002, 2006, and 2007 must all be above 85%, for the clause to resolve as true. This condition will then remain true until the occupancy of all the detectors drop below, 50%. This reduces waffling in the timing plan decision if just a single lane drops below 50% for several cycles. Certainly more experimentation and research is necessary to provide guidelines on the setting of entry and return thresholds. Regardless of their exact values, it is important to make sure that the entry and exit thresholds are not exactly the same number. If the entry threshold is 85%, the exit threshold should be at least, say, 75% or lower. If the begin and return thresholds are the same value, much more frequent switching between plans will occur resulting in significant performance degradation.

Persistence Time Thresholds

Begin and return conditions for each clause each have persistence time thresholds, which can be changed independently, to make sure that the entry or release condition is persistent for a minimum amount of time before the clause is evaluated as true or false. The approach is currently designed so that the entry and release conditions for all clauses in the logic use the same values for the persistence time. This could be modified, but we find no empirical evidence or theoretical justification for different persistence thresholds. However, some experimentation does reveal that it may be useful to configure the “begin” and “return” persistence times differently. In particular, it seems to be more effective to configure a very low threshold for “begin” persistence and a longer time threshold for “return”. In our testing, entry thresholds of one minute and return thresholds of three to four minutes seemed to provide reasonably responsive operation. This rule of thumb specifically refers to the use of detector occupancy as trigger inputs, as TOSI, SOSI, and queue length estimates were not tested as extensively.

Online Performance Evaluation Framework

For the research project, this experimental congestion management logic tool was integrated with Virtual D4 traffic control software and the Vissim simulation system as illustrated in Figure 53. This system can use detector occupancy, TOSI/SOSI, and queue length as inputs to logic clauses. Field integration of the logic tool using TOSI/SOSI and queue estimation algorithms with real-world traffic controllers would be necessary for any field implementation. The system is part of the deliverables for this project and thus can be redistributed to anyone by NCHRP.

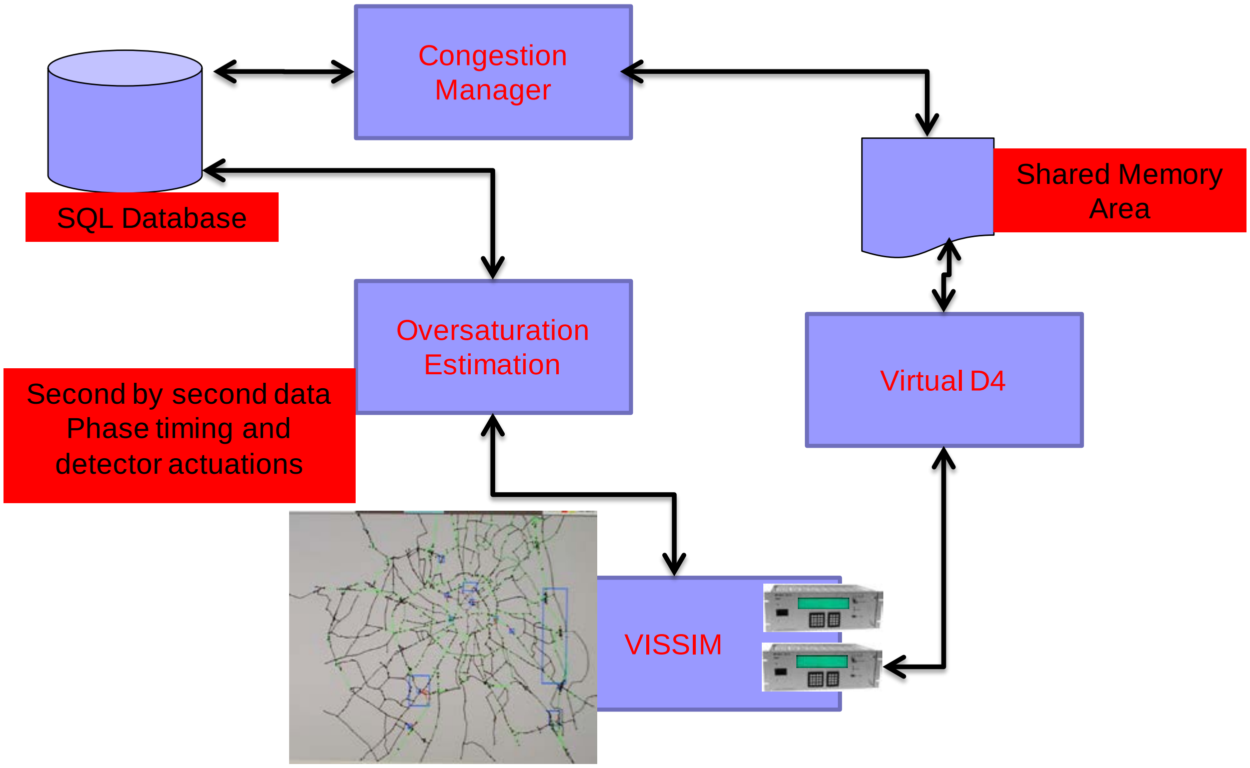


Figure 53. Online oversaturation management research software integration

The online process is implemented in a software-in-the-loop framework as illustrated below.

- Vissim simulation model
- Queue length and oversaturated conditions estimator module
- Congestion manager module
- D4 Virtual traffic controller module

Vissim is used to represent the "real world" movements of drivers in the case study locations and their responses to traffic control strategies. The traffic control strategies are implemented in Vissim using the D4 virtual traffic controller. D4 is real-world software used to control intersections in San Francisco and San Jose, CA, among other locations including Windsor, ON. This software-in-the-loop approach (SILS) replaces the internal approximate control logic of the modeling software with the algorithms of the actual traffic controller, thereby allowing simulation software to be used with full controller functionality at faster-than-real-time speed. This has been shown in previous testing to improve simulation performance to better than 3:1 real-time speeds.

The first step of the evaluation process, as shown above in Figure 53, is to obtain access to the traffic signal phase timing information and detector actuations inside of the simulation model and use that data to estimate queue lengths and compute oversaturation estimates. The queue length and oversaturated conditions measurement module was developed by the University of Minnesota. This module then sends these estimates of queue length and oversaturation intensity on a cycle-by-cycle basis to the congestion manager as illustrated below in Figure 54.

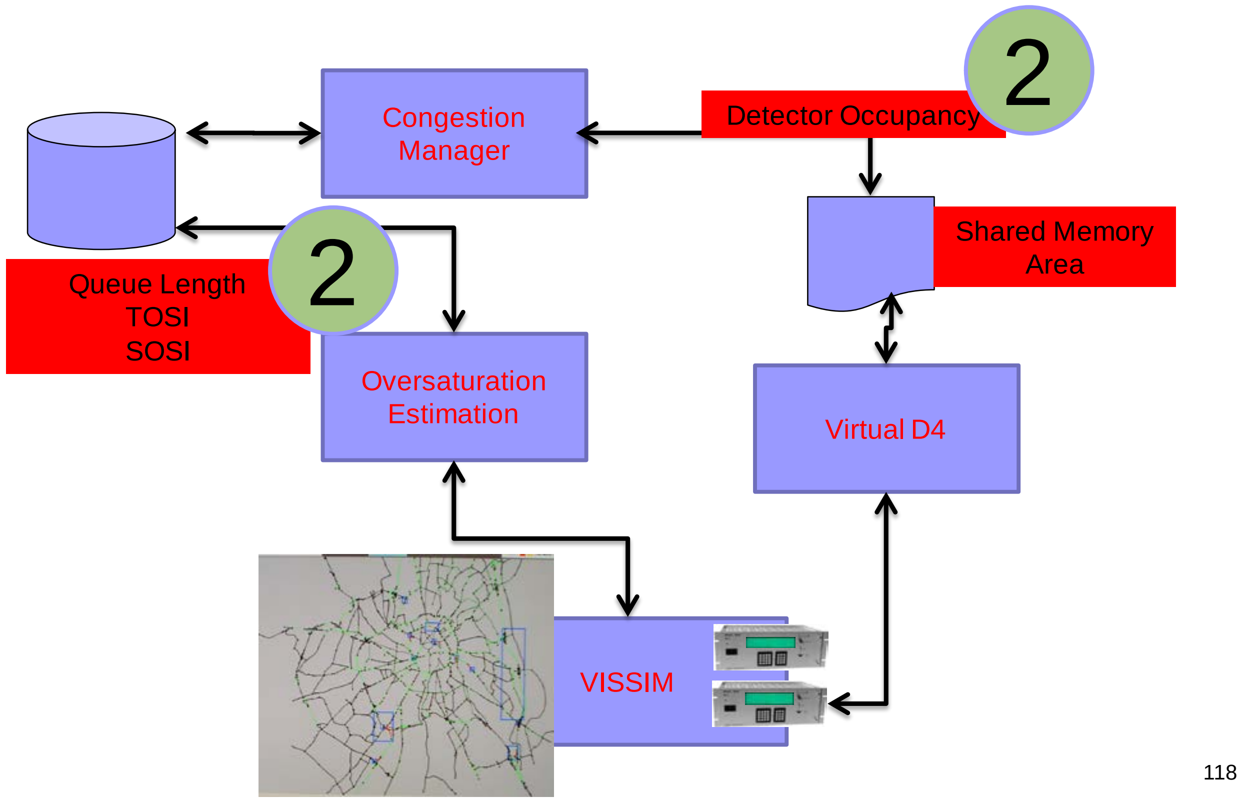


Figure 54. Research software integration – Step 2

The congestion manager uses these measurements in the generic if/then logic to determine which action plans should be enabled. If a plan change or other control action is required, the congestion manager communicates this information to the Vissim simulation by sending a control message to the D4 virtual controller(s) as illustrated below in Figure 55. D4 has a very simple shared memory interface that allows any external program to send commands by writing the desired plan information to the shared memory.

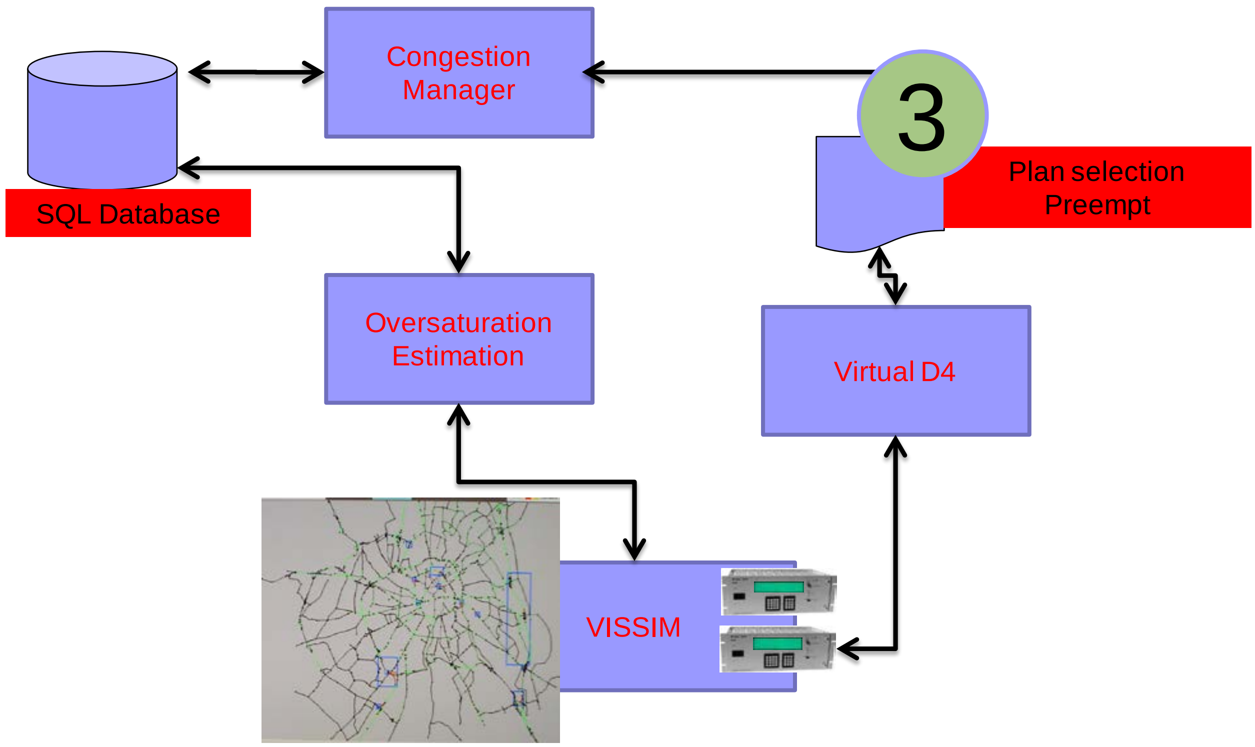


Figure 55. Research software integration – Step 3

D4 then acts on the plan commands sent by the congestion manager and implements the requested change to the traffic control strategy, such as phase omits, phase reservice, split and cycle time modifications, simultaneous offsets, and other strategies, by running the newly-requested plan. This step is illustrated in Figure 56. Note that this approach is not adaptive; these alternative plan designs must be pre-loaded on the controller and set up appropriately before-hand by the engineer.

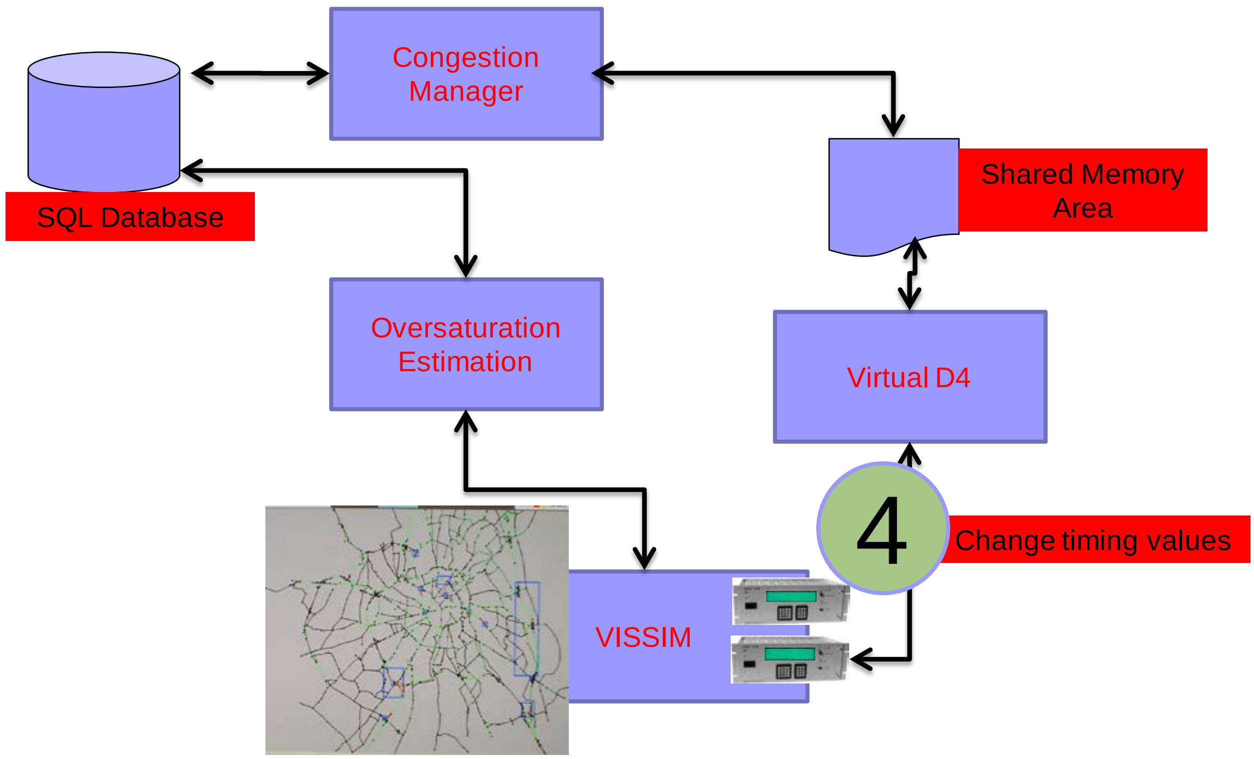


Figure 56. Research software integration – Step 4

After the new plans are implemented during the oversaturated conditions, the Vissim model collects the performance data such as delays, throughput, stops, and so on for effectiveness evaluation as identified below in Figure 57.

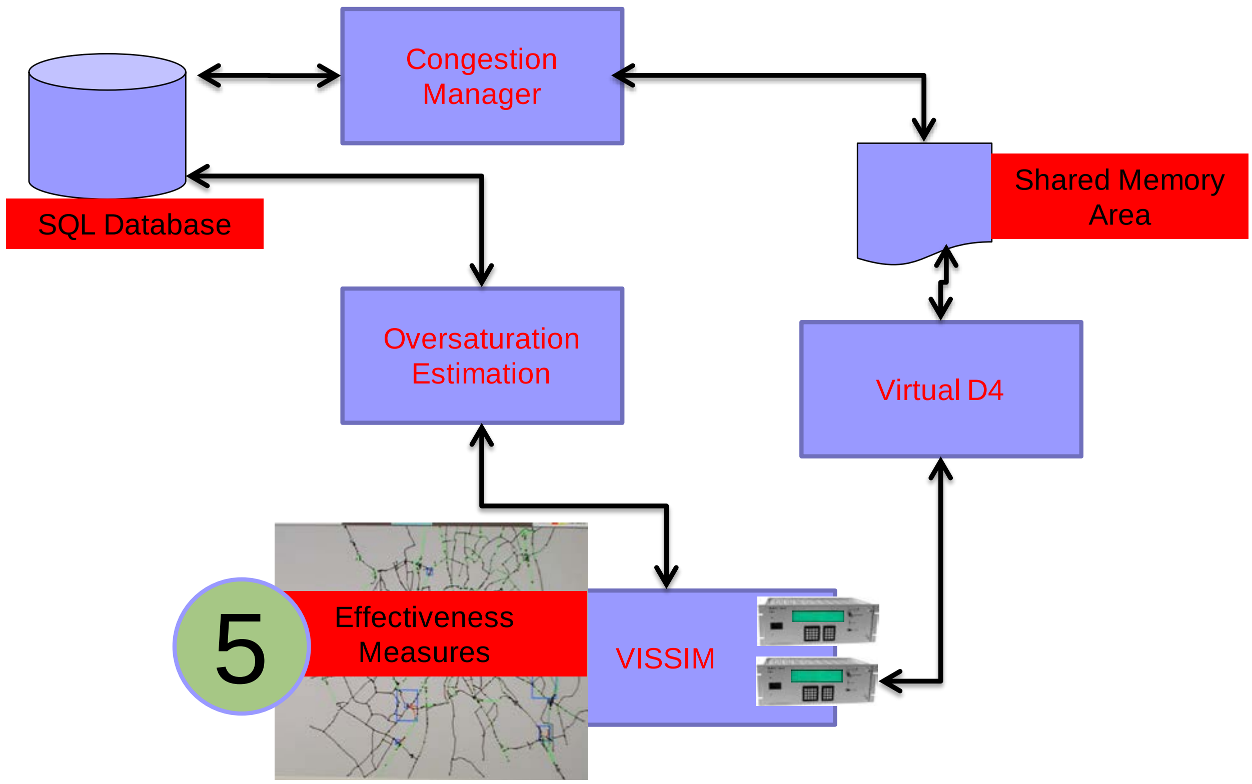


Figure 57. Research software integration – Step 5

A similar process is applied in the offline case, except the congestion manager module is not used to dynamically change signal timing strategies based on the congestion estimates. In the offline case, the changes to signal timing patterns are simply programmed ahead of time based on the analysis of the baseline conditions and implemented in the simulation model at the pre-defined times. In either case, multiple runs will be conducted to make sure that statistical variations in the effectiveness of a given strategy are captured.

Summary

In this section we described a process and a software tool for implementing mitigation strategies in an online manner for non-recurrent conditions. The first step of the process is to design the mitigation strategies based on observation of the condition and load those congestion plans into the field controllers in the network of interest. Based on measurement of TOSI, SOSI, queue length, or detector occupancy, the online tool can select the best matching timing plan or plans based on simple if…then logic. A truth table is used to map out all of the possible combinations of detector conditions that can lead to one plan or another being selected. The tool was integrated with the Vissim microscopic simulation system with the D4 SILS virtual traffic controller. The tools (minus Vissim and the D4 SILS) are part of the deliverables of the project.

In Chapter 3, we provide an example application of the tool to mitigate oversaturated conditions in a real-world situation. The performance results of this test were mixed, primarily because of the extremely challenging nature of the chosen problem. The process and the tool, however, were proven to be a viable method to tie together the design of mitigation strategies with the measurement of oversaturation severity estimates and queue length measurement.

Further in Chapter 3, we also describe an experimental methodology for direct calculation of green time adjustments from TOSI and SOSI measures on an oversaturated route. This procedure was applied and tested in an offline manner for two test cases in this project. However, the procedure may have promise for online application as well. This potential research and development effort is left for future work.

Chapter 3: Test Applications

In Chapter 2, we presented the three primary areas of new research and development that were conducted during this project:

- Development of quantitative measures for the severity of oversaturated conditions (queue length, TOSI, SOSI)
- Development of a process for generating and analyzing timing plans for mitigating oversaturated conditions
- Development of a tool for online application of mitigation strategies using TOSI and SOSI measurements to select alternative timing plans

In this chapter we describe several test cases conducted as part of the project to validate the research theories listed above. Two of the test networks (Reston Parkway in Herndon, VA and the Post Oak area in Houston, TX) were used in the development and testing of the multi-objective strategy development and evaluation methodology. Two other networks (TH55 in Minneapolis, MN and the Pasadena, CA downtown network) were used in development and testing of strategies directly related to TOSI and SOSI, including the development of an analytical procedure for directly adjusting green times. Two other test cases (an arterial in Surprise, AZ and a small network in Windsor, ON) were used to test the direct application of the guidance methodology developed in Task 6. All of the test applications were simulated using the Vissim simulation system with either the RBC controller or the Virtual D4 controller. While route proportions and demand flows were changed over time, no dynamic traffic assignment was used, i.e. vehicles in the simulation did not react to the congestion conditions by changing their route, destination, or forgo travel. The characteristics of these test cases are summarized in Table 11.

Table 11. Summary of test case attributes

Test Case	Config	Number of Ints	Spacing (ft)	speed (mph)	Typical phasing	Test duration	Causation	Types of symptoms	Recurrence	Critical locations	Types of mitigation	Components tested
Reston Parkway; Northern VA	Arterial with freeway interchange	14	500 to 3300	40	4, 6, 8	3 hours	Demand	All	Recurrent	2	Cycle, splits, offsets, phase reservice, gating	Timing plan development framework
Post Oak area of Houston, TX	Network	16	400 to 1800	30-40	2, 4, 6, 8	3 hours	Demand	All	Recurrent	8	Cycle, splits, offsets, phase reservice, gating	Timing plan development framework
TH55; Minneapolis, MN	Arterial	5	500 to 2600	55	4	1 hour	Preemption	Spillback, overflow queuing	Non-recurrent	2	Green extension, green truncation	Calculation of TOSI, SOSI, and queue length
Downtown grid; Pasadena, CA	Grid	22	400 to 1000	25-35	2, 4, 6, 8	2 hours	Light-rail; demand	Spillback, overflow queuing	Recurrent	5	Splits, offsets	Calculation of TOSI, SOSI, and queue length
Border Tunnel Entrance; Windsor, ON	Small network	9	400 to 800	25-35	2, 4, 6	45 min, 1 hour, 2 hours	Incident	All	Non-recurrent	1	Many	Online feedback tool; TOSI/SOSI
Surprise, AZ	Arterial	6	2600	45	8, 6	1.5 hours, 3 hours	Planned event	All	Both	1	Many	Application of guidance

Arterial Test Case: Application of the Multi-Objective Timing Plan Development and Evaluation Framework

The Reston Parkway arterial network is located in Reston, VA between Herndon and Vienna near the entrance to the Dulles International Airport. The network is significantly oversaturated during peak periods. The network consists of 14 intersections with a total length of 16,572 ft (3.1 miles). The spacing between intersections ranges from 524ft to 3,309 ft. The speed limit for the main arterial is 45 mph. Side street speed limits range from 15 mph to 45 mph. The arterial operates in a coordinated fashion during the day using traditional forward progression offsets. System detector count data was provided from The Northern Region Operation (NRO) of Virginia Department of Transportation (VDOT) which operates the traffic signal system in Reston Parkway.

The current system employs 170 controllers and the Management Information System for Transportation (MIST) central software. Actual detector data from this network was taken from August 11th, 2009 to September 10th, 2009. These detectors cover almost the whole network. MIST compiles system detector data in 15-minute intervals. During the evening peak period (2:30 P.M. to 8:00 P.M.) the arterial network operates with three different timing plans. The system operates a 130s cycle until 3:00 P.M., shifts to a longer cycle of 180s, and then transitions to a 140s cycle at 6:00 P.M. A sub-network of five intersections was selected for application of the timing plan development and evaluation framework presented in Chapter 2. This portion of the arterial has significant recurrent oversaturation due to the interchange with the Dulles Toll Road ramps as shown in Figure 58.

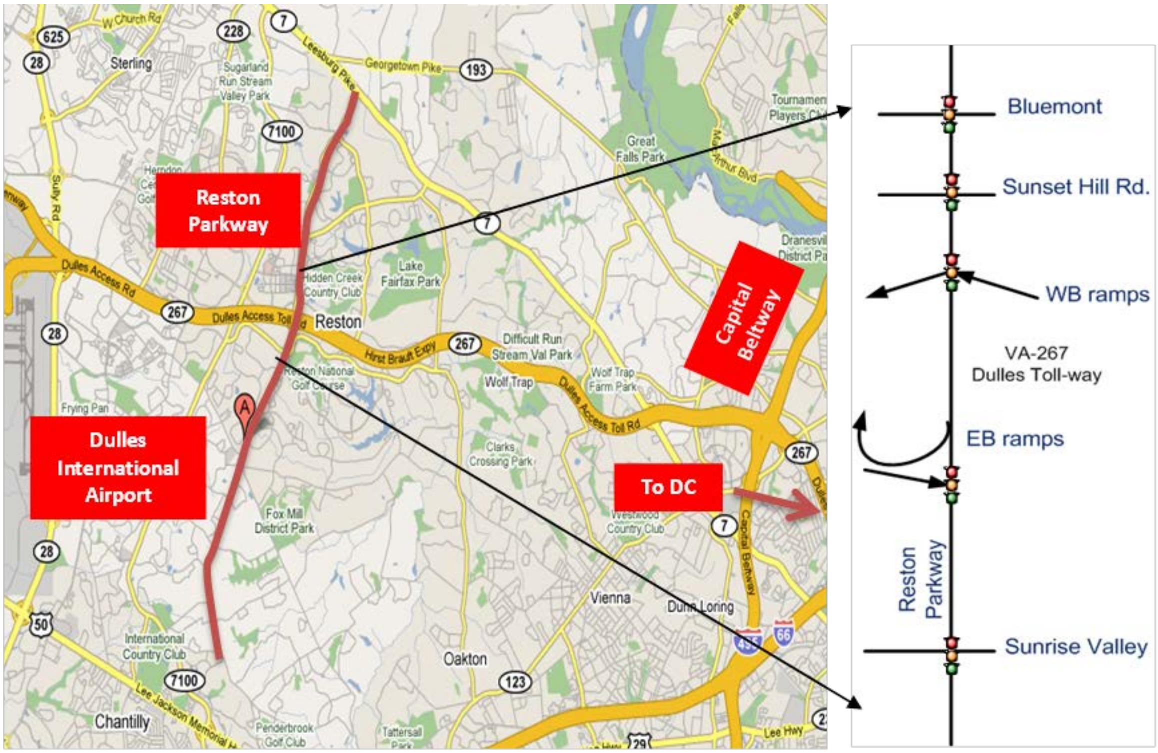


Figure 58. Reston Parkway network

Significant changes to the traffic patterns for both demand and directional distribution occur in the selected sub-network during the peak period. Turning volumes become very heavy and the dominating direction changes from south to north. The relatively short link length at the interchange with the toll road contributes to the oversaturated situation.

This network was coded in the Vissim microscopic traffic simulation model. The network was calibrated to the P.M. peak period traffic using field data collected from Monday June 1st to Wednesday June 3rd, 2009. The data collected included link travel times, travel speeds, queue lengths, and queue discharge headway.

Traffic Patterns on Reston Parkway

During the P.M. peak period, traffic patterns change rapidly in the Reston Parkway arterial network. Figure 59 through Figure 61 illustrate the changes in the arterial northbound and southbound traffic counts at various intersections during different points of the peak period. These changes in volume patterns emphasize the need to identify the critical routes in the system.

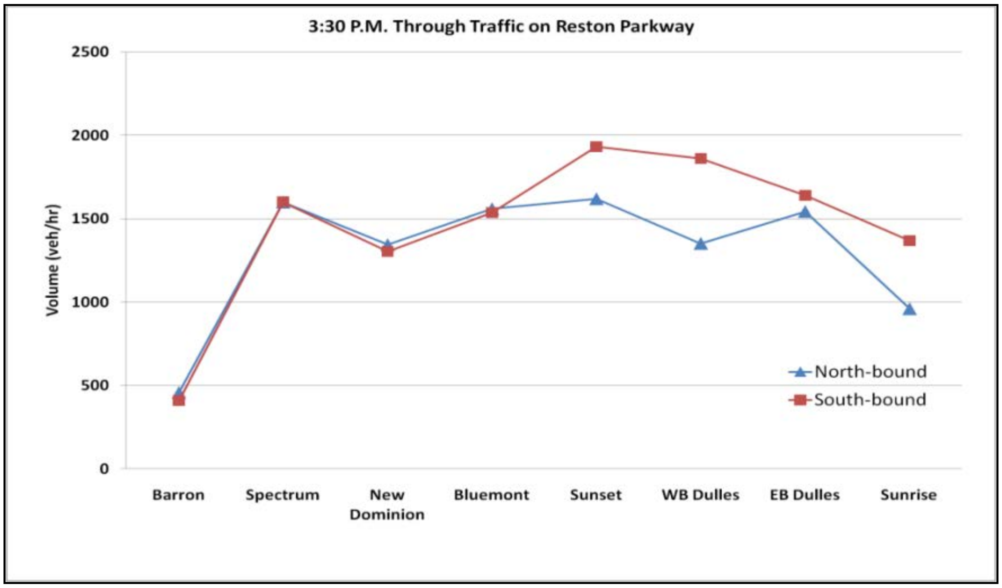


Figure 59. Changes in traffic patterns in Reston Parkway at 3:30 P.M.

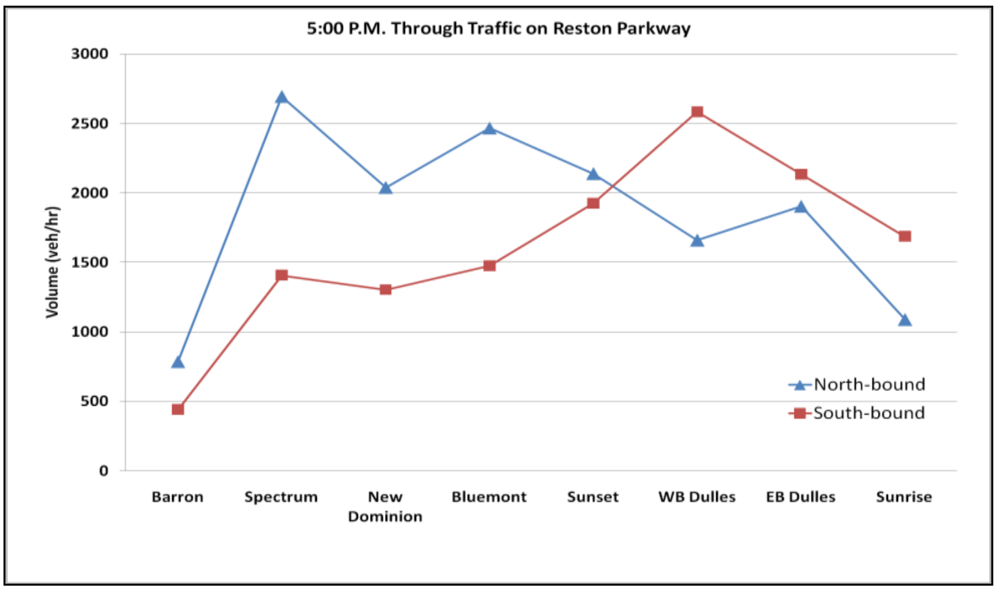


Figure 60. Changes in traffic patterns in Reston Parkway at 5:00 P.M.

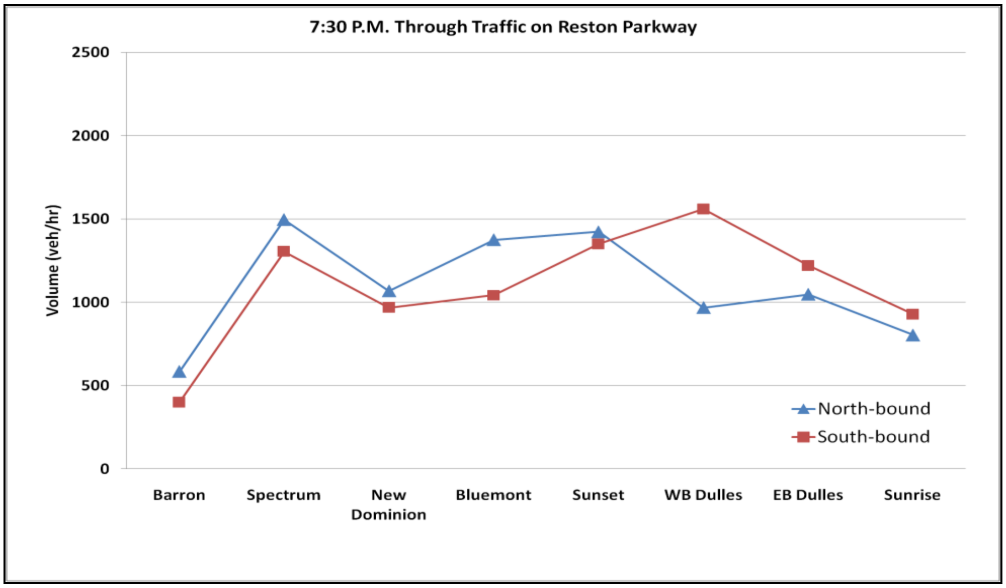


Figure 61. Changes in traffic patterns in Reston Parkway at 7:30 P.M.

Critical Route Scenarios

Six critical route scenarios were established based on the volume analysis of the actual observed counts. Throughout the day, each of the six scenarios may occur, thus requiring different signal timing plans. These scenarios are illustrated in Figure 62. The analysis of the volumes included (but is not presented in this report): correlation analysis, cluster analysis, and pattern recognition techniques. For each critical route scenario, background traffic was determined and the volume on each critical route during the peak period was estimated. This flow estimate for each critical route accounted for the maximum observed volume from the month of system detector data that was analyzed. A profile of demand volumes on each critical route was developed. The arrival demand on each critical route as well as the background traffic was modified in 15-minute increments.

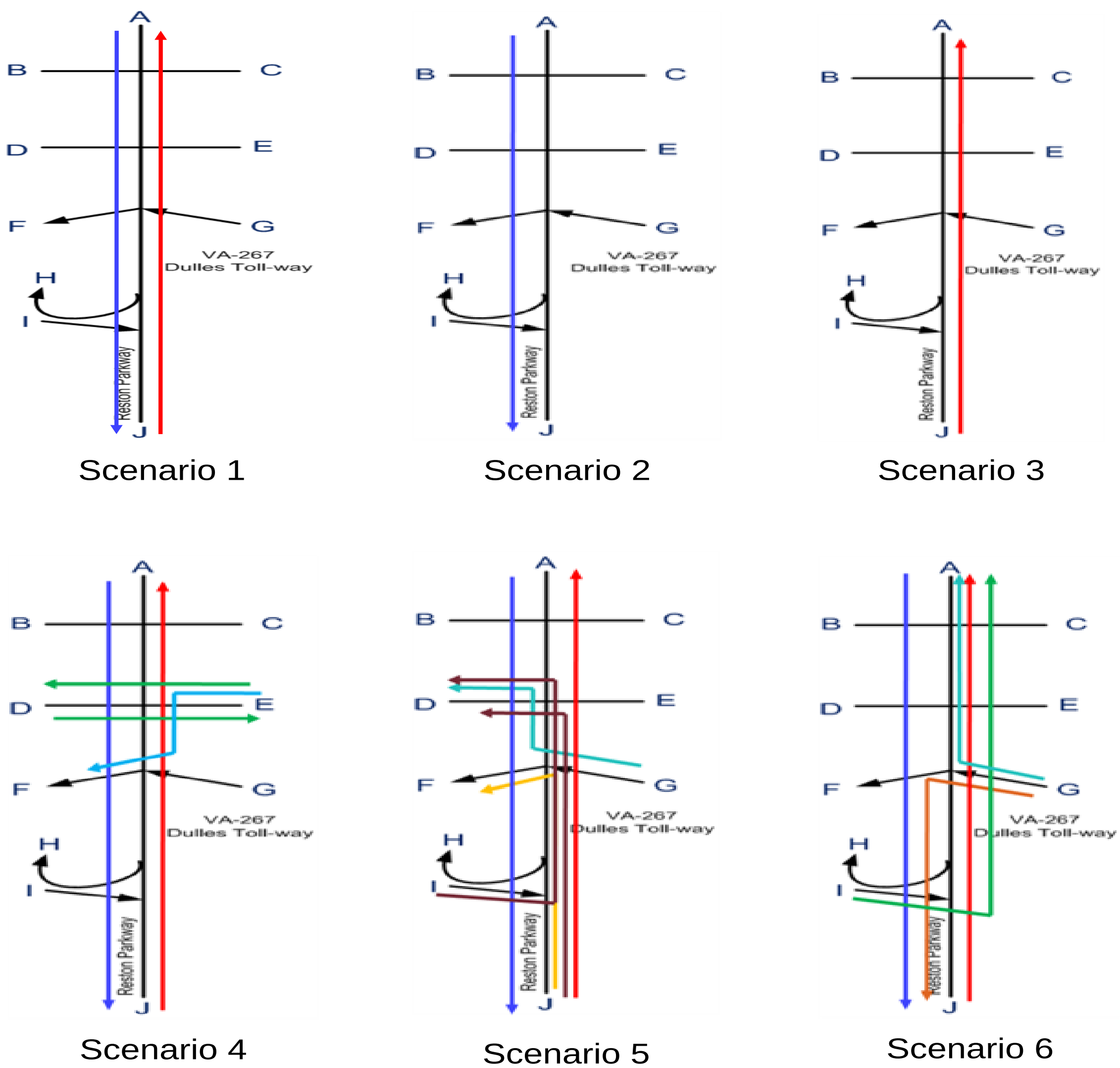


Figure 62. Critical route scenarios on the Reston Parkway network

In addition, the volume profiles were adjusted to account for the *delayed demand* that might not have been able to enter the network due to oversaturated conditions, and therefore would not have been reflected in the system detector data. Figure 63 illustrates the adjustment of the volume profile for an example movement. The adjustment of the profile increased the peak volume significantly, which is critical to be taken into account in the design of the mitigation timing plans. These adjustments were applied to each of the critical routes for each scenario shown in Figure 62.

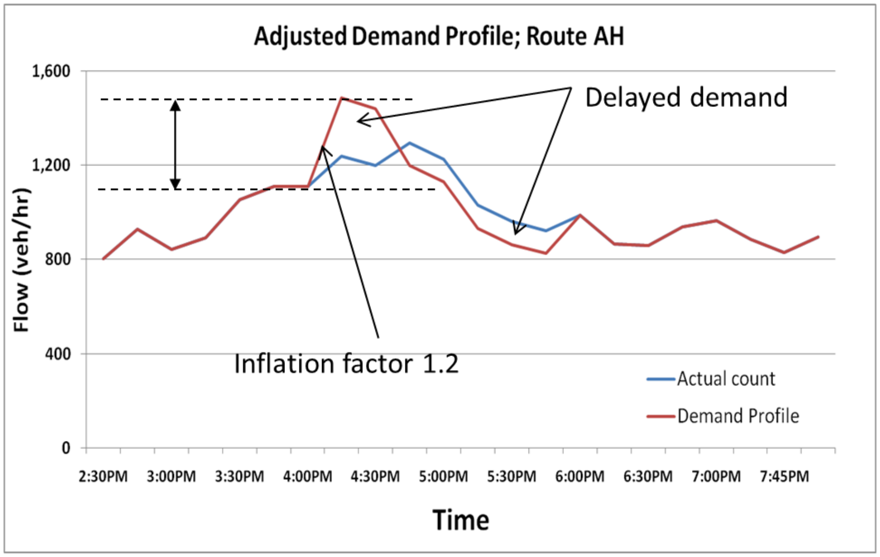


Figure 63. Adjusted demand profile for Route AH to account for demand unrepresented in the system detector counts

Illustration Using of Critical Routes to Determine Mitigation Strategies

In this section we describe the use of critical routes to develop mitigation strategies. This discussion will focus on Scenario 5 since it represents one of the most challenging situations with severe oversaturation and symptoms. Figure 64 identifies the attributes of this scenario. Critical routes, oversaturated symptoms, and proposed control strategies are illustrated in Figure 64.

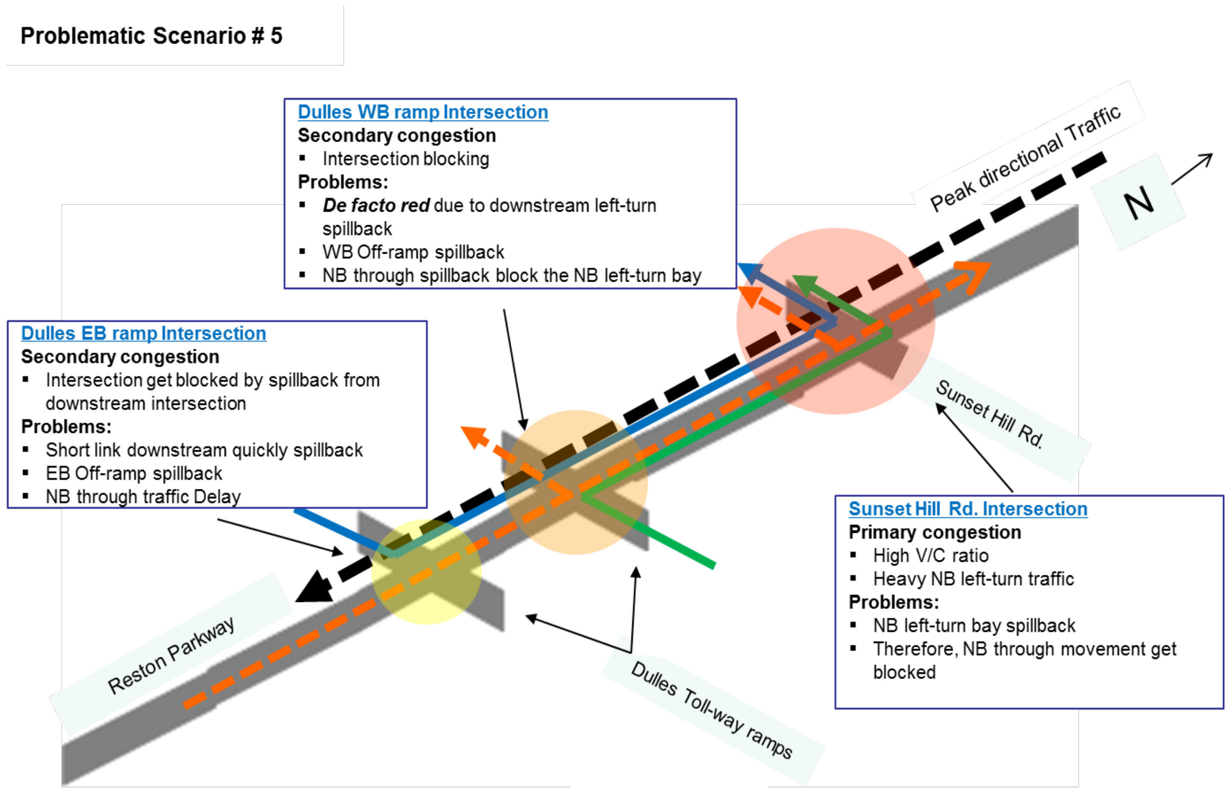


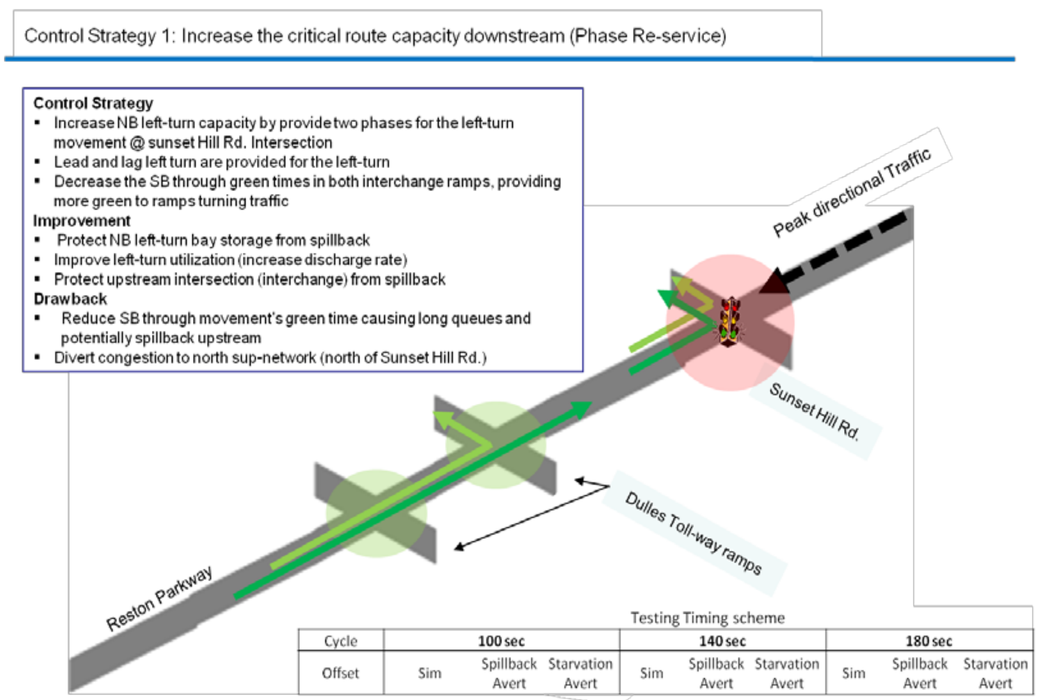
Figure 64. Problematic Scenario 5: critical movements and diagnosis

In this scenario the oversaturated approach northbound on the Parkway at Sunset Hill Road for left-turn vehicles causes secondary congestion at the two upstream intersections at the eastbound and westbound toll road ramps. In addition, the primary peak direction of through traffic is southbound on the parkway. This combination of flows provides a rather challenging situation for the design of mitigation timing plans.

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				

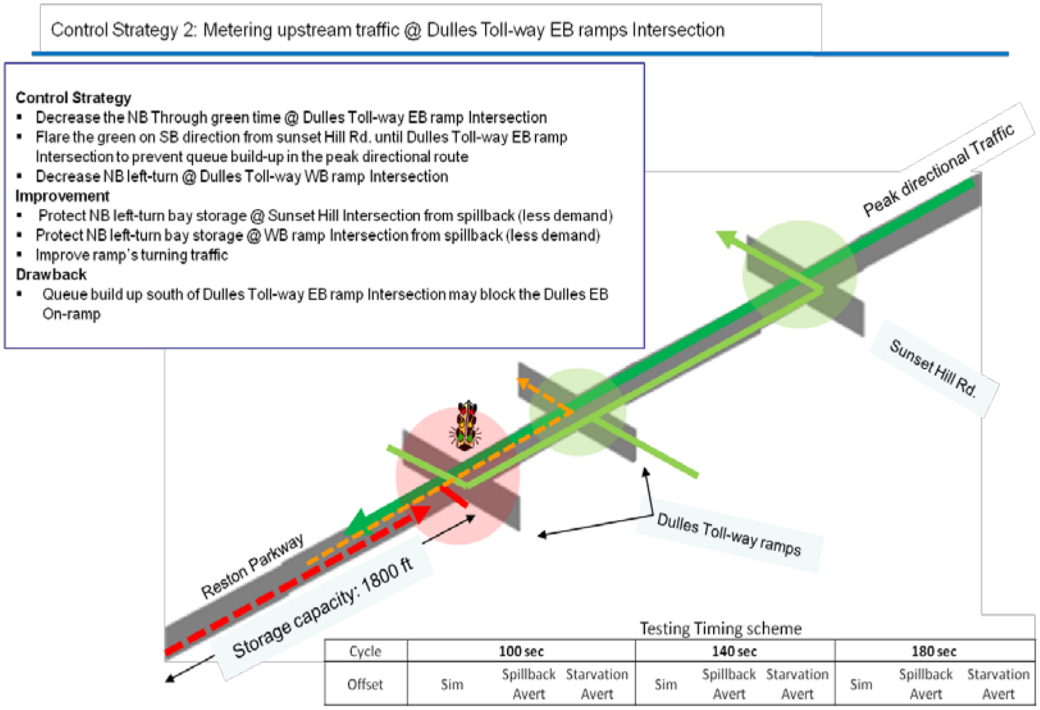
Figure 65. Key attributes of Scenario 5

Based on observation of the critical routes in this scenario, we first determine that one of two different canonical operational strategies will be applied in this test case, either a metering strategy for the northbound critical routes or a phase reservice strategy for the northbound left turn at Sunset Hill Road. Both of these methods have strengths and weaknesses. A comparison of the two methods is illustrated in Figure 66 (a) and (b) below.



(a) Northbound left-turn phase reservice at Sunset Hill Road intersection

Figure 66. Comparison of phase reservice and metering strategies



(b) Metering of northbound traffic at the Dulles eastbound ramp intersection

Figure 66. (continued)

The main strength of the phase reservice strategy is that the additional green time for the left turn alleviates the spillback on the northbound approach at Sunset Hill Road which is the primary source of oversaturation on the arterial. The main drawback is that the green time for the southbound route will be significantly restricted and will create southbound queues north of the Sunset Hill Road intersection. The main strength of the gating strategy is that by storing vehicles on the approach link, the queues on the left turn at Sunset Hill do not grow as quickly. The main drawback of this approach is that the queue on this link can block the right turn to enter the eastbound toll road. Figure 67 illustrates the storage on the link that was selected for the metering strategy.

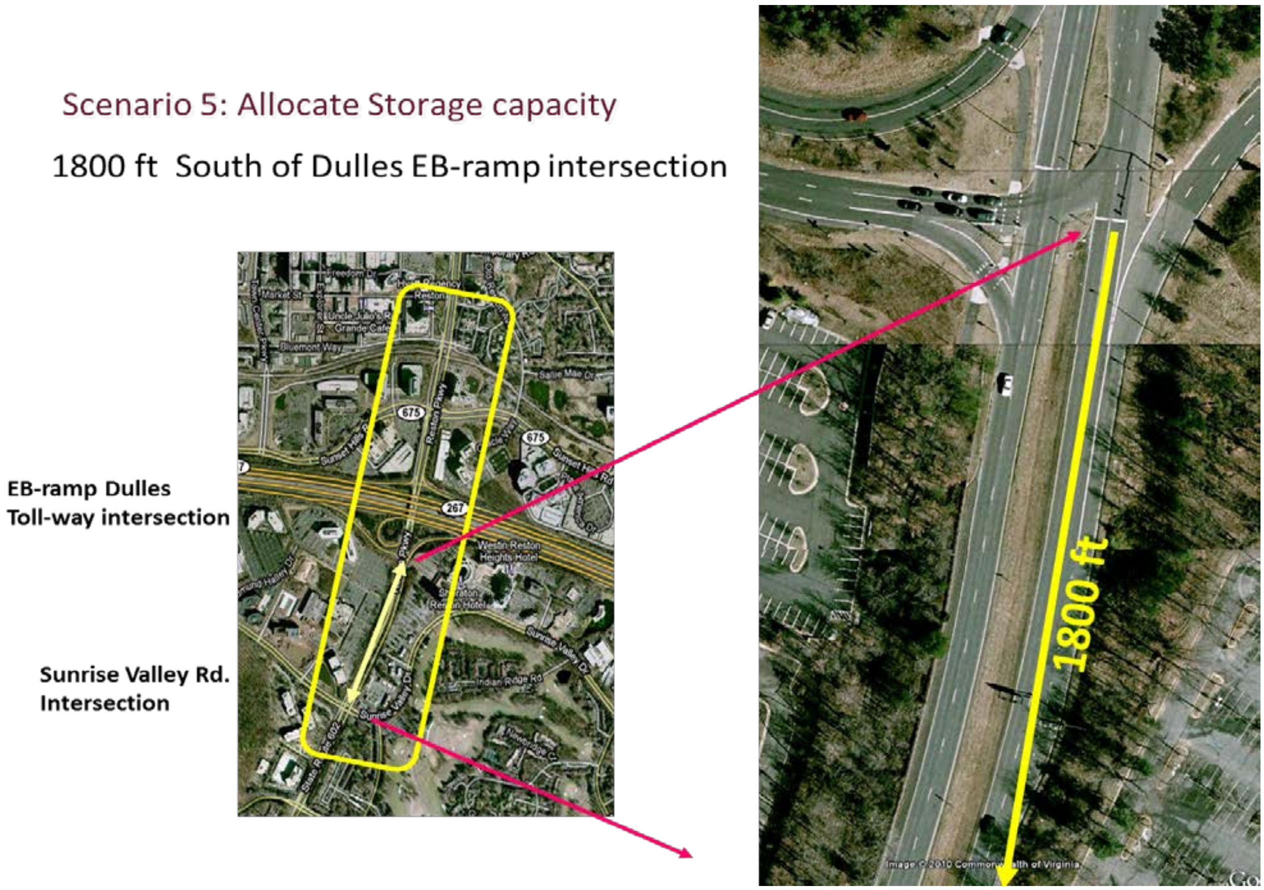


Figure 67. Storage capacity of the link used for metering in the Reston Parkway network

The timing plan generation framework presented in Chapter 2 was used to generate timing plan parameters for the proposed control strategies. This process is described further in the next section.

Cycle Length Calculations

The first step of the optimization and evaluation process is to calculate feasible cycle lengths for each intersection in the network. The cycle length calculation procedure presented in Chapter 2 was used to compute the maximum cycle length that does not result in spillback at each intersection. The resulting maximum cycle times are shown in Table 12. Note that the maximum cycle times for each intersection vary widely so a judgment procedure is necessary to determine cycle times that can be implemented on all of the intersections in the system. Note also that the equation from Roess (1994) almost always estimates a longer maximum cycle than the Lieberman equation. In order to test the performance of several different cycle times, we choose several canonical cycles in the range that meet most of the requirements. If the minimum cycle time of these maximums was used, the evaluated cycles would be less than 85s. This was considered to be unrealistic to implement because of the number of phases at the larger intersections.

Table 12. Maximum cycle length before spillback occurs on critical network links

	Network Links		Length	Storage	Max C	Max C
	From	To	Ft	Veh/lane	(Lieberman)	(Roess)
1	Cameron	Spectrum	830	35	195	587
2	Spectrum	Bowman	710	30	177	174
3	Bowman	Dominion	670	28	103	150
4	Dominion	Bluemont	800	33	215	352
5	Bluemont	Sunset Hill	640	27	87	100
6	Sunset Hill	Dulles WB	710	30	115	188
7	Dulles WB	Dulles EB	600	25	85	129
8	Dulles EB	Sunrise Valley	1800	75	232	236

Table 13 illustrates some of the critical geometric data for this test scenario. These distances are critical in the computation of the offsets on these links.

Table 13. Critical network links and left-turn bay storage lengths

Arterial Network Links	Length (ft.)	SB Left-turn bay (ft.)	NB Left-turn bay (ft.)
South Lake Rd.– Sunrise valley Rd.			435
Sunrise valley Rd. – Dulles EB ramps	1800	720	-
Dulles EB ramps - Dulles WB ramps	710	-	645
Dulles WB ramps – Sunset Hill Rd.	710	-	480
Sunset Hill Rd. – Bluemont Way	640	300	380
Bluemont way – New Dominion	800	390	

Table 14 lists the parameters of the shockwave model that were used to calculate the minimum and maximum offset values. From this feasible range, three sets of offset values were chosen for evaluation (i.e., progression offsets, starvation-avoidance offsets, and spillback-avoidance offsets) in each of the six different critical route scenarios.

Table 14. Shockwave Modeling Parameters

Parameters Value	
Discharge rate (veh/hr/ln)	1,500
Desired speed (ft./sec)	36
Queue discharge wave speed (ft. /sec)	18
Acceleration rate (ft./sec^2)	4
Deceleration rate (f/sec^2)	4
Start-up loss time (sec)	3
Average vehicle length (ft)	25
Peak Hour Factor	0.9
Lane Utilization Factor (Fu):	1

These combinations of offsets are referred to with the following acronyms:

- ***Sim***: Simultaneous (zero) offsets
- ***Max***: Offsets to prevent starvation at downstream intersection
- ***Min***: Offsets to prevent spillback at upstream intersection
- ***Med***: Offsets that are the medium value between Max and Min

Design of Splits and Offsets

Figure 68 presents the split-offset calculation procedure developed in Chapter 2 for the mitigating oversaturation. The initial splits at each intersection are calculated based on the v/c ratio for each phase. These splits are then modified in an iterative procedure by estimating the expected queue lengths for each approach for v/c either greater than or less than 1.0. To mitigate the oversaturation on the critical route(s), the degree of saturation values for each approach is constrained by a pre-specified threshold. The resulting queue-to-link ratios are then used to determine the bounds on the values of the offsets that will avoid spillback and starvation. Figure 69 through Figure 71 illustrate the offset values obtained using this procedure for the southbound route and the two routes together.

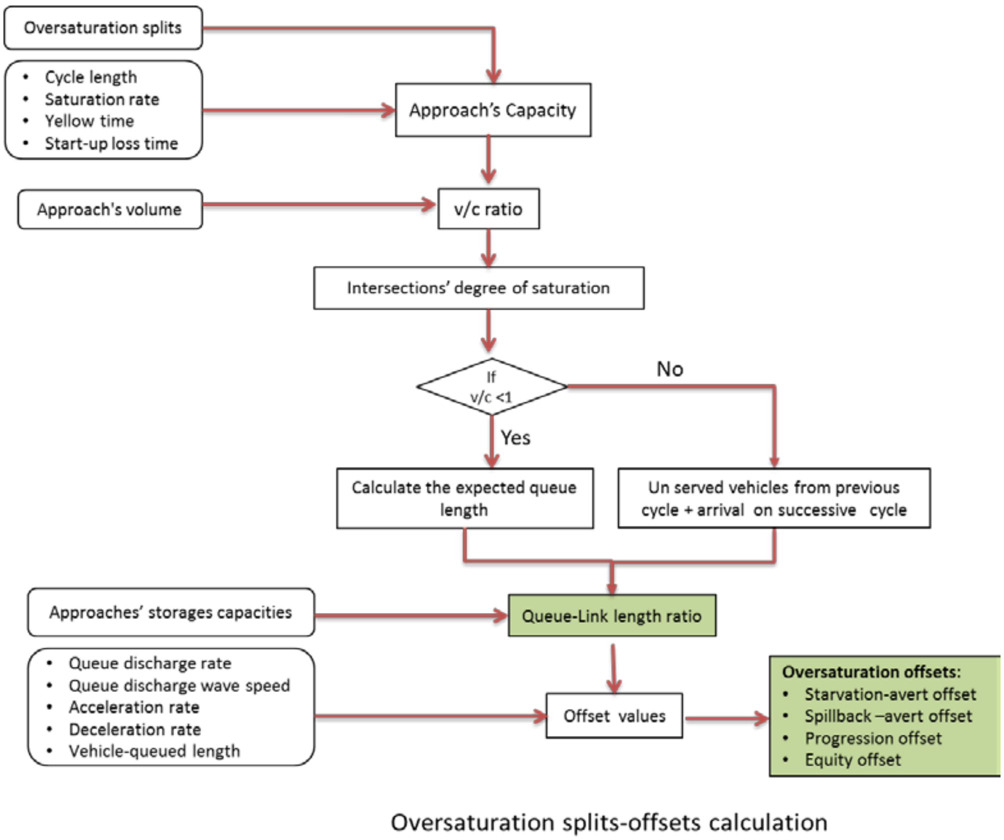


Figure 68. Split-offset calculation procedure

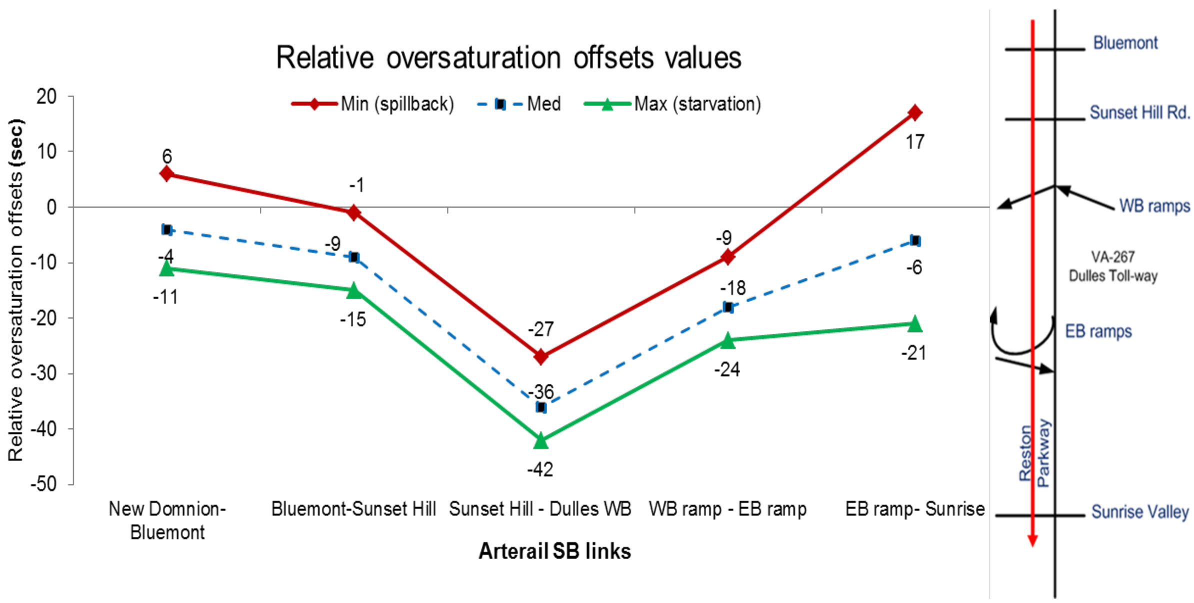


Figure 69. Oversaturation offsets for the southbound critical route

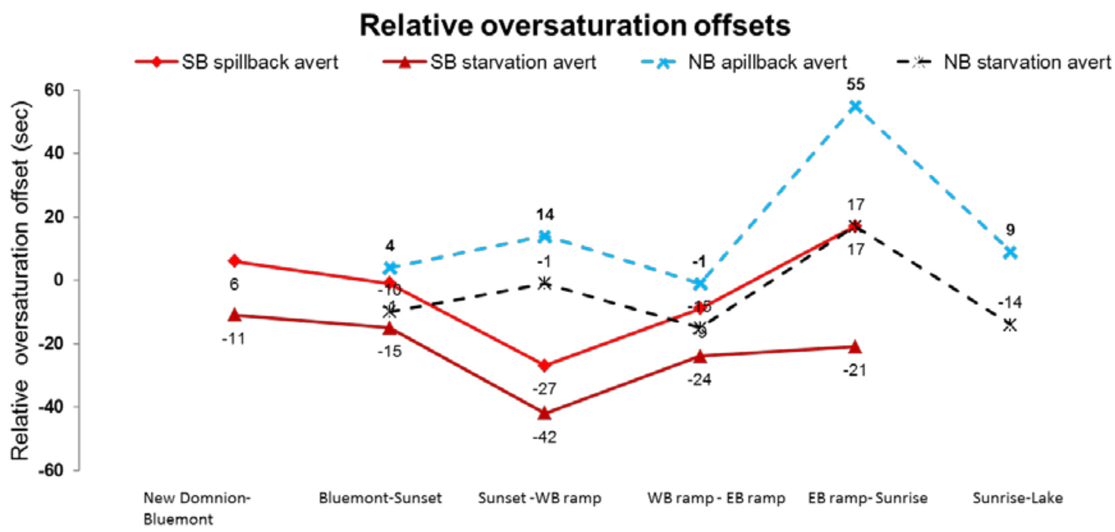


Figure 70. Oversaturation offsets for both southbound and northbound critical routes

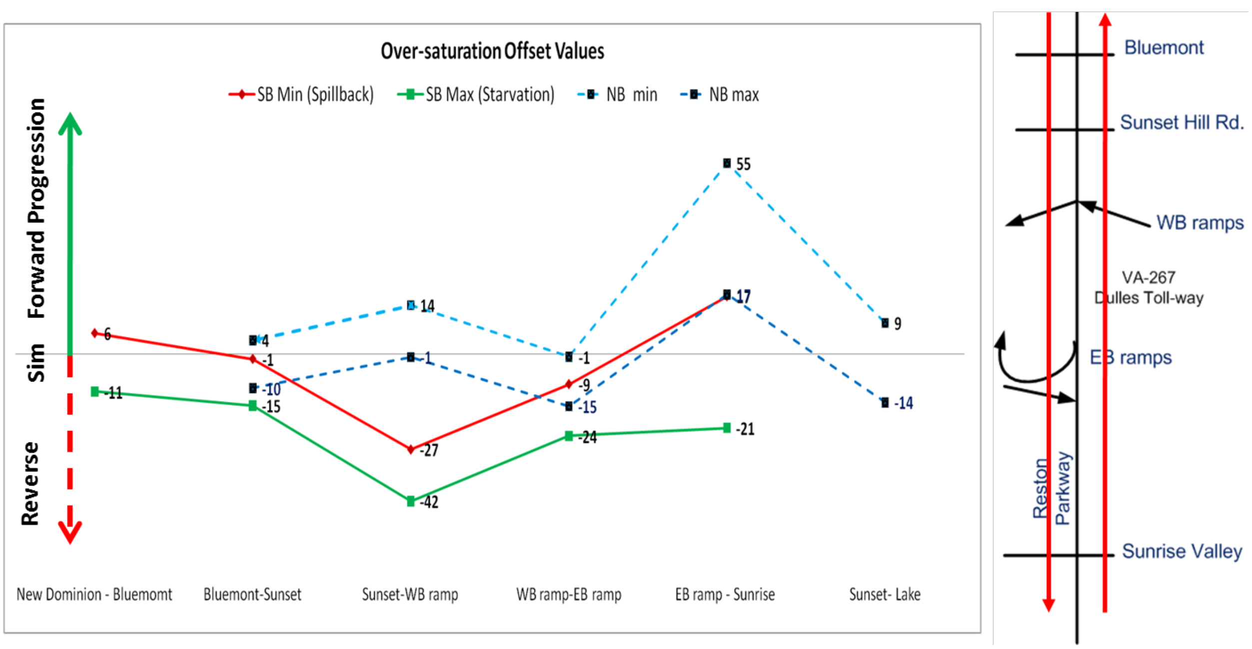


Figure 71. Scenario 5 offset design values (min and max), for northbound and southbound progression

By obtaining the maximum and minimum values of the offsets using this procedure, a feasible region is identified that satisfies both objectives of spillback and starvation avoidance. The offsets obtained from this procedure can vary from negative to positive values, depending on the overflow queue length constraints, link length, and green split time. Note that in this scenario, it is not possible to find offsets that ensure that spillback and starvation are avoided in both directions. In this scenario, the most critical route, the northbound route, is selected.

Simulation Experiment

Timings plans were designed for six volume and critical routing scenarios (denoted scenarios 1, 2, 3, and so on). Three cycle length values (100s, 140s, 180s) were selected for evaluation and four types of offsets (simultaneous, minimum, medium, and maximum) were evaluated, where minimum and maximum refer to the spillback-avoidance and starvation avoidance respectively. Each combination of cycle time and offsets was then given a unique number (i.e. Strategy 1, 2, and so on) are shown in Table 15. These 12 combinations of timing plans were also compared to the current baseline timings used by VDOT in the real world. Furthermore, the combination of cycle time, offsets, and splits was then combined with either metering in the northbound direction or phase reservice at the critical intersection. Table 16 illustrates the components of the control strategies for Scenario 5. For the metering strategy we chose to reduce the green time for the metered approach by 20% from its previously calculated value. Five runs, with different seed numbers, were conducted in Vissim to compute the average and variance of the performance measures for each strategy. The scenario lasts for five hours. Thirty minutes of simulation prior to the peak period, and approximately one hour of simulation after the peak period to allow enough time for the queues to dissipate and return the system to steady-state operation. Pareto fronts are then calculated to determine which timing plans are non-dominated on the three optimization objectives.

Table 15. Combination of cycle time and offset values for each strategy

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)

Table 16. Control strategy combinations with metering or phase reservice

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

Simulation Results and Evaluation

The simulation output results were collected every 15 minutes of the simulation time including total system and link-by-link delay, total number of stops, and total system throughput measured by the total number of vehicles leaving the network during the time period. Figure 72 illustrates the four-dimensional Pareto front obtained for Scenario 5. These diagrams directly demonstrate the strategy that is optimal for a specific 15 minute time period during the simulation.

The Pareto front diagrams present only the optimal non-dominated solutions among all of the tested control strategies for a particular volume scenario. In both diagrams, the x-axis (horizontal) represents the total system delay time, the y-axis (extending into the depth of field) represents the total system stops, and the z-axis (vertical) represents the total system throughput. In the top diagram, the color of each dot is the simulation time. Dark colors represent early times in the simulation (the loading period) and light colors represent the end of the simulation time (recovery period). In the bottom diagram, the color of each dot represents the cycle time of the strategy that is dominant at that time in the simulation. Blue colors represent low cycle times (in this example, 100s) and red colors represent high cycle times (in this example, 180s).

For this specific test case, Figure 72 (top) can be interpreted as follows: during the processing regime timing plans that are designed to maximize throughput (orange color scale refers roughly to the processing regime) are performing best. However, timing plans that are optimized for minimizing total system delay and total stops perform the best on all three performance scales during the loading and recovery regimes (white and yellow color scale represent the recovery regime, dark red and brown represent the loading regime).

In Figure 72 (bottom), illustrates timing plans with longer cycles (dark red color scale) that were designed to maximize the throughput are dominant during the processing regime. Timing plans with short cycles (blue color scale) that were designed to minimize delay are dominant in both the loading and recovery regimes of the scenario. Control strategies with the medium cycle length (140s) with different offset values (light blue, green, and yellow) are not performing well in this particular scenario since they appear infrequently as a dominant timing plan during any 15-minute time period. It is clear from this analysis that applying just one timing plan for an entire scenario, when it lasts for an appreciable amount of time (e.g. over an hour), is not recommended. Different approaches to mitigation are needed during the three regimes of operation (loading, processing, and recovery).

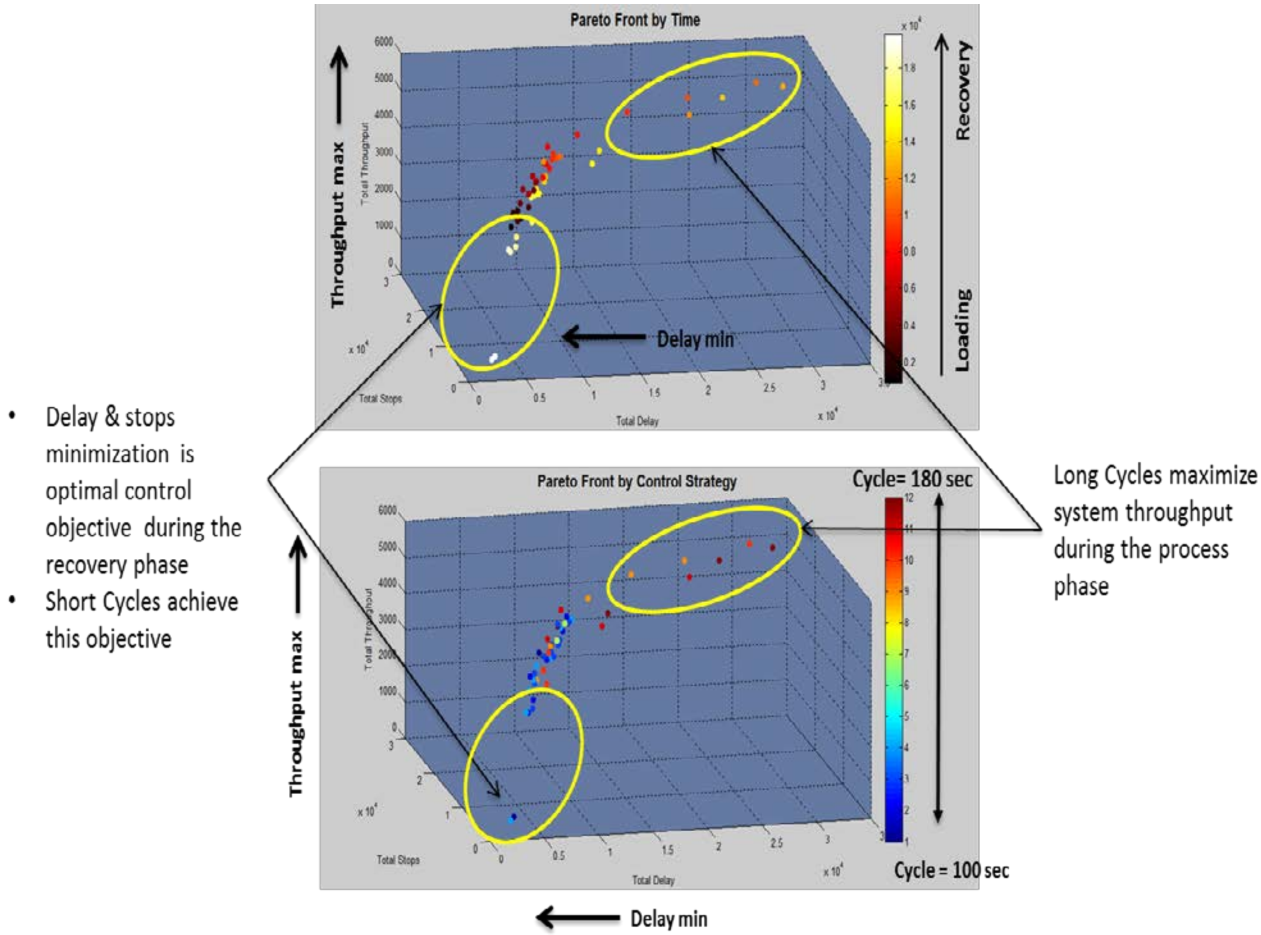


Figure 72. Illustration example of the Pareto front for Scenario 5: metering strategy

Pareto Front Analysis

In this section we present the results of the Pareto front analysis for the 25 strategies (12 combinations of cycle time and offsets with either phase reservice or metering plus the baseline timing plan) on the six critical route scenarios. In general, it seems strategies that include metering, longer cycles perform better during the recovery regime. On the other hand, when using phase reservice at the critical intersection, timing plans with shorter cycles dominate both the loading and recovery regimes. If you consider only the throughput objective, the graphics tend to illustrate that plans with shorter cycles perform poorly during the loading regime and then system throughput starts to improve as demand increases in the network. Toward the end of peak period the timing plans with shorter cycles return to their poor performance. These temporal diagrams convey invaluable information to the decision makers and analysts regarding system performance, but take some investment of the reader to adequately digest them. In the following sections we have provided some summary bullets for each test scenario to quickly summarize the information depicted in the diagrams. The percentage improvement statistics are relative to the baseline timing plan used in the real world (130s cycle with forward progression offsets; no metering or phase reservice considerations).

Scenario 5 Results: Pareto Front

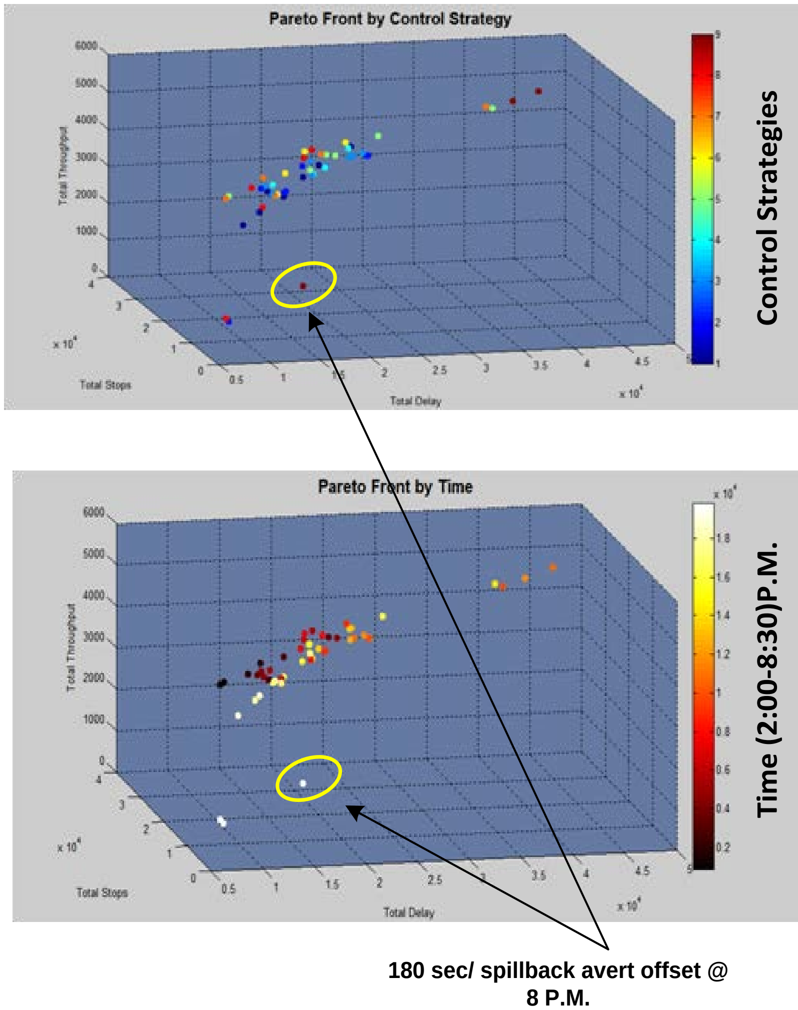


Figure 73. Scenario 5 with phase reservice

Routing Scenario 5 with phase reservice control findings:

- Minimum cycle lengths and mid-value offsets minimize delay
- Medium cycle lengths and mid-value offsets maximize throughput
- The phase reservice strategy has lower total throughput than the metering strategy
- Throughput maximization strategies reduce delay by 29% and increase throughput by 1%
- Delay minimization strategies reduce delay by 29% and reduce throughput by 1%

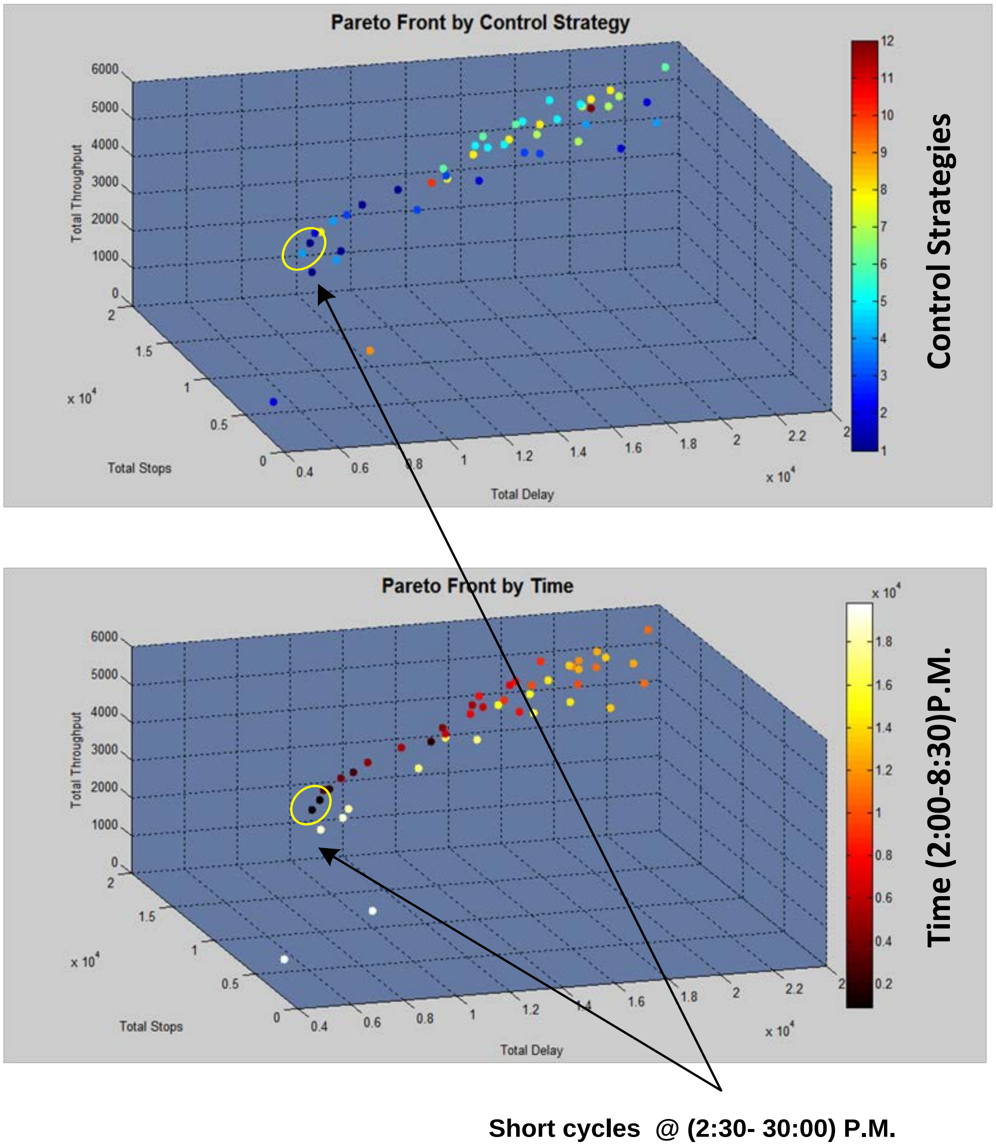


Figure 74. Scenario 5 with metering

Routing Scenario 5 with metering control findings:

- Short cycle with min/mid offset minimizes delay and stops
- Medium cycle length maximizes throughput
- Longer cycle length increases delay significantly
- Throughput maximization strategies reduce delay by 13% and increase throughput by 11%
- Delay minimization strategies reduce delay by 35% and increase throughput by 7%

Table 17. Scenario 5: total improvement over the baseline strategy

Strategy	Cycle (sec)	Delay	Stops	Throughput
Phase Re-service				
1	100	0.26	0.45	-0.01
2	100	0.26	0.45	-0.01
3	100	0.29	0.42	-0.01
4*	100	0.29	0.6	-0.01
5	140	-0.17	0.07	-0.01
6	140	-0.13	0.22	-0.01
7	140	-0.14	0.15	-0.01
8	140	-0.09	0.18	-0.02
9	180	-1.17	-0.25	-0.08
10	180	-1.42	-0.19	-0.12
11	180	-0.90	-0.29	-0.07
12	180	-1.22	-0.29	-0.41
Metering				
1	100	0.35	0.45	0.07
2	100	0.35	0.45	0.07
3	100	0.34	0.42	0.08
4	100	0.33	0.42	0.06
5	140	0.06	0.07	0.1
6	140	0.17	0.22	0.1
7*	140	0.13	0.15	0.11
8	140	0.14	0.18	0.09
9	180	-0.18	-0.25	0.11
10	180	-0.15	-0.19	0.08
11	180	-0.24	-0.29	0.06
12	180	-0.24	-0.29	0.06

The following figures illustrate the dominant timing plan during each 15-minute period of the scenario for each of the control objectives based on the average performance on each objective over the five simulation runs. One thing to note is how the optimal plan for some of the objectives moves around considerably. It would be unreasonable to consider switching between the timing plans as shown here because of undesirable transitioning effects. It is also important to note that each scenario using a particular strategy evolves during the peak period differently. So the performance at any 15-minute period for a particular strategy is dependent on the performance of that strategy in the previous 15 minutes. However, the presented demand profiles are the same in every case (common random number seeds), so the trend in performance of a particular timing strategy is reasonably consistent.

For the timing plans that include phase reservice, optimal strategies that were designed to minimize delay appear to be stable during most the of the peak period (Strategies 3 and 4). On the other hand, strategies that are optimal for throughput maximization show instability as illustrated in Figure 75. For timing plans that included metering, delay minimization and throughput maximization optimal strategies appear to coincide with each other for most of the peak period (the green diamonds are overlaid behind the blue diamonds).

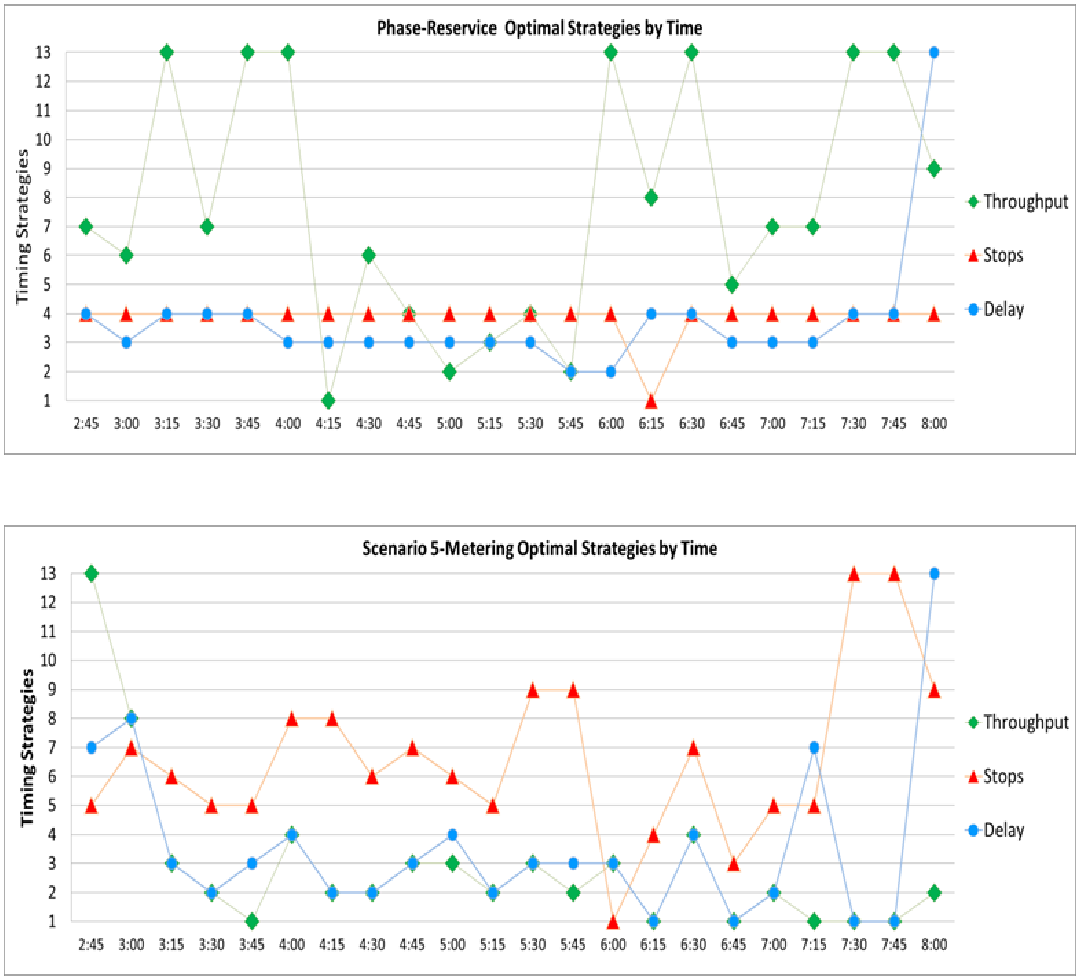
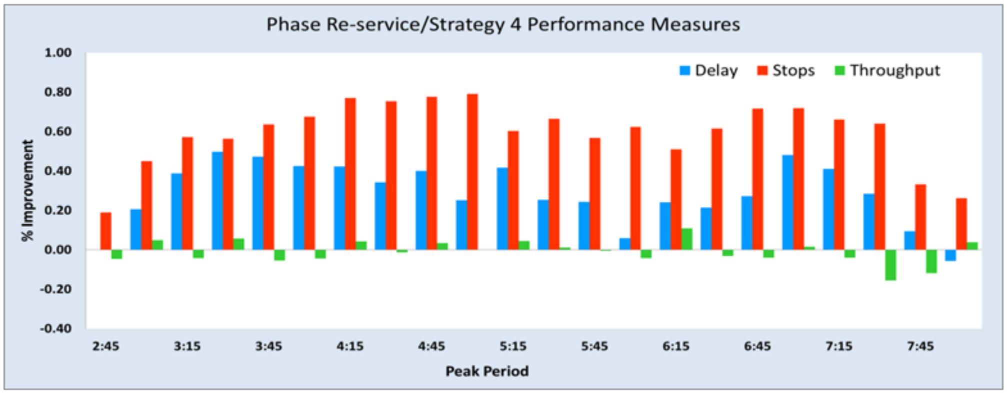


Figure 75. Scenario 5: optimal control strategies for each 15-minute period

Finally the mitigation strategies were evaluated for each 15-minute interval with respect to the timing plan implemented in the real world. The percentage improvements for each measure are presented in Figure 76. Table 18 presents the total average improvement percentages during the entire peak period. For each scenario in the following sections, we present a representative plot for a scenario/strategy combination. Full performance data from the study is available from NCHRP.



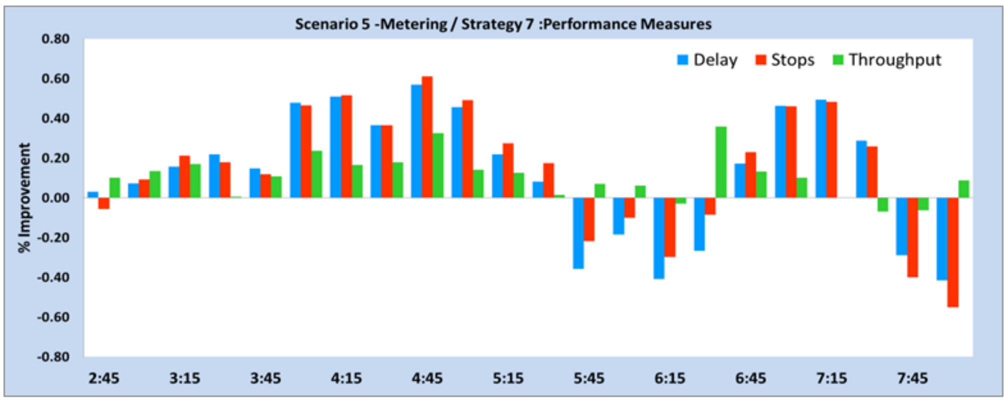


Figure 76. Scenario 5: Example performance profiles of a mitigation strategy versus the baseline timing plan

Results for the five other critical route scenarios are listed below for further illustration of the performance comparisons between the mitigations.

Scenario 1 Results

The mitigation strategies in this scenario were optimized for just two critical routes: northbound and southbound on Reston Parkway. Metering and phase reservice were not applied at Sunset Hill because those turning routes were not considered critical in this scenario. Therefore only 12 timing strategies are considered instead of 24 as in Scenario 5. The Pareto front results are presented in Figure 77.

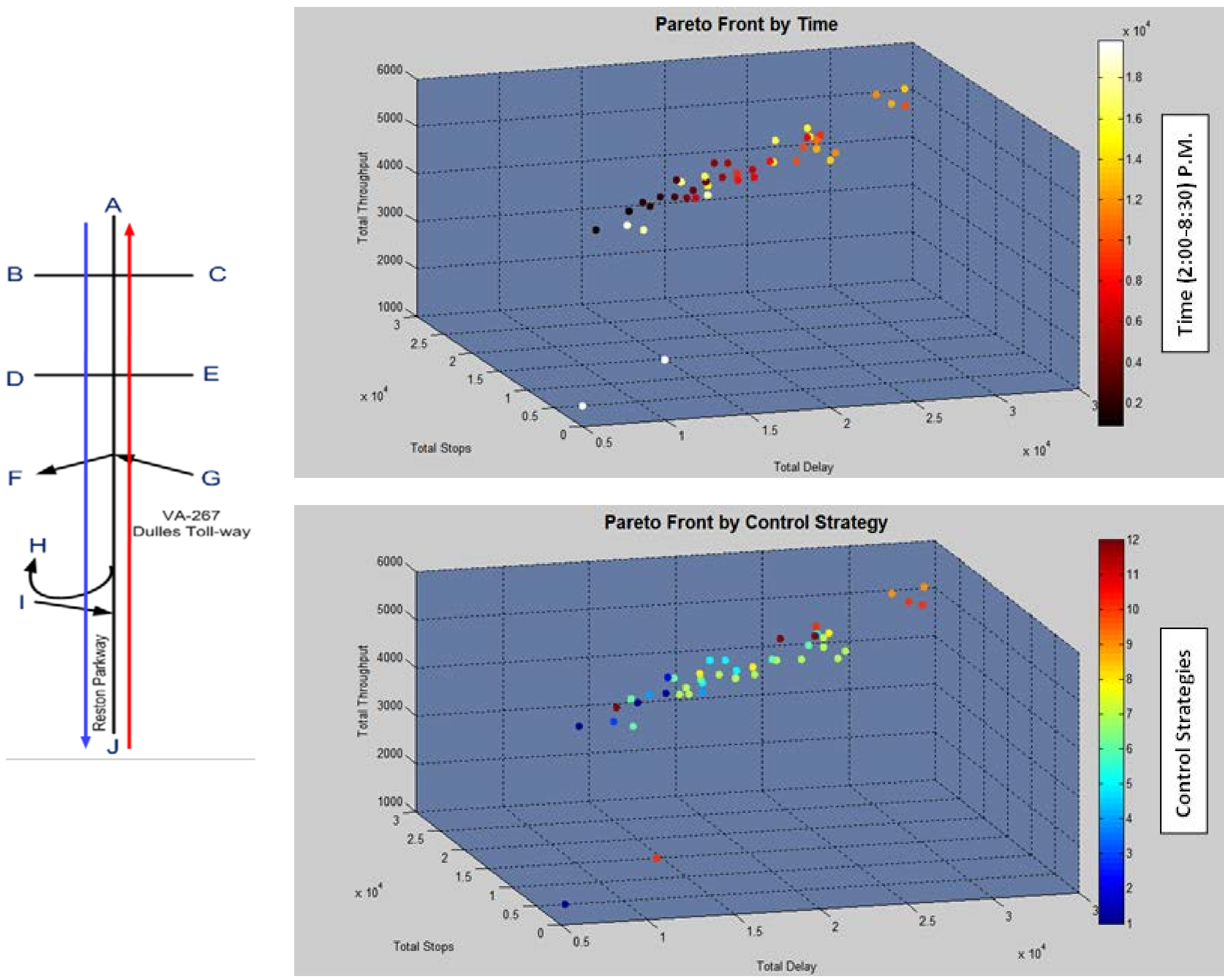


Figure 77. Scenario 1: non-dominated strategies

Routing Scenario 1 findings:

- Medium cycles with min/mid offsets maximize throughput
- Medium cycles with min/mid offsets dominate most other strategies
- During all three regimes, throughput maximization is the optimal control objective
- Throughput maximization strategies reduce delay by 16% and increase throughput by 13%
- Delay minimization strategies reduce delay by 41% , but reduced throughput by 4%

Table 18. Scenario 1: total improvement over the baseline strategy

Strategy	cycle	Delay	Stops	Throughput
1	100	0.39	0.26	-0.04
2		0.22	0.2	-0.04
3		0.41	0.2	-0.04
4		0.39	0.25	-0.04
5	140	0.16	0.2	0.13
6		0.16	0.19	0.12
7		0.22	0.25	0.11
8		0.21	0.23	0.11
9	180	0.06	0.16	0.08
10		0.06	0.16	0.08
11		0.03	0.13	0.08
12		-0.07	0	0.12

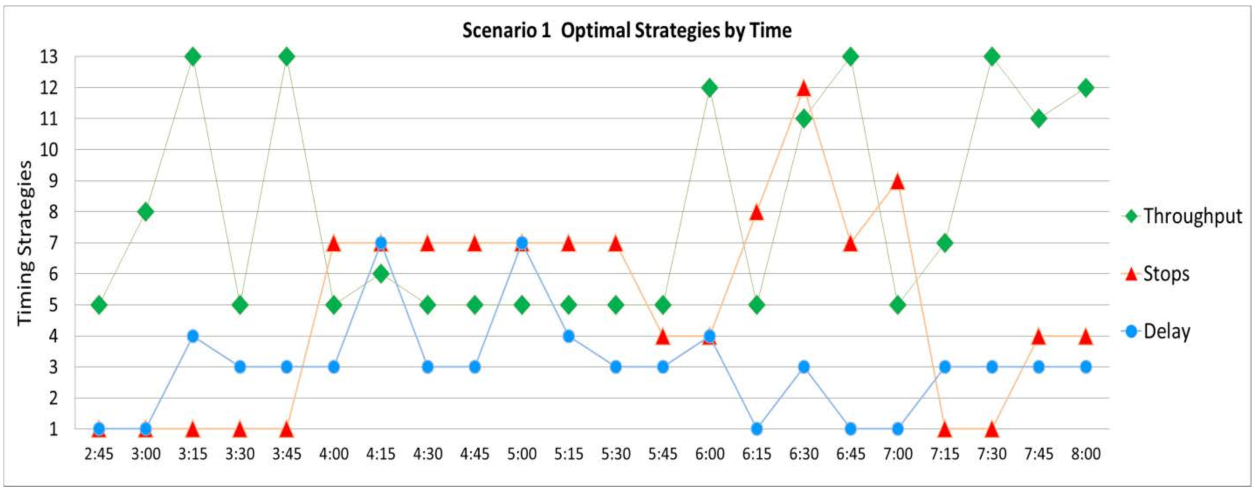


Figure 78. Scenario 1: optimal control strategies by time

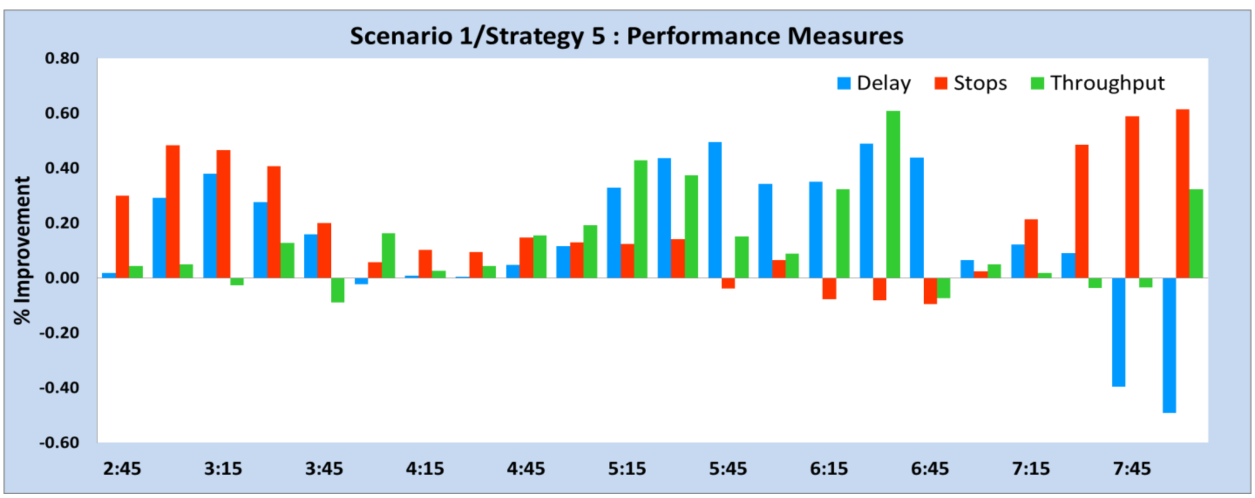


Figure 79. Scenario 1: example Strategy 5 improvement percentage over the baseline

Scenario 2 Results

The mitigation strategies in this scenario were optimized for just one critical route: southbound on Reston Parkway. Metering and phase reservice were not applied at Sunset Hill because those turning routes were not considered critical. Therefore only 12 timing strategies are considered instead of 24 as in Scenario 5. The Pareto front results are presented in Figure 80.

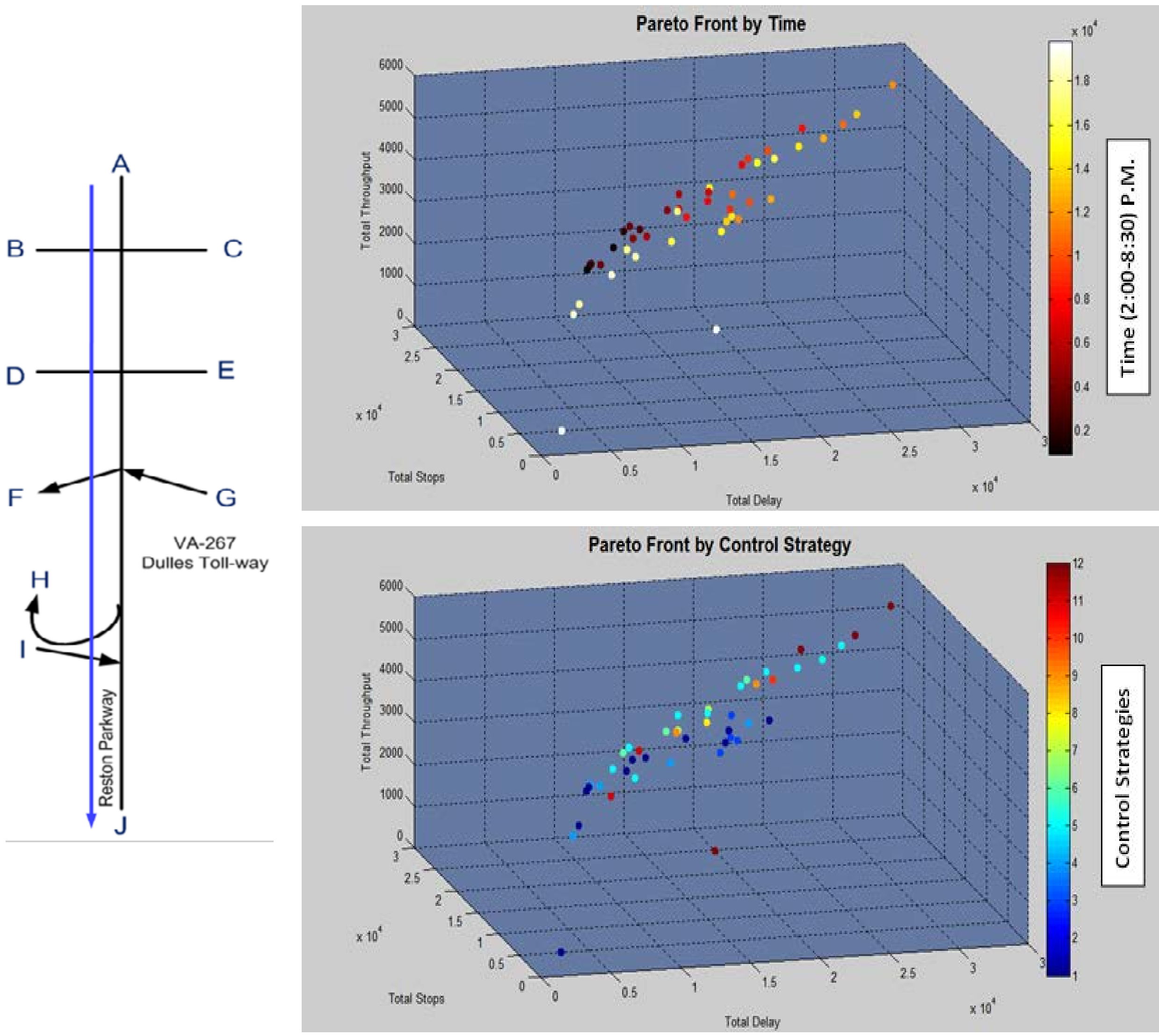


Figure 80. Scenario 2: Pareto fronts

Routing Scenario 2 findings:

- Short and medium cycles with min/mid offset maximize throughput
- During the processing regime, throughput maximization is the optimal control objective
- During the loading and recovery regimes, delay minimization is the optimal control objective
- Throughput maximization strategies reduce delay by 13% and increase throughput by 11%
- Delay minimization strategies reduce delay by 35% and increase throughput by 7%

Table 19. Scenario 2: total % improvement over the baseline plan

Strategy	Cycle	Delay	Stops	Throughput
1	100	0.35	0.45	0.07
2		0.35	0.45	0.07
3		0.34	0.42	0.08
4		0.33	0.42	0.06
5	140	0.06	0.07	0.1
6		0.17	0.22	0.1
7		0.13	0.15	0.11
8		0.14	0.18	0.09
9	180	-0.18	-0.25	0.11
10		-0.15	-0.19	0.08
11		-0.24	-0.29	0.06
12		-0.24	-0.29	0.06

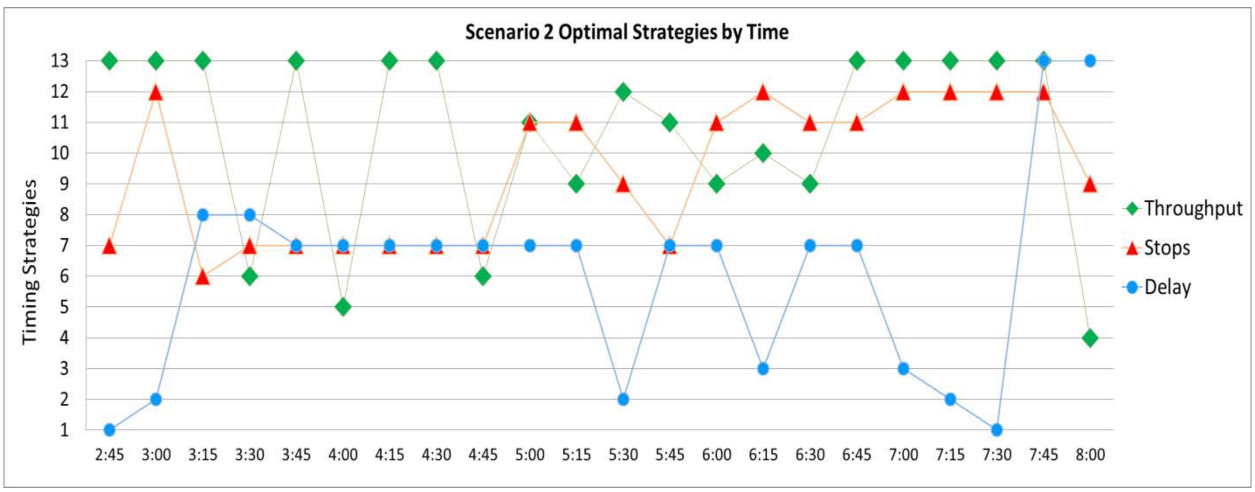


Figure 81. Scenario 2: optimal control strategies by time

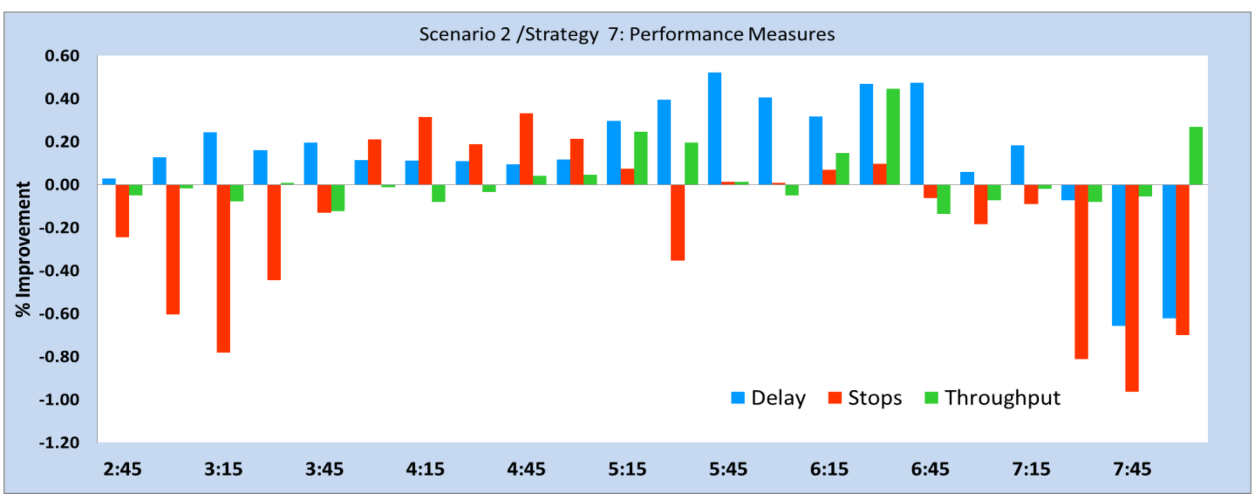


Figure 82. Scenario 2: strategy 7 improvement % by time

Scenario 3 Results

The mitigation strategies in this scenario were optimized for just one critical route: northbound on Reston Parkway. Metering and phase reservice were not applied at Sunset Hill because those turning routes were not considered critical in this scenario. Therefore only 12 timing strategies are considered instead of 24 as in Scenario 5. The Pareto front results are presented in Figure 83.

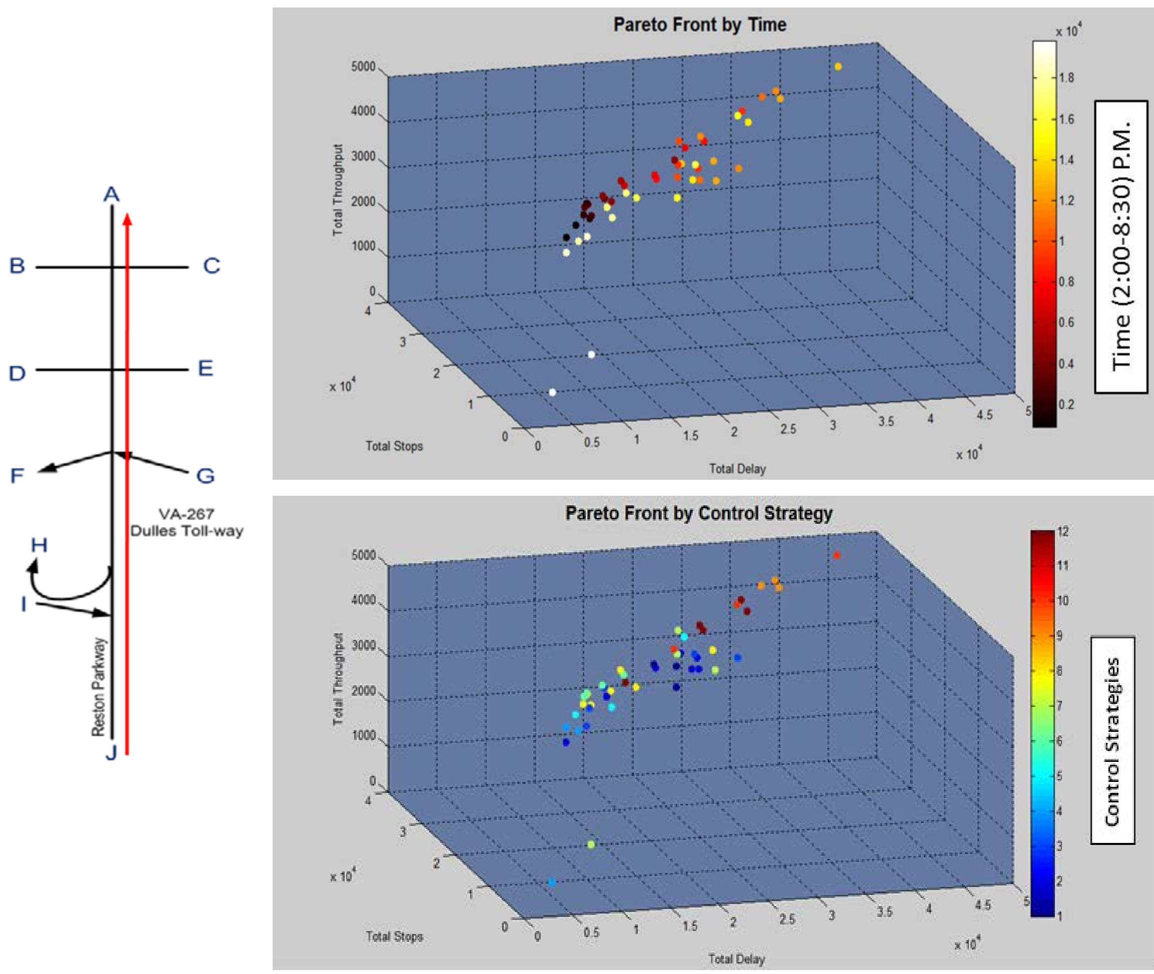


Figure 83. Scenario 3: Pareto fronts

Routing scenario 3 findings:

- Long cycles with min/ mid offset maximize throughput
- During the processing regime, throughput maximization is the optimal control objective
- During the loading and recovery regimes, delay minimization is the optimal control objective
- Throughput maximization strategies reduce delay by 21% and increase throughput by 13%
- Delay minimization strategies reduce delay by 49% , and increase throughput by 9%

Table 20. Scenario 3: total % improvement over baseline plan

Strategy	Cycle	Delay	Stops	Throughput
1	100	0.49	0.57	0.09
2		0.48	0.56	0.09
3		0.49	0.56	0.09
4		0.51	0.58	0.09
5	140	0.43	0.47	0.09
6		0.45	0.49	0.08
7		0.20	0.25	0.13
8		0.14	0.21	0.12
9	180	0.20	0.25	0.13
10		0.14	0.21	0.12
11		0.21	0.27	0.13
12		0.13	0.17	0.12

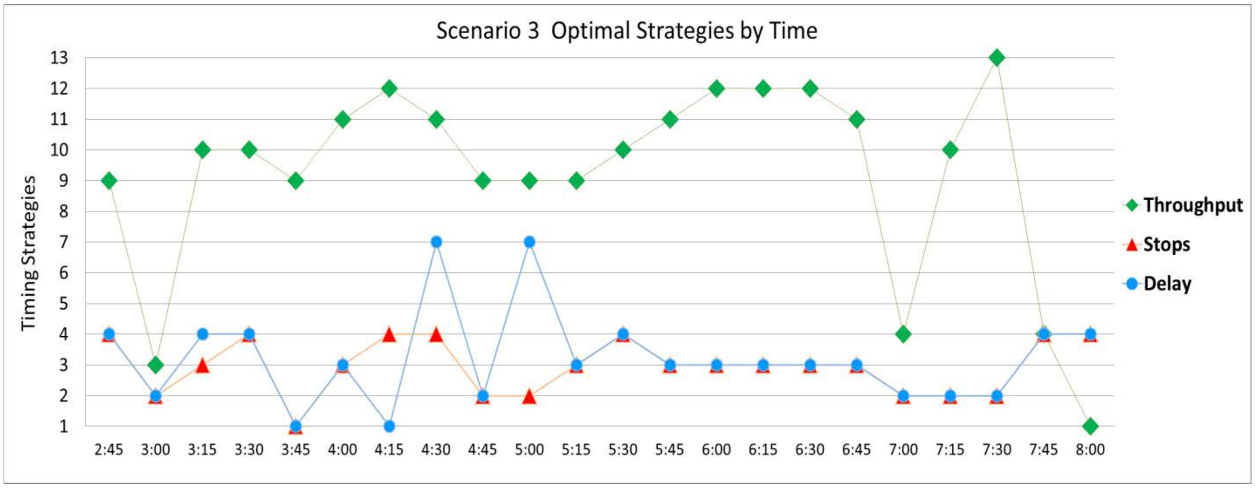


Figure 84. Scenario 3: optimal control strategies by time

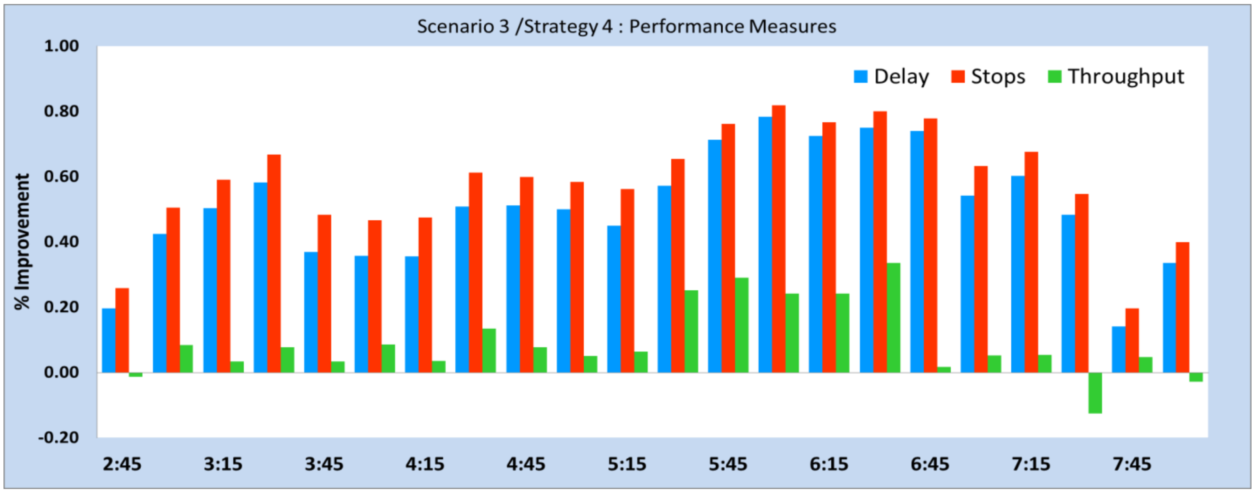


Figure 85. Scenario 3: Strategy 4 improvement % by time for performance measures (delay, stop, and throughput)

Scenario 4 Results

The mitigation strategies in this scenario were optimized for five critical routes: northbound and southbound on Reston Parkway, eastbound and westbound at Sunset Hill Road, and the left turn from Sunset Hill Road on to the toll road. Metering and phase reservice were not applied at Sunset Hill because those turning routes were not considered critical. Therefore only 12 timing strategies are considered instead of 24 as in Scenario 5. The Pareto front results are presented in Figure 86.

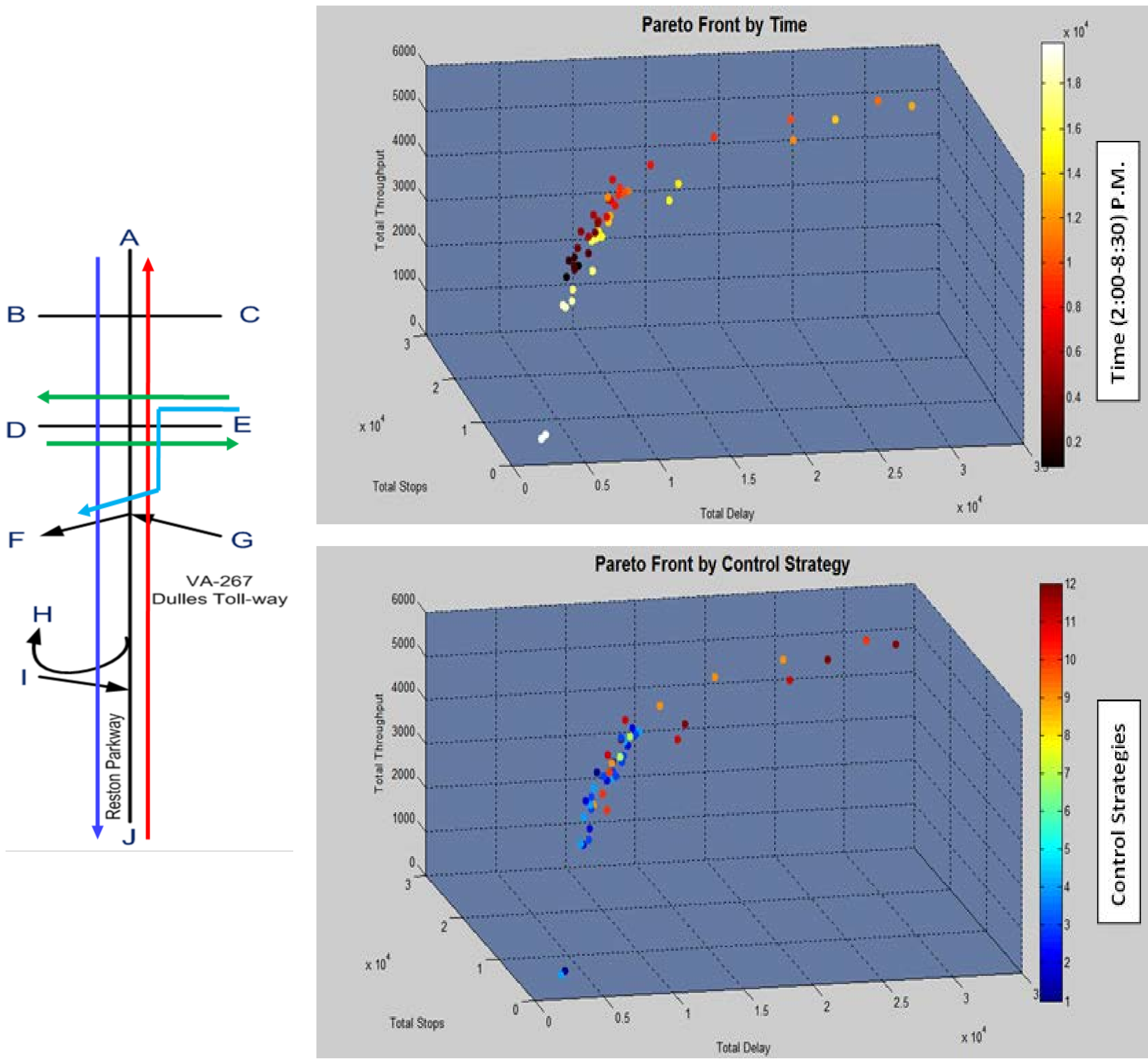


Figure 86. Scenario 4: Pareto fronts

Routing Scenario 4 findings:

- Long cycles with min/ mid offset maximize throughput
- Short cycles minimize delay during both loading and recovery regimes
- During the processing regime, throughput maximization is the optimal control objective
- During the loading and recovery regimes, delay minimization is the optimal control objective
- Throughput maximization strategies reduce delay by12% and increase throughput by 11%
- Delay minimization strategies reduce delay by 45% , and increase throughput by 8%

Table 21. Scenario 4: total improvement % over baseline plan

Strategy	Cycle	Delay	Stops	Throughput
1	100	0.08	0.24	0.07
2		0.04	0.20	0.07
3		0.07	0.22	0.07
4		0.05	0.20	0.07
5	140	0.02	0.06	0.10
6		0.45	0.49	0.08
7		0.07	0.13	0.10
8		0.12	0.19	0.11
9	180	-0.46	-0.53	0.09
10		-0.40	-0.47	0.09
11		-0.46	-0.53	0.09
12		-0.11	-0.12	0.09

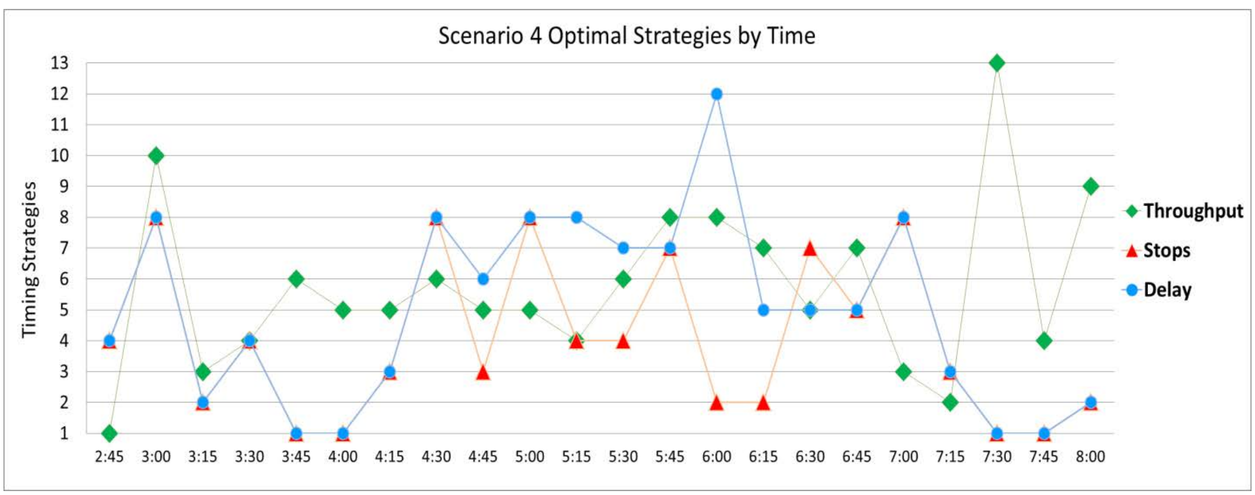


Figure 87. Scenario 4: optimal control strategies by time for performance measures (delay, stop, and throughput)

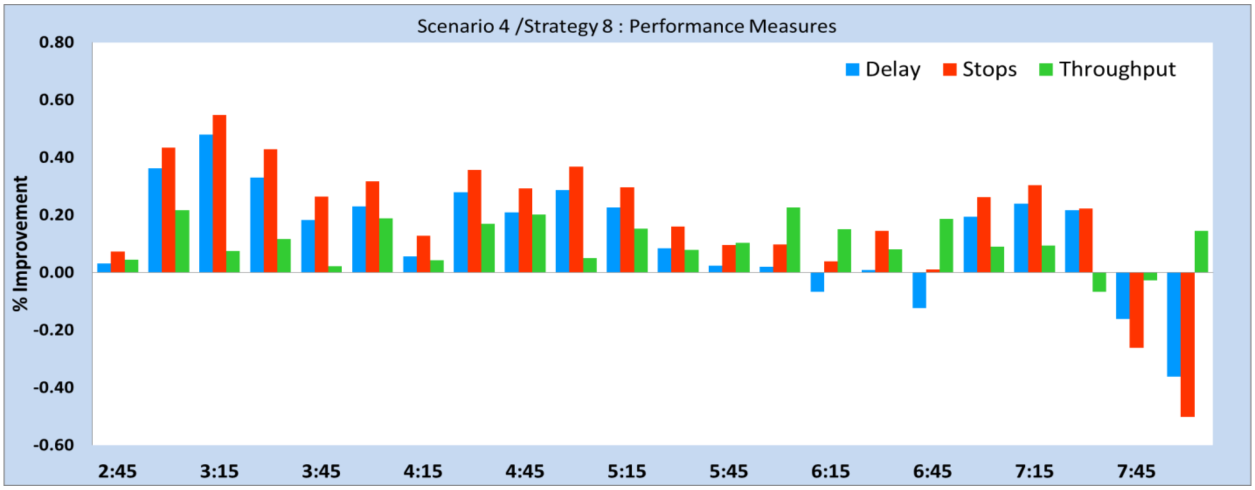


Figure 88. Scenario 4: Strategy 8 improvement % by time for performance measures (delay, stop, and throughput)

Scenario 6 Results

The mitigation strategies in this scenario were optimized for five critical routes: northbound and southbound on Reston Parkway, left and right turns from the westbound toll road off-ramp and the left turn from the eastbound toll road off-ramp on to northbound Reston Parkway. Metering and phase reservice were not applied at Sunset Hill Road because those turning routes were not considered critical. Therefore only 12 timing strategies are considered instead of 24 as in Scenario 5. The Pareto front results are presented in Figure 89.

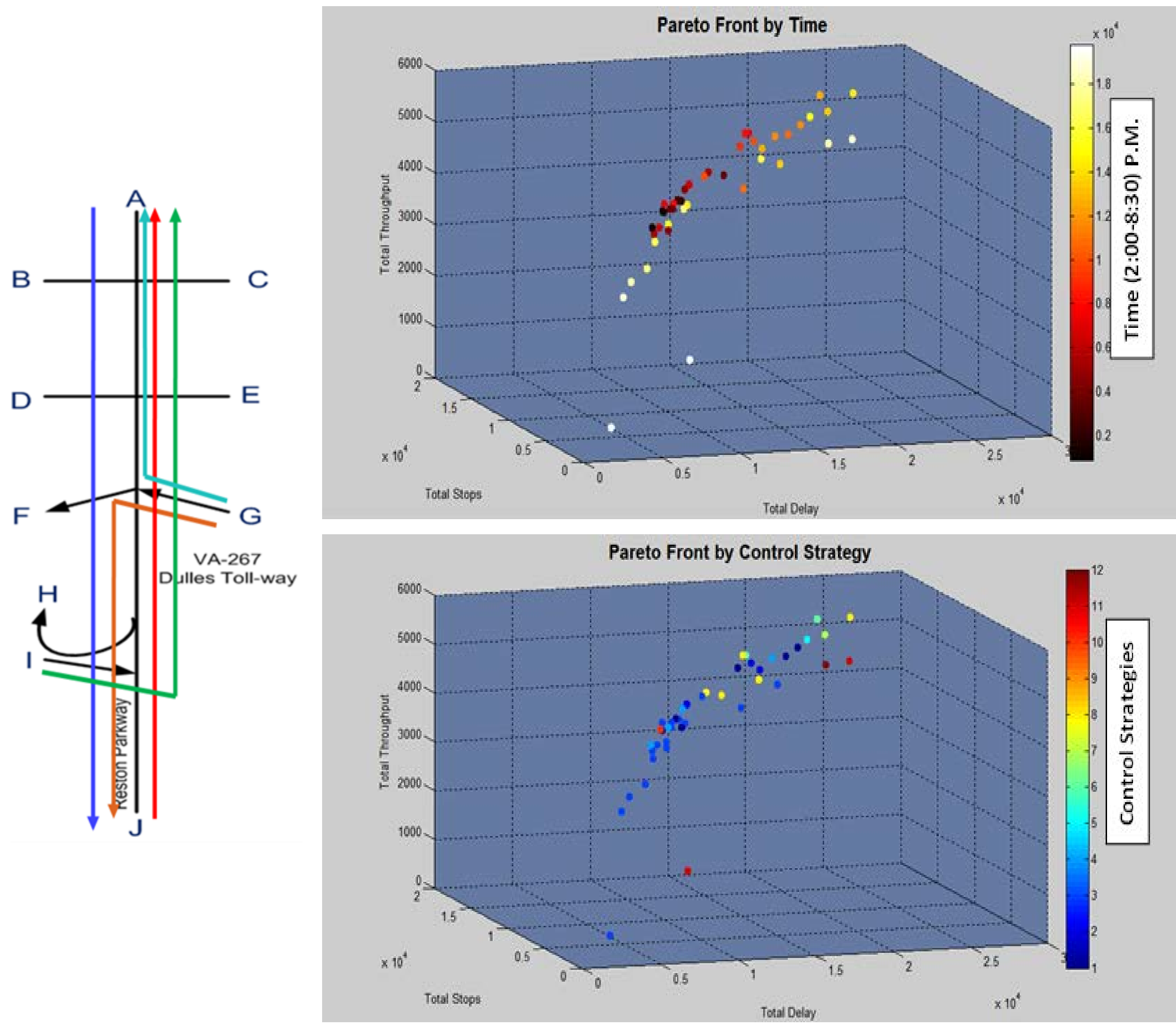


Figure 89. Scenario 6: network performance measures of the optimal control strategies during peak period

Routing Scenario 6 findings:

- Short cycles dominate the optimal solutions of this routing scenario
- During the process phase, throughput maximization is the optimal control objective
- During recovery phases, optimal solutions consist of both control objectives (delay-min and throughput-max)
- Throughput maximization strategies reduce delay by 29% and increase throughput by 19%
- Delay minimization strategies reduce delay by 38% and increase throughput by 17%

Table 22. Scenario 6: total improvement % over baseline plan

Strategy	Cycle	Delay	Stops	Throughput
1	100	0.20	0.30	0.08
2		0.19	0.27	0.09
3		0.15	0.25	0.11
4		0.18	0.28	0.07
5	140	0.29	0.33	0.19
6		0.33	0.38	0.17
7		0.38	0.42	0.17
8		0.29	0.32	0.17
9	180	0.10	0.07	0.18
10		0.15	0.12	0.19
11		0.22	0.22	0.18
12		0.15	0.12	0.19

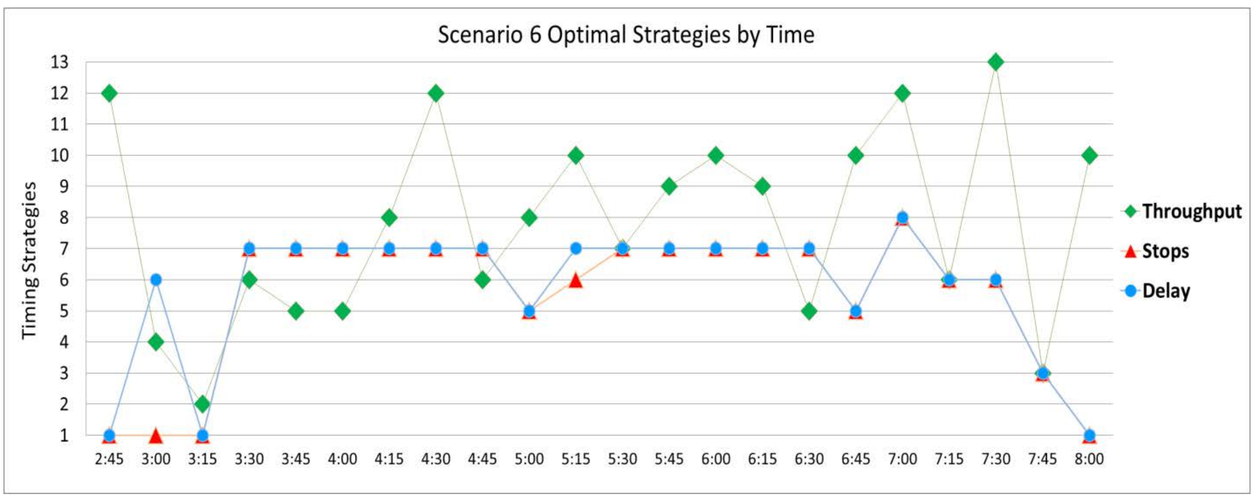


Figure 90. Scenario 6: optimal control strategies by time

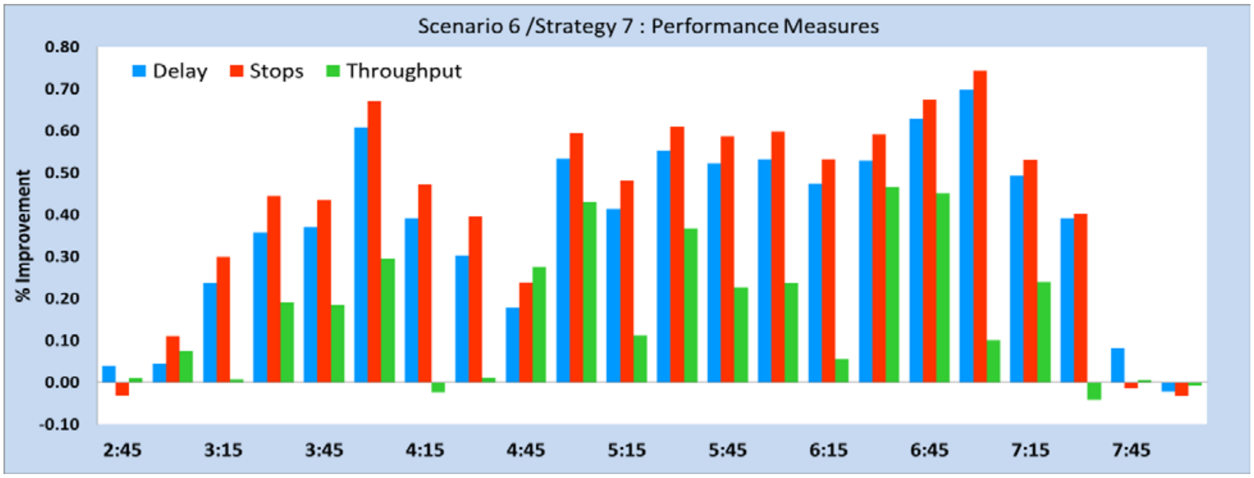


Figure 91. Scenario 6: Strategy 7 improvement % by time

Lessons Learned and Guidance from the Reston Parkway Case Study

The Reston Parkway case study was used to illustrate the process of developing timing plans for mitigating oversaturated conditions using the methodology developed in Chapter 2 and the evaluation of the performance of those plans on multiple performance measures. The following is a list of major findings from the Reston Parkway Study:

- **Critical route determination is essential for determination of optimal strategies**. A variety of results were obtained for the six different critical route scenarios that were optimized. In some of the scenarios, all of the mitigation timing plans outperformed the baseline plan which was not designed to perform for that set of critical flows.
- **Cycle length should be determined to accommodate maximum queues**. However, the findings for several of the scenarios indicate that longer cycle times result in worse performance in total delay but improve throughput. Side streets will be degraded to improve over-all system performance.
- **Offsets should be determined within a range to prevent spillback and starvation**. The performance results indicate that the procedure offered in Chapter 2 for determining offsets are effective in improving over-all system performance measures over a procedure which assumes forward progression and obtaining maximum bandwidth is an optimal method.
- **Queues are to be considered as constraints in the optimization problem; throughput and delays are objectives**. As part of the optimization approach, the queue-link ratio must be fixed to a specific value to solve for an offset/split combination that minimizes both spillback and starvation. This procedure is approximate since an average value of the queue ratio is used to solve one specific value of the offset. If the queue dissipates or grows due to fluctuation in demand, conceptually a new offset is needed for that situation. However, in solving for a fixed set of timing parameters that can be implemented with existing traffic controllers, this process is the most reasonable that could be expected.

- **During the network loading and processing regimes, optimal strategies maximized throughput. During the recovery regime optimal strategies minimized total delay**. For most of the scenarios of this test case, this trend held true. In other test cases during this project, minimize delay strategies also had merit during the loading regime and maximize throughput strategies dominated during the loading period. This test case also did not consider scheduling of plans during the three regimes, so the results assume the same timing plan is running during the entire peak period. A switching strategy is needed to enable the best timing plan at the best time, either via online logic or by time of day schedule.
- **High cycle lengths are associated with higher delay without substantially increasing total throughput**. This matches findings from other research and other tests during this project, that shorter cycles have better total throughput performance than were previously assumed because of the higher percentage of lost-time. The reduction in the interaction of spillback queues using shorter cycles seems to outweigh the penalty of more frequent phase changes.

Summary and Conclusions

In this case study we applied the signal timing plan design and analysis framework described in Chapter 2 to a real-world network. This framework utilizes knowledge of the critical routes in the network as a key first step in the design of timing plans. The framework takes into account some basic underlying constraints on cycle, offsets, and splits for minimizing detrimental effects of oversaturation (spillback and starvation). The optimization approach generates cycle, split, and offset values that minimize the degree of saturation of the approaches on the critical routes. This procedure was then applied to several scenarios on the Reston Parkway model and evaluated. Results revealed that the use of the signal timing plan design framework led to significant improvement of system performance when measured by total system throughput, total system delay, and total stops.

A three-dimensional temporal Pareto front diagram was developed to illustrate the non-dominated solutions throughout the peak period. The results indicate that since the problem is so dynamic, no one solution is optimal all of the time. This innovative type of diagramming can be used to assist in the analysis of the performance of control strategies in order to distill rational guidance for practitioners.

In the next section, we apply the process again to a more complicated test network in the Post Oak area of Houston, TX and document the results and findings.

Network Test Case: Application of the Multi-Objective Evaluation Process with Explicit Consideration of Operational Regimes

In this section, we describe a second test case in which we applied the methodology for design of timing plans that consider oversaturated conditions. The network shown in Figure 92 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 in oversaturated conditions. In the first test case, we learned that different plans have different levels of effectiveness in each of the three regimes of operation (loading, processing, and recovery). In this test case, we expanded our approach to explicitly consider the TOD schedule of the implementation of three successive timing plans during the scenario. This scenario also significantly increases the level of complexity by considering an oversaturated network with multiple, interacting critical routes.

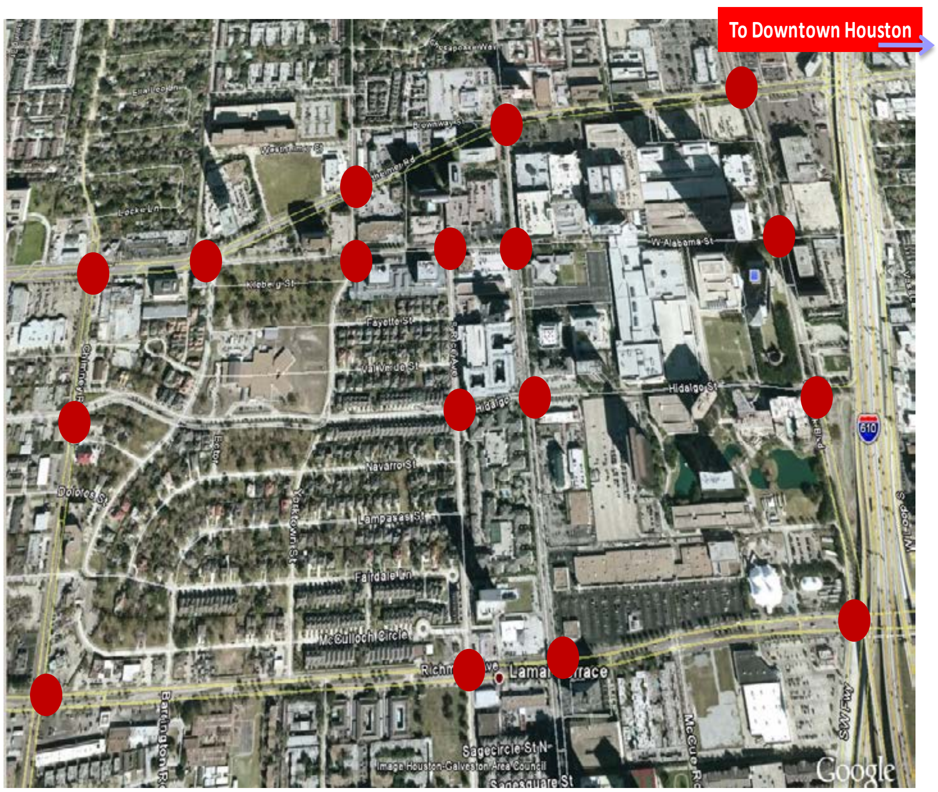


Figure 92. Post Oak area of Houston, TX

Background

Houston, Texas, is the fifth largest metropolitan area in the U.S. Traffic congestion continues to grow as the urban area expands. The Post Oak / Galleria Mall area (also referred to as Uptown) experiences serious oversaturated conditions during peak periods. Adjacent to the I-610 Loop Freeway on the west side of downtown Houston, the queuing due to on- and off-ramp traffic is significant. Westheimer Road, on the north side of the sub-network shown below, is one of the most heavily traveled arterials in the region. During freeway incidents, the area falls under additional pressure as motorists divert to arterials parallel to the I-610 to continue their commute towards I-10 and northwest Houston. The study network is adjacent to the I-610 Loop and US59 interchange as shown in Figure 93. The West Loop is Houston's second busiest freeway (after the Southwest Freeway). The I-610 is often referred to as the West Loop parking lot because of the severe congestion that occurs between I-10 and US59.

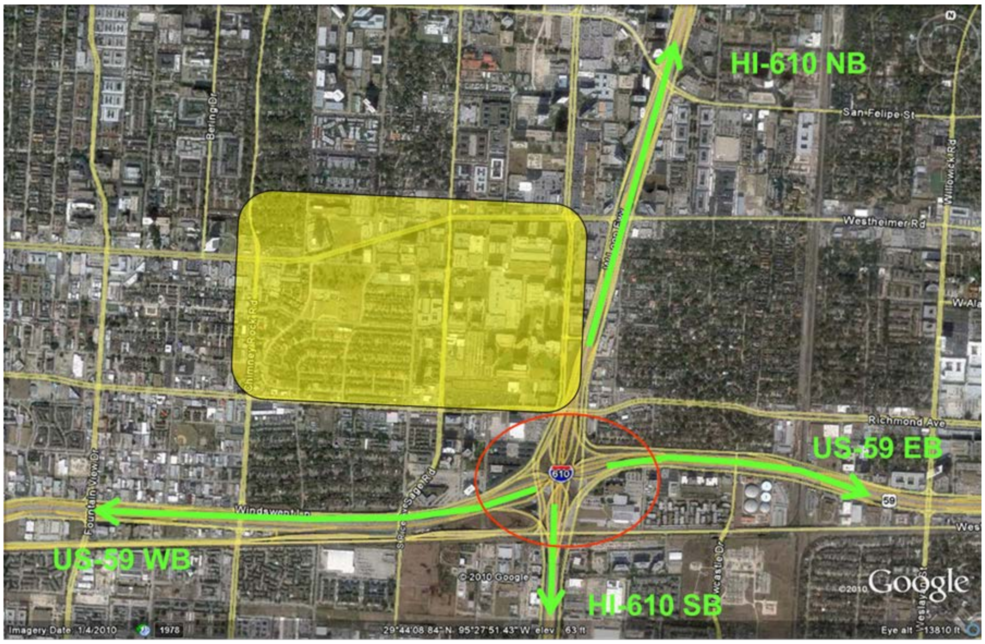


Figure 93. I-610 loop and US 59 interchange, Houston, TX

The Uptown/Galleria district of Houston, Texas, is a dynamic, urban community in Houston's West Loop, about five miles west of Downtown surrounding The Galleria, a large, upscale, indoor mall and commercial office complex.

Figure 94. Skyline view of Uptown Houston

About 42,000 residents live in the Houston Galleria's apartments, modern townhouses, and single-family homes and there is five million square feet of retail space. With the growing demand for residential property in Uptown Houston, developers have increased their activity in the area. Uptown is also host to Houston's largest hotels, which host about 20 million visitors a year. The area's commercial activities and proximity of major highways generates different traffic patterns during A.M. and P.M. peak periods.

Development of Critical Route Scenarios

The first step in the process to develop effective control strategies is to identify the critical oversaturated routes through the network. Once the critical routes, movements, and bottleneck links are identified, a wide range of control strategies are evaluated to reduce the detrimental effects of these problematic symptoms.

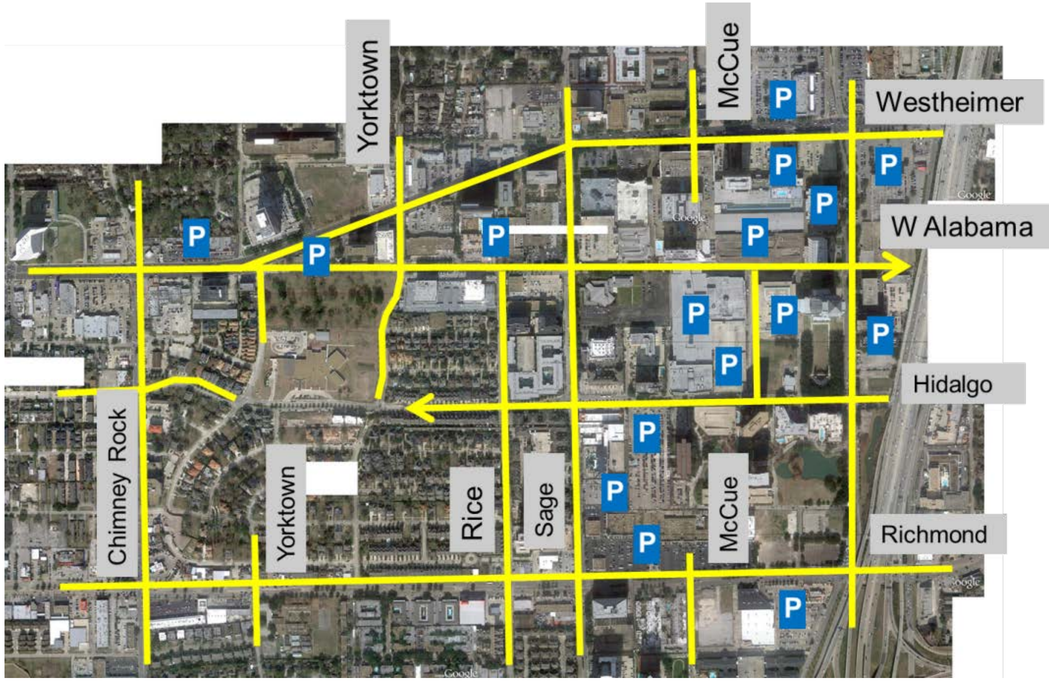


Figure 95. Parking lot facilities in Post Oak Network

During the P.M. peak period, a large number of trips are generated from inside the network, as employees leave their offices. The large number of parking lot facilities in the area (illustrated in Figure 95) and their substantial capacities are indicators to the significant contribution of these egress flows to the P.M. traffic load. In addition to these routes, traffic passes through the network east/west to and from the I-610 loop and south towards US-59. In the analysis of this network, we considered these two routing scenarios separately. First, a critical route scenario representing vehicles passing through the network was developed. Second, a critical route scenario representing the traffic flows generated from the parking garages was developed. Both scenarios were evaluated for the P.M. peak traffic conditions.

Scenario 1: Routes Passing Through the Area to Other Destinations

This scenario was developed considering that the routes passing through the network on the main arterials (i.e., Westheimer, West Alabama, Richmond Avenue, Post Oak, and Chimney Rock) are the critical routes. The proximity of the network to one of the major interchanges in Houston area results in significant oversaturation on these routes. These major routes are illustrated in Figure 96 through Figure 98. The following routes are the main access routes to I-610 and US-59:

1- EB on Westheimer Road to I-610 NB
2- WB on West Alabama Street to SB frontage road that leads to I-610 SB as well US-59 EB and WB
3- SB on Rice Avenue leads to US-59 EB and WB

4- SB on Sage Road leads to US-59 EB and WB
5- SB on Post Oak Boulevard leads to US-59 EB and WB
6- EB on Richmond Avenue to I-610 NB

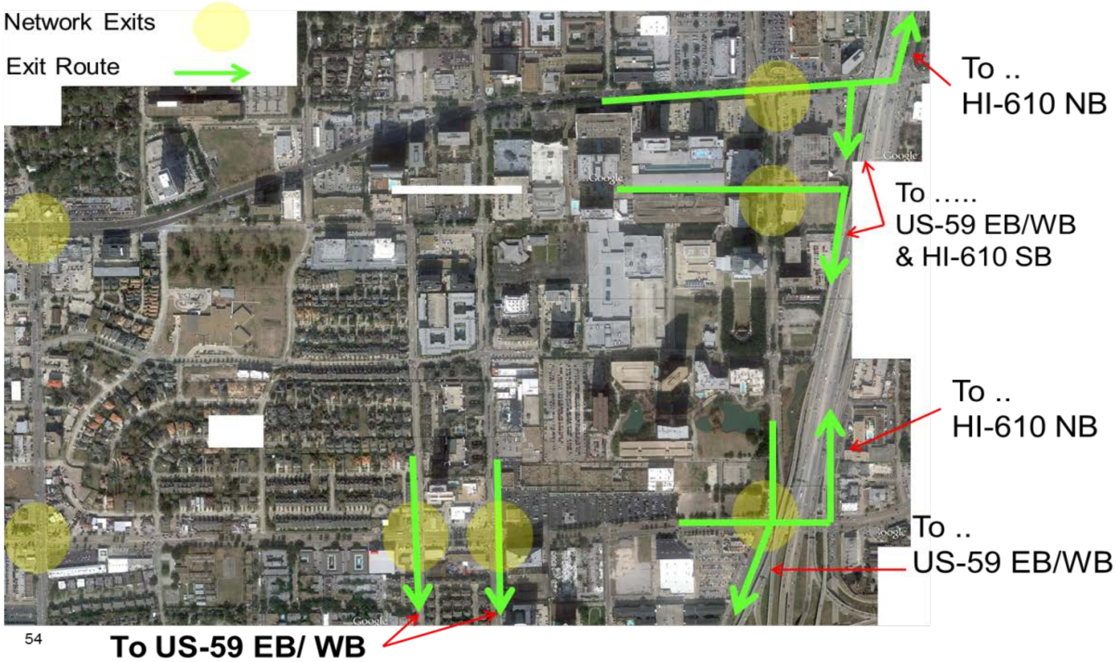


Figure 96. Network exits and routes to the highways interchange ramps

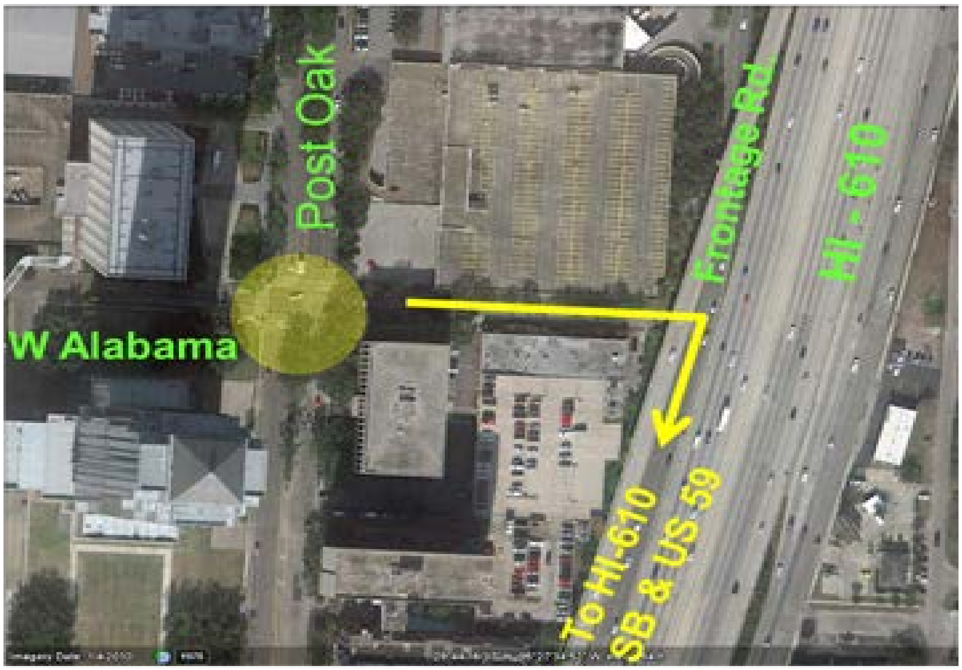


Figure 97. Post Oak Ave./W Alabama Ave.; exit to I-610 southbound and US 59 ramps

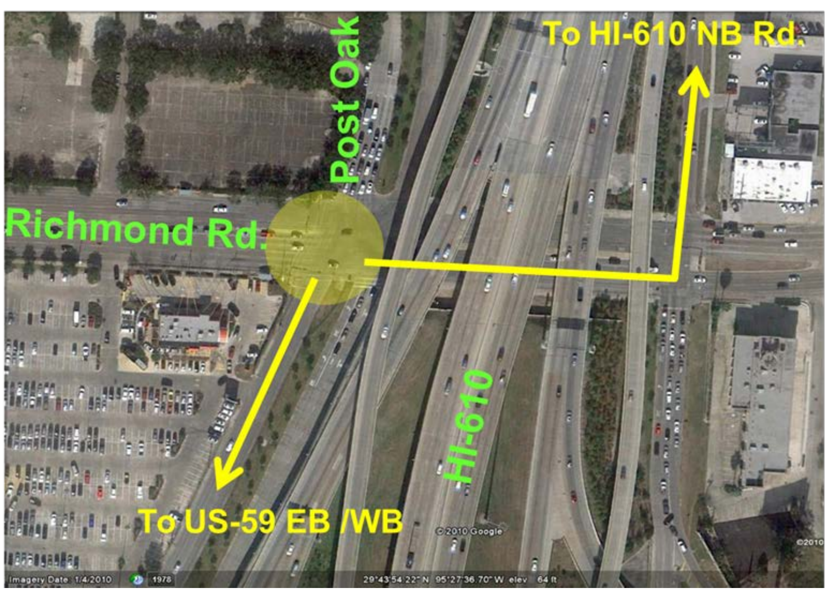


Figure 98. Richmond Ave. and Post Oak Blvd.; exits to I-610 northbound and US-59 ramps

System detector volumes were obtained for a large number of the approaches in the network from Harris County, TX. This data was analyzed to identify the critical routes in the network. To identify critical routes, we performed correlation analysis on the detector data. Correlated movements were clustered together to construct critical routes. The critical routes identified by this process were then verified and complemented with knowledge provided by local practitioners. Table 23 lists the critical routes that resulted from this process. These routes are identified graphically on Figure 99.

Table 23. Critical routes for Scenario 1

Route Number	Route Description
2	WB on Westheimer Rd.
3	WB on Westheimer Rd., left on Chimney Rock Rd., SB on Chimney Rock Rd.
4	EB on Westheimer Rd.
5	EB West Alabama St. to I-610 frontage road SB.
9	Exit from I-610 SB, LT on Post Oak Blvd, SB on Post Oak Blvd., and RT on Richmond Ave.
10	EB on Richmond Ave.
13	SB on Post Oak Blvd.
19	WB on Richmond Ave.

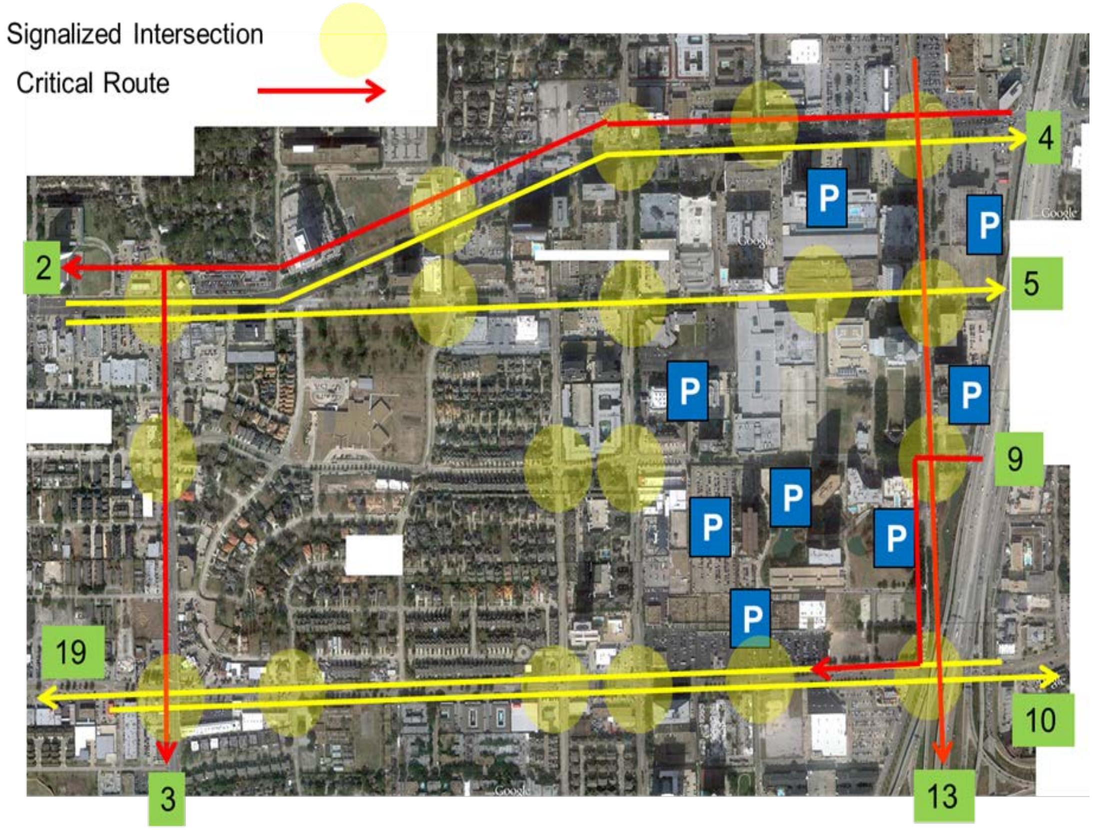


Figure 99. Critical routes for Scenario 1

Development of Arrival Demand Profiles on Critical Routes

In this step, critical routes for the scenario were assigned the maximum volumes they can handle taking into account the background traffic (traffic on non-critical routes) that also traverses the same links but with different destinations. An optimization procedure was applied to generate a feasible solution (i.e., volume profiles) for each critical route. The volume optimization process considers the minimum and maximum observed volume for each movement during all of the peak periods in the system detector data that was analyzed. The maximum possible volume is assigned to the critical route, but takes into account the background traffic flow on each link that is not traveling on the critical route. The volume profiles for each of the critical routes in this scenario are shown in Figure 100.

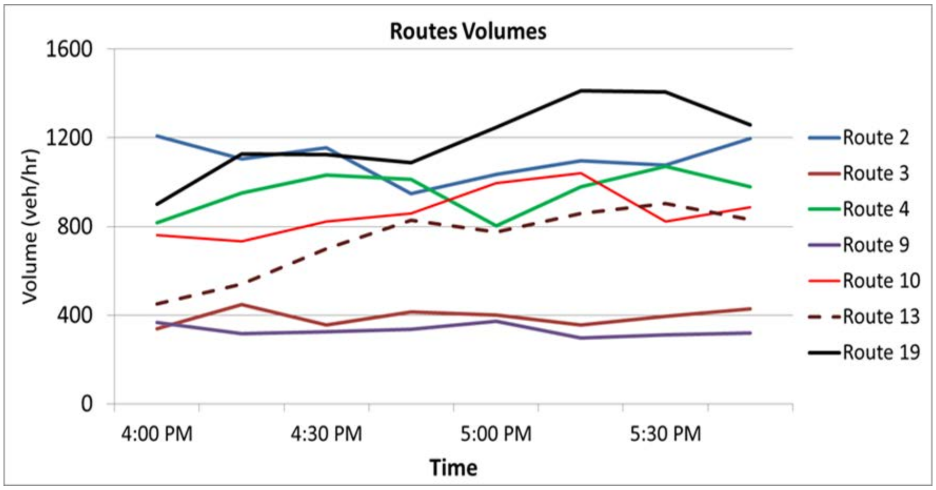


Figure 100. Volume profiles for critical routes in Scenario 1

Once the critical routes were established the corresponding critical movements throughout the network could be identified where critical routes use common links. This is illustrated for three critical routes in Figure 101. Both routes and the movements will be considered later when developing the control strategies. For each critical movement, we protect the v/c ratios to be as low as possible to ensure that the critical routes experience the minimum levels of TOSI and SOSI that are possible.

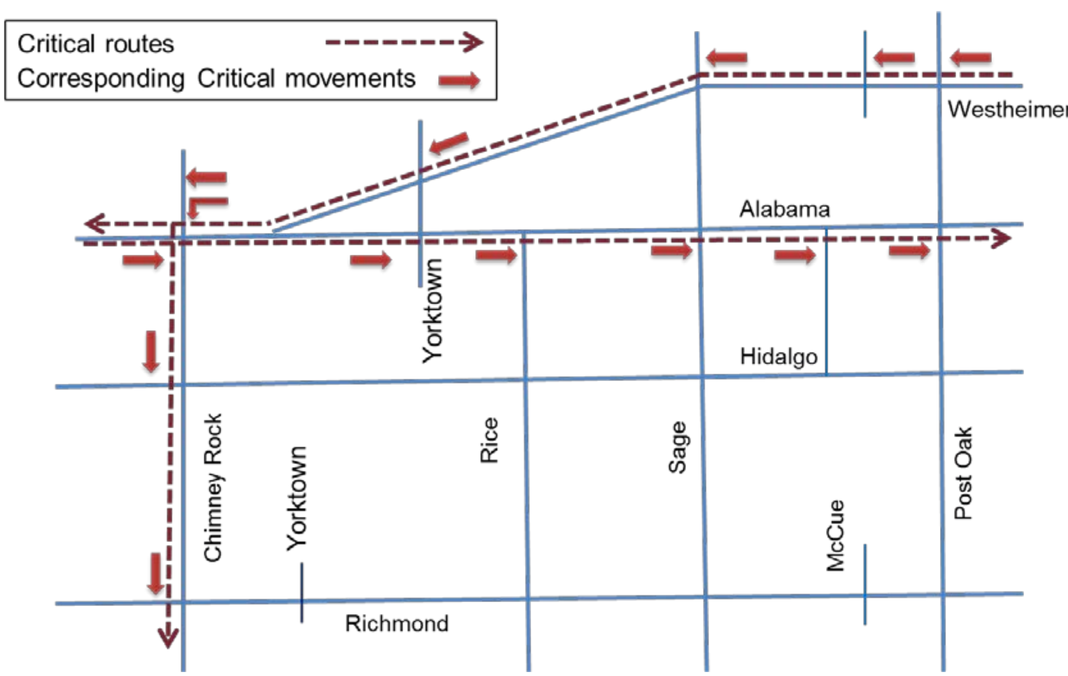


Figure 101. Critical routes 2, 3, 5 and the corresponding critical movements

Scenario 2: Critical Routes Generated from Traffic Inside the Network

In Scenario 2, traffic generated from inside the network (i.e., parking lots) that is leaving the network is considered to be the critical route flows. Seven routes were identified based on the volume correlation analysis and the views and insights of local practitioners. As in the analysis of Scenario 1, the volume analysis consisted of analyzing the correlation of detector data during the peak period. Compatible movements on a possible route that were correlated to be higher or lower during the same time periods were clustered together to form a critical route. After identifying the critical routes, volumes on these routes were maximized to the highest limits they can carry. Any remaining traffic volume from the actual count data on a link that is part of a critical route is considered as background traffic. The background traffic consists of both non-critical routes traffic and the stable fraction of the traffic on the critical routes during the congested period. The minimum and maximum observed volumes are used to establish network background traffic. Table 24 and Figure 102 describe and illustrate Scenario 2 critical routes (i.e., outbound routes).

Table 24. Critical routes for Scenario 2 (outbound routes)

Route Number	Route Description
6	NB on Sage Rd., RT EB on Westheimer Rd.
7	WB on Alabama St., LT NB on Post Oak Blvd., and RT EB on Westheimer Rd.
8	WB on Hidalgo Rd.
11	SB on Post Oak Blvd., and LT EB on Richmond Rd.
12	SB on Sage Rd., LT EB on Richmond Rd., and RT SB on Post Oak Blvd.
14	SB on Sage Rd.
20	NB on Post Oak Blvd.

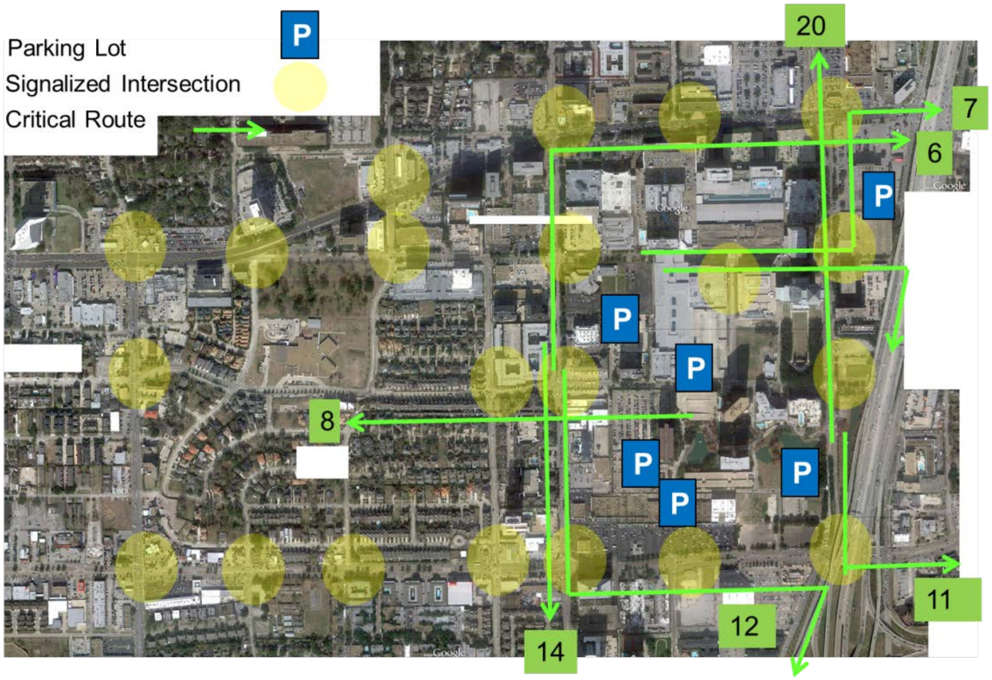


Figure 102. Critical routes for Scenario 2

The volume profiles of outbound routes (illustrated in Figure 103) show that the traffic volumes start to increase substantially around 4:30 P.M. This surge in volume is due to the trips originating from the commercial offices. Shortly thereafter, the critical routes quickly become oversaturated and remain congested until the end of the peak period after 6:00 P.M.

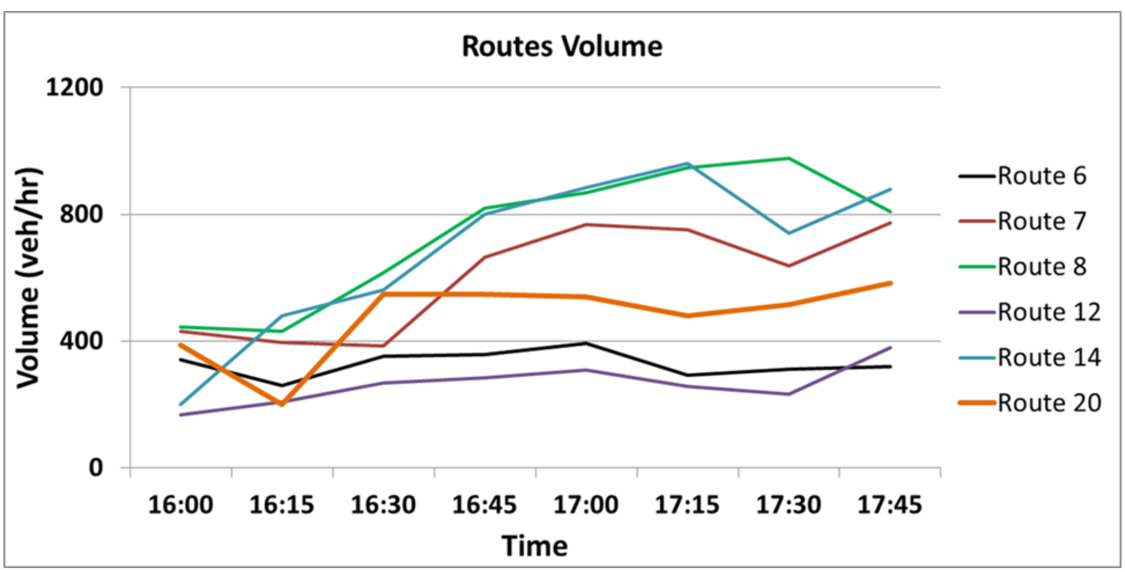


Figure 103. Volume profiles for critical routes in Scenario 2

In the next section, we discuss common issues of oversaturation in the network under both critical route assumptions. Following this discussion, we discuss the control strategies that were applied to each routing scenario, and the timing plans and parameters that were generated from applying the optimization methodology.

Development of Control Strategies for Scenarios 1 and 2

During the peak hours, the Post Oak network faces serious recurrent oversaturation due to the trips generated from the commercial office parks combined with the pass-through traffic en route to and from the freeways. The symptoms of the oversaturation are characterized by starvation, storage blocking, and entry and exit driveway blocking. Initial evaluation using Vissim simulation software indicated that the dominant symptoms affecting the traffic operation in the network were widespread spillback (SOSI > 0), overflow queue formation (TOSI >0), and secondary congestion at driveways. In particular, many left-turn storage bays spill back into the lanes for the through movements. The left-turn spillbacks are caused by inadequate green times and limited storage capacities.

Problematic Symptoms of Oversaturation in the Network

The following problematic symptoms of the oversaturation were observed in the analysis of the operation of the baseline signal timing strategy:

- Spillback and de facto red
  - Intersection of Richmond Ave. and Sage Rd.
  - Intersection of Richmond Ave. and Rice Ave.
  - Intersection of Yorktown and West Alabama St.
- Queue formation
  - WB and SB at Intersection of Post Oak Blvd. and West Alabama St.
  - SB at Intersection of Richmond Ave. and Sage Rd.
  - SB and NB at Intersection of West Alabama St and Sage Rd.
  - EB on Westheimer @ Post Oak intersection, McCue intersection, and Chimney Rock.
- Storage blockage
  - W. Alabama
- Entry drive blockage on:
  - W. Alabama St. between Sage Rd. and Post Oak Blvd.
  - EB Westheimer Rd. between Yorktown and Sage Rd.
  - Post Oak between Hidalgo and Richmond

These symptoms are illustrated in Figure 104 and Figure 105 below.

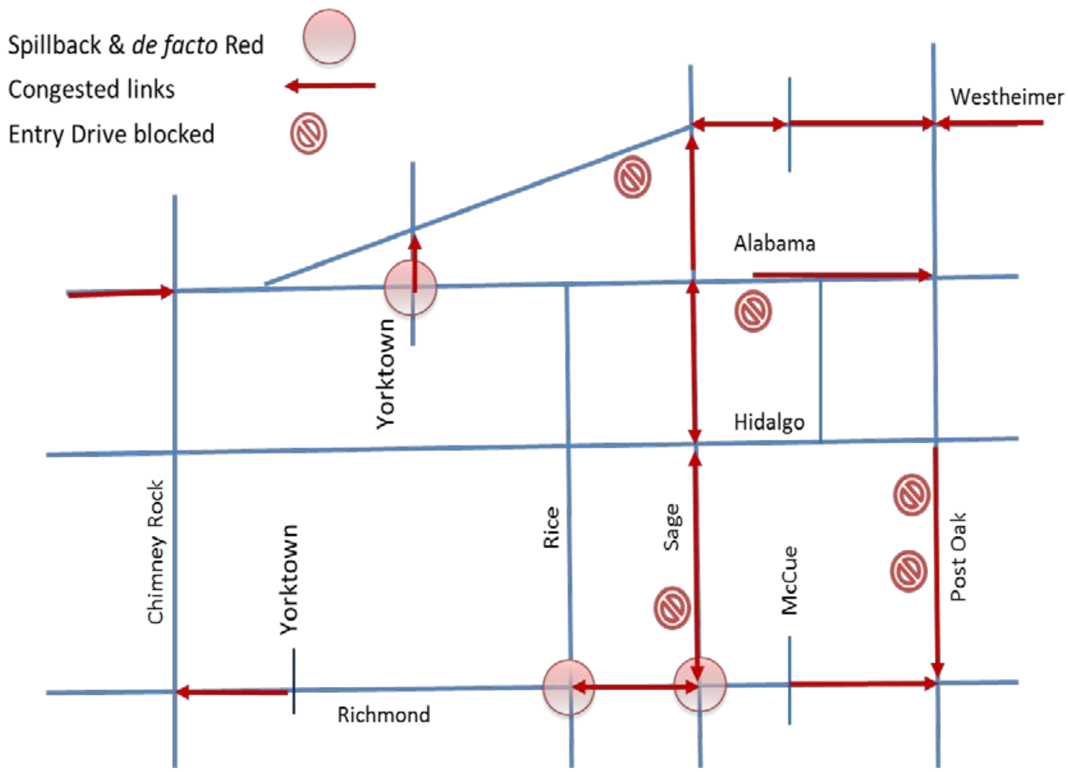


Figure 104. Symptoms of oversaturation in the Post Oak network

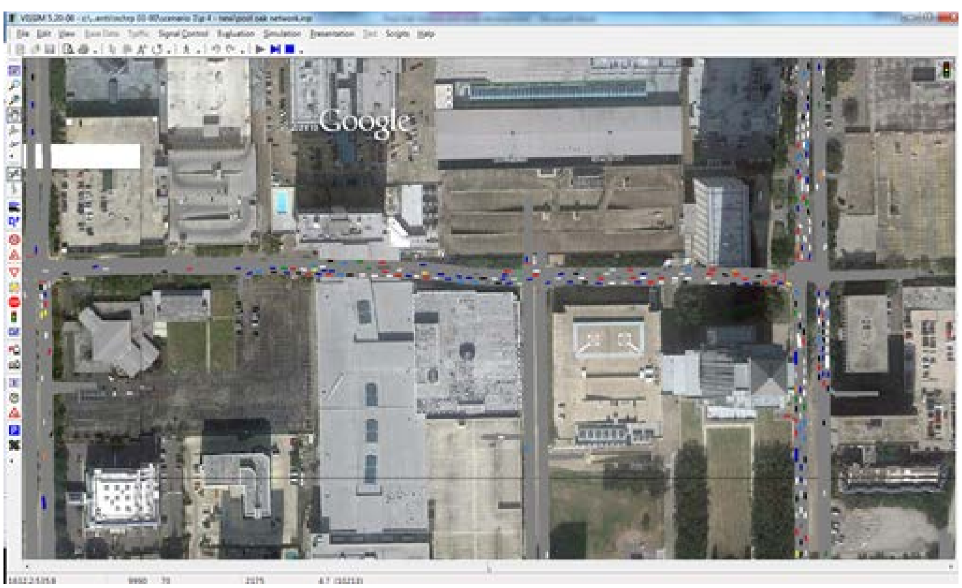


Figure 105. Simulation snapshot showing queue spillback along a critical route

Since persistent, recurring spillback of left-turn bays was identified as one of the dominating factors influencing the operation during the peak period, it was essential to develop a control strategy that explicitly addresses this problem. The left-turn phase reservice strategy was chosen to provide the extra capacity to the left-turn movements. To prevent spillback on short links, we also chose to test the use of the strategy to flare the green. Flaring the green provides extra capacity to downstream intersections in order to clear existing queues by restricting the upstream green times and holding those vehicles in upstream queues.

The arterials on the perimeter of this network are characterized by long spacing between intersections and thus high capacities to store vehicles upstream of the bottleneck areas. While the spacing on the perimeter of the network allows the use of large cycle lengths, the implementation of such large cycles would not be feasible for the interior intersections that are much more closely spaced. In particular, applying longer cycles in the interior locations would cause build-up of queues that would block the many entry and exit driveways from the commercial buildings.

Therefore, we determined that a critical component of the mitigation strategies would be to run interior and closely-spaced intersections on half cycles. Table 25 and Figure 106 summarizes the control strategies that were tested during the queue management regime, which intersections these strategies were applied, the improvements expected from applying the strategy, and the operational objectives of each method. These mitigation strategies were applied only during the processing regime.

Table 25. Control strategies applied in Scenario 1

Control Strategy	Location	Expected Improvement	Operational Objective
Left-turn phase Reservice	WBL Richmond Ave. & Rice Ave.	Prevent storage blocking	Queue management
	NBL Yorktown & Westheimer Rd.		
"Flare the green"	EB on W. Alabama St.	Clear downstream queues	Queue management
	WB on Hidalgo St.	Improve progression	&
	WB on Richmond Ave.		Throughput maximization
Half-Cycling	W Alabama St. & Yorktown	Improve green utilization	Queue management
	W Alabama St. & Sage Rd.	Prevent starvation	&
	W Alabama St. & McCue St.	Prevent secondary congestion	Delay minimization
	Hidalgo St. & Sage Rd.		
	Hidalgo St. & Rice Ave.		
	Richmond Ave. & Rice Ave.		
	Richmond Ave. & Sage St.		
Right turn overlap	Post Oak & Richmond Post Oak & Westheimer	Prevent storage blocking	Queue management
Gating			

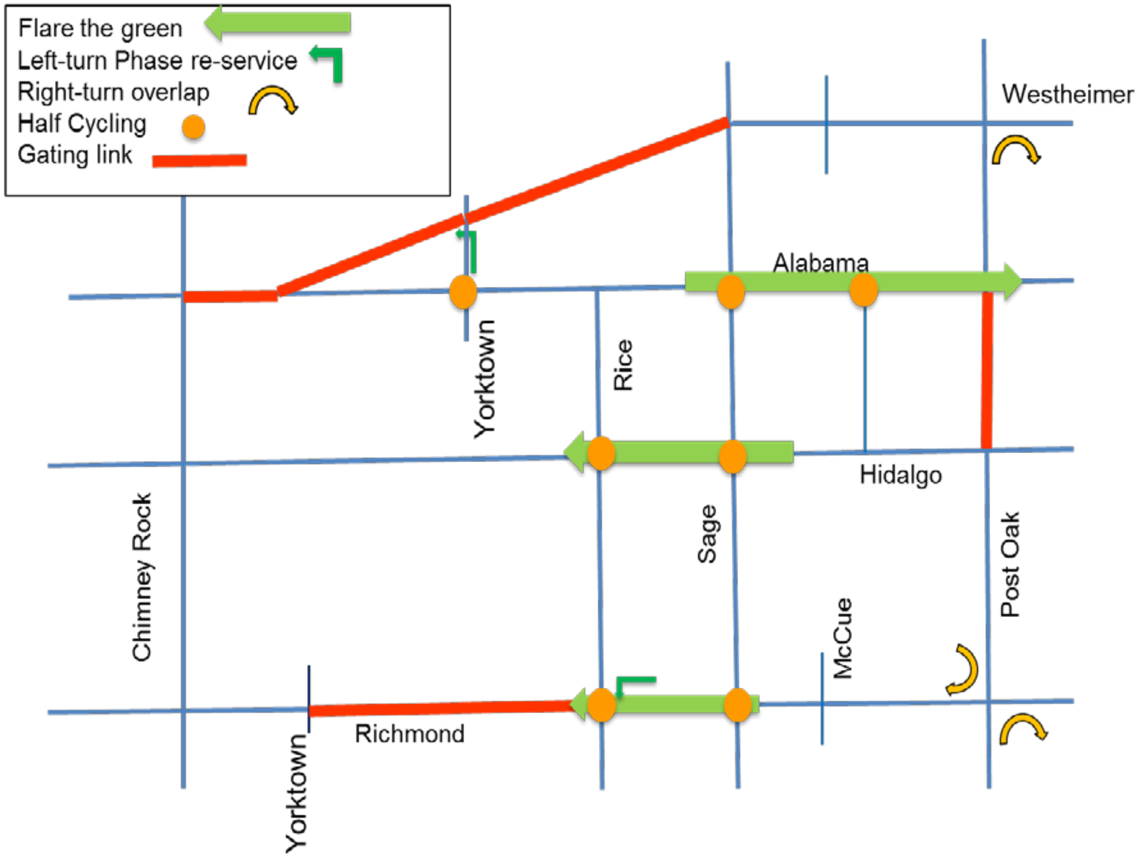


Figure 106. Control strategies applied in Scenario 1

Scenario 1: Timing Plan Development

Two strategies were generated based on the assumptions of Scenario 1 that routes through the network are the critical routes. Both strategies explicitly account for the three regimes of oversaturated operations (loading, processing, and recovery). The objectives of the first timing strategy were to minimize delay during the loading regime, implement queue control strategies during the processing regime, and then maximize throughput during the recovery regime. This is denoted later as (min delay-manage queues-max throughput). A second timing strategy was developed to maximize throughput during both the loading and recovery regimes and manage queues during the processing regime. This strategy is denoted as (max throughput-manage queues-max throughput). These strategy combinations were then tested in simulation.

The cycle, splits, and offsets of the timing plans during each of these regimes were generated using the timing plan development framework presented in Chapter 2. Since the volume profiles for critical movements were used to generate timing plans instead of the movements with the highest volumes, multiple timing plans were generated for each time bin in the volume profile (i.e., a total of 17 plans for each control objective resulting in a total of 51 timing plans that were evaluated in this procedure).

An optimization procedure was then conducted to identify the optimal timing plans and their optimal switching points to meet the assumed control objectives in each regime and duration constraints for each timing plan. Figures 107 and 108 illustrate the resulting schedules of timing plans and switching times between the plans in each strategy. Table 26 describes the characteristics of the two strategies.

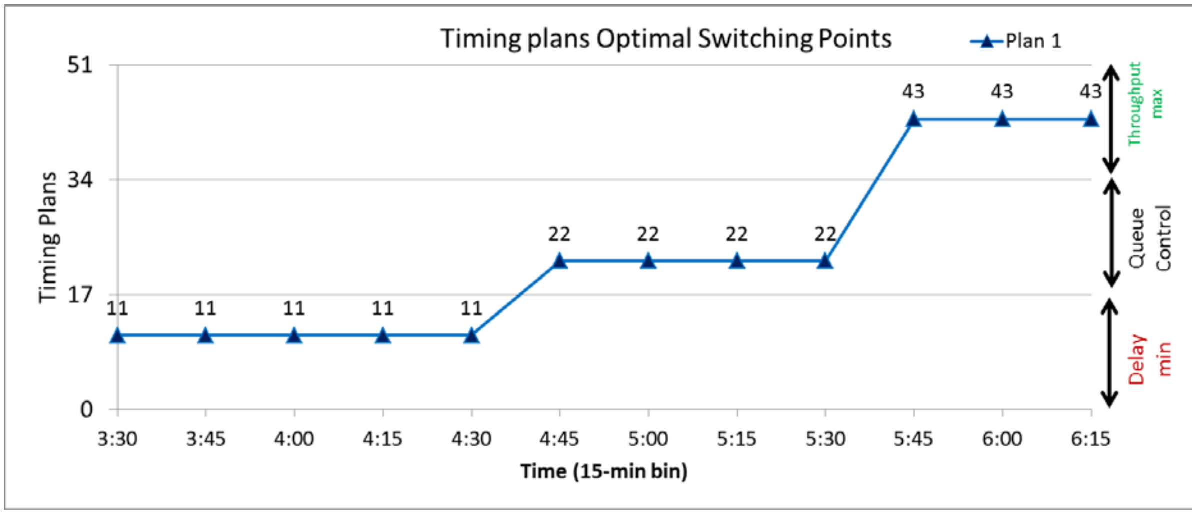


Figure 107. Min Delay-Queue Management-Max Throughput Strategy (Strategy 1)

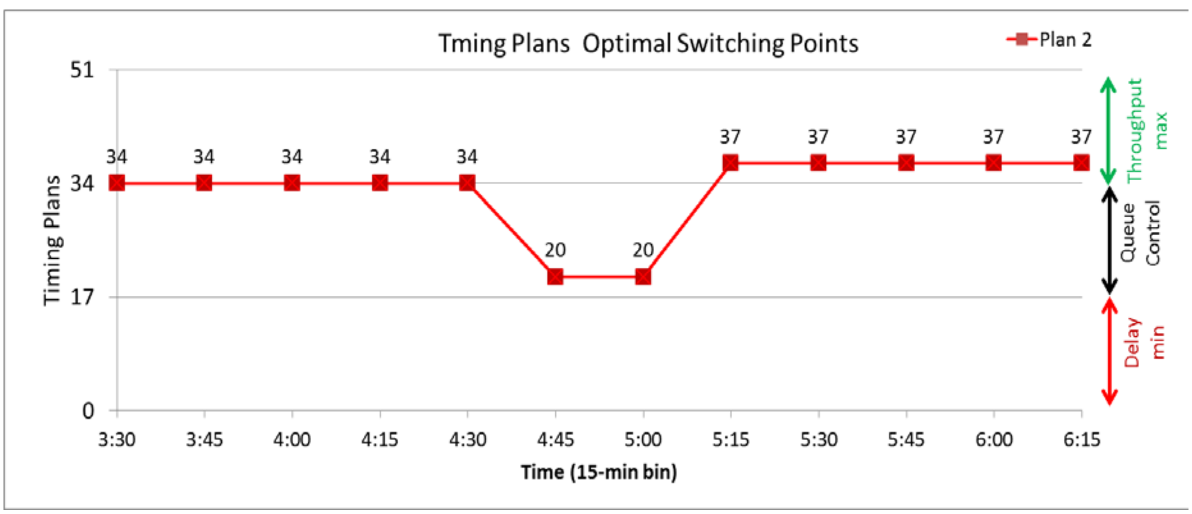


Figure 108. Max Throughput-Queue Management- Max Throughput Strategy (Strategy 2)

Table 26. Description of strategies

Strategy No.	Loading Regime			Processing Regime			Recovery Regime		
	Operation Objective	Timing Plan #	Starting time	Operation Objective	Plan #	Starting time	Operation Objective	Plan #	Starting time
1	Delay min	11	3:30	Queue management	22	4:30	Throughput max	43	5:30
2	Throughput Max	34	3:30	Queue management	20	4:30	Throughput max	37	5:00

Timing Plans (1-17) are plans that are designed for the delay-minimization objective, timing Plans (18-34) are plans that are designed for the queue management objective, and Plans (35-51) are plans that are designed for the throughput maximization objective. Table 27 shows the cycle lengths and timing plans numbers used in each control strategy.

Table 27. Cycle lengths used in each strategy

	Timing Plan #	Cycle length (sec)
Strategy 1	11	150
	22	90
	43	160/80
Strategy 2	34	120
	20	100/50
	37	160/80

Scenario 1: Simulation Experiment

The Post Oak network was coded in Vissim. The ring barrier controller (RBC) was used to operate all of the intersections in the network and implement the control strategies. Data collection points were placed at all of the entry and exit points and along the critical routes in the network. A 30-minute warm-up period was used before data collection was started. Thirty minutes of simulation time after the peak period was added as well to clear the network of queues at the end of the simulation. Each control strategy was evaluated via five simulation runs with different seed numbers. Figure 109 shows a snapshot of the coded network. Vissim generates the following performance parameters that were used to evaluate the control strategies:

- System total delay
- Average delay per vehicle
- System total stopped delay
- Average stopped delay per vehicle
- Average number of stops per vehicles
- Average speed
- Number of stops
- Intersection throughput
- System traffic load

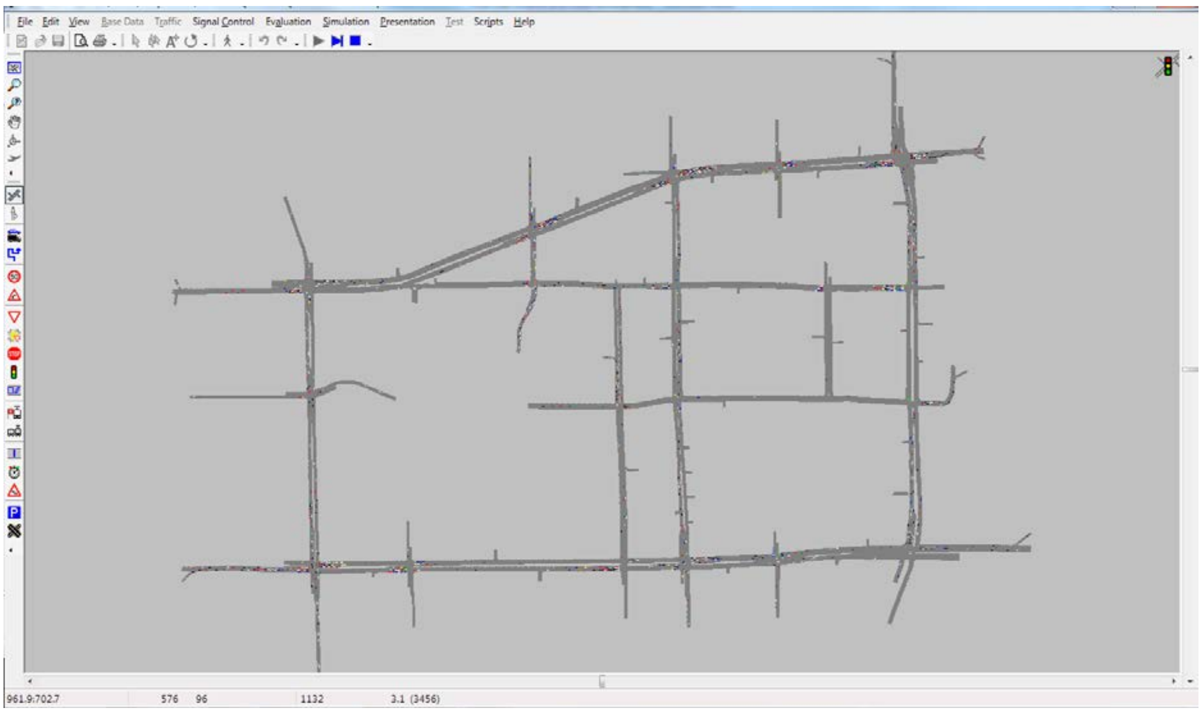


Figure 109. Post Oak network modeled in Vissim

Scenario 1: Results for Critical Routes Passing Through the Network

The results of Strategies 1 and 2 are presented in Table 28. The average value and standard deviation (expressed as a percentage) for each high-level summary performance measures are presented in the table. Modest improvements over the baseline timing plan are obtained for most of the performance measures with only Strategy 2 (max throughput-manage queues-max throughput) resulting in a higher number of average and total stops.

Table 28. Performance evaluation of strategies on Scenario 1

	Baseline		Strategy 1		Strategy 2	
Performance Measure	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
Total delay time [hr]	3772.57	3.27%	3327.30	2.47%	3626.78	3.07%
Average delay time per vehicle [s]	227.76	2.02%	208.77	3.81%	219.06	3.31%
Total stopped delay [hr]	2645.00	4.36%	2420.48	2.02%	2448.31	3.92%
Average stopped delay per vehicle [s]	159.68	4.93%	144.91	3.05%	147.88	4.35%
Average number of stops per vehicle	11.65	7.72%	10.52	2.58%	12.33	2.98%
Average speed [mph]	9.77	4.01%	8.79	2.14%	8.52	1.34%
Total number of stops	694,942	5.90%	632,772	4.93%	734,998	5.43%

Figure 110 presents the improvements of the strategies and the baseline control plan for each performance measure. These results compare the strategies against a poorly-designed timing plan (75s cycle length with simultaneous offsets, with splits allocated proportionally to each phase demand) because the baseline timing plan that was provided by the City of Houston, TX is already a rather efficient operation. While oversaturation occurs in the network and is significantly debilitating for mobility during the peak period, these results indicate that conditions could be worse

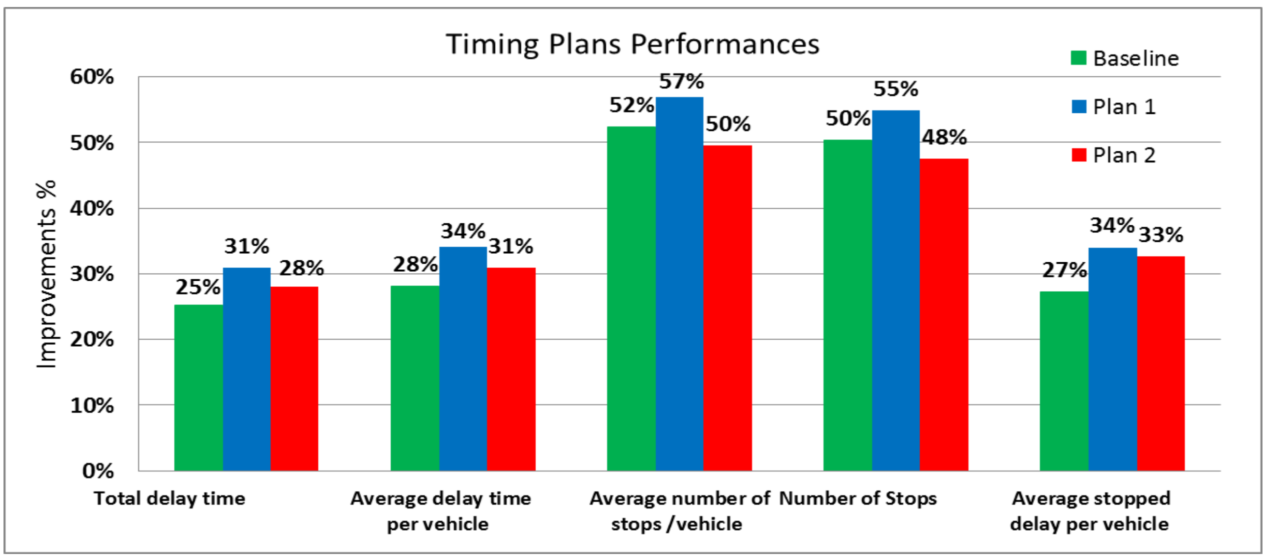


Figure 110. Network summary performance measures for Scenario 1

Figure 111 and Figure 112 present the throughput performance of Strategy 1 and Strategy 2 at the intersection level. These Figures compare the throughput at each intersection for the strategy with the baseline over the peak period. Negative values on these figures indicate that the baseline plan is more efficient at processing vehicles and positive values indicate that the mitigation control strategy is more efficient at processing vehicles at that intersection during the peak hour.

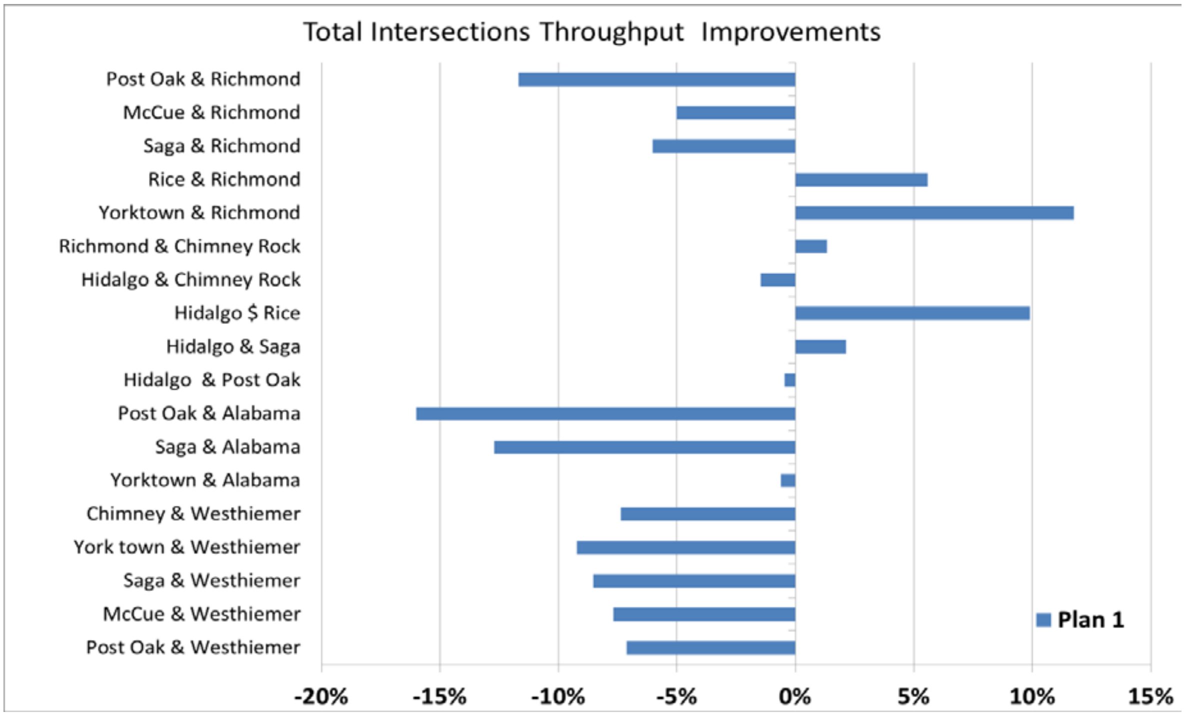


Figure 111. Intersection throughput improvement for Strategy 1

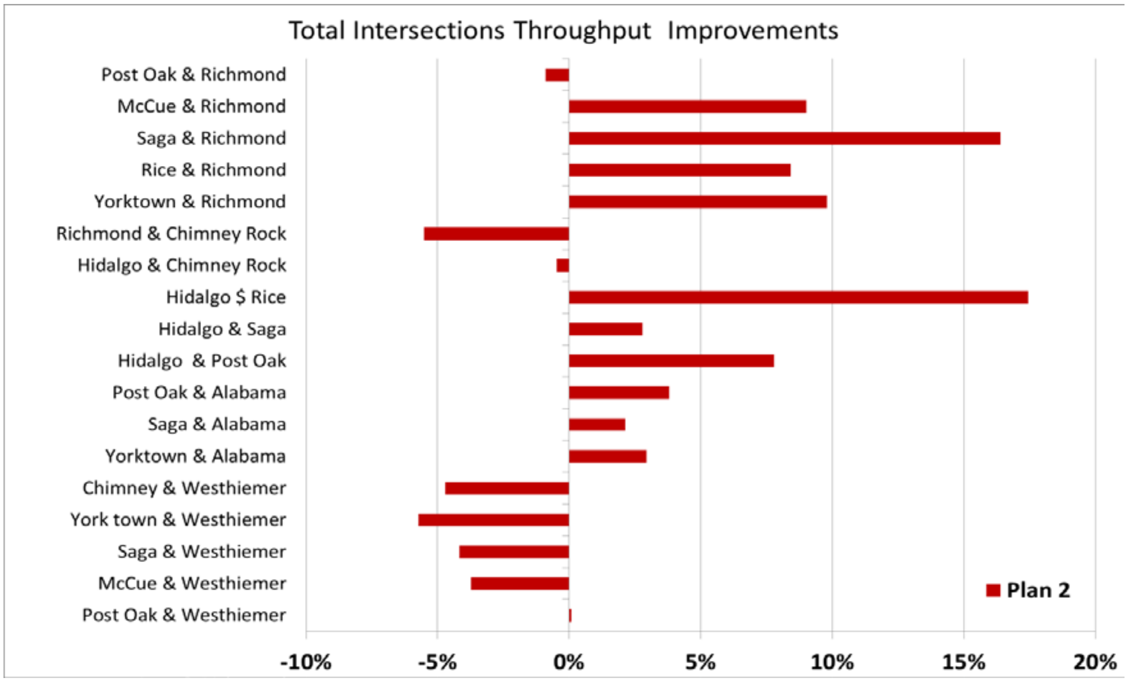


Figure 112. Intersection throughput improvement for Strategy 2

Figure 113 combines the information from the previous two figures on one display to compare the two strategies against each other. Strategy 2 demonstrates better throughput performance, as it improved the throughput of more interior intersections than Strategy (1). This is a validating result since Strategy (2) was designed to maximize system throughput before and after the processing regime. Strategy (1) was designed to minimize delay during the loading regime and maximize throughput during the recovery regime. Recall also that in Strategy 2 interior network intersections were run with half cycles. This type of operation reduced the length of overflow queues during red and improved green time utilization by reducing storage blocking and the resulting starvation symptoms. However, neither strategy could improve the throughput of intersections along Westheimer Road because in both cases the links on Westheimer Road were designated as gating links.

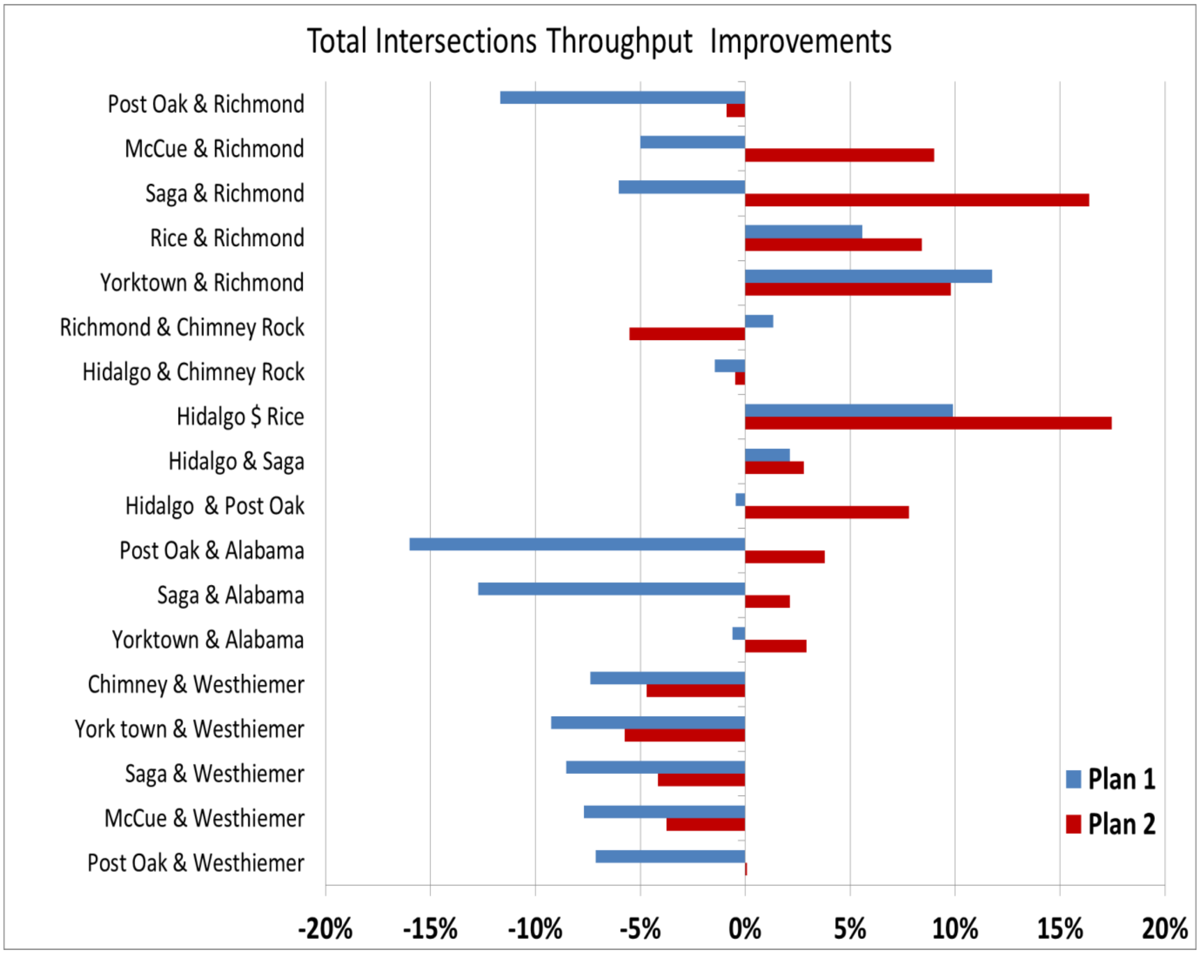


Figure 113. Comparison of intersection throughput between the two strategies

Network traffic loads (total vehicles in the system) during the peak period varied significantly for each control strategy. The total number of vehicles in the system (network load) was calculated by subtracting the total traffic exiting the system from the total traffic entering the system. Traffic loading trajectories for the three control strategies are illustrated in Figure 114 below. As shown, the baseline timing plan is effective in processing vehicles through the system during the loading. However, as the volumes continue to increase during the processing regime and queues start building up rapidly the baseline timings splits became inadequate to process the peak load and a large number of vehicles are gated outside of the network area. As the volumes begin to reduce during the recovery period, the latent demand then fills the network considerably and it takes much longer to clear those overflow queues from the system. Strategies 1 and 2 do a better job of spreading the traffic load over the peak period by more efficiently utilizing the space inside the network for storing vehicles. During the loading regime, both mitigation strategies protected the critical routes from becoming saturated early, since the v/c ratios of the phases on the critical routes were all minimized in the timing plan design. At 4:30 P.M., both strategies shifted to queue management strategies (i.e., phase reservice, flare the green, and half-cycling). At 5:00 P.M., Strategy 2 shifted back to throughput maximization control while Strategy 1 shifted to throughput maximization control at 5:30 P.M. The different switch points for the two strategies were determined based on the optimization process described earlier.

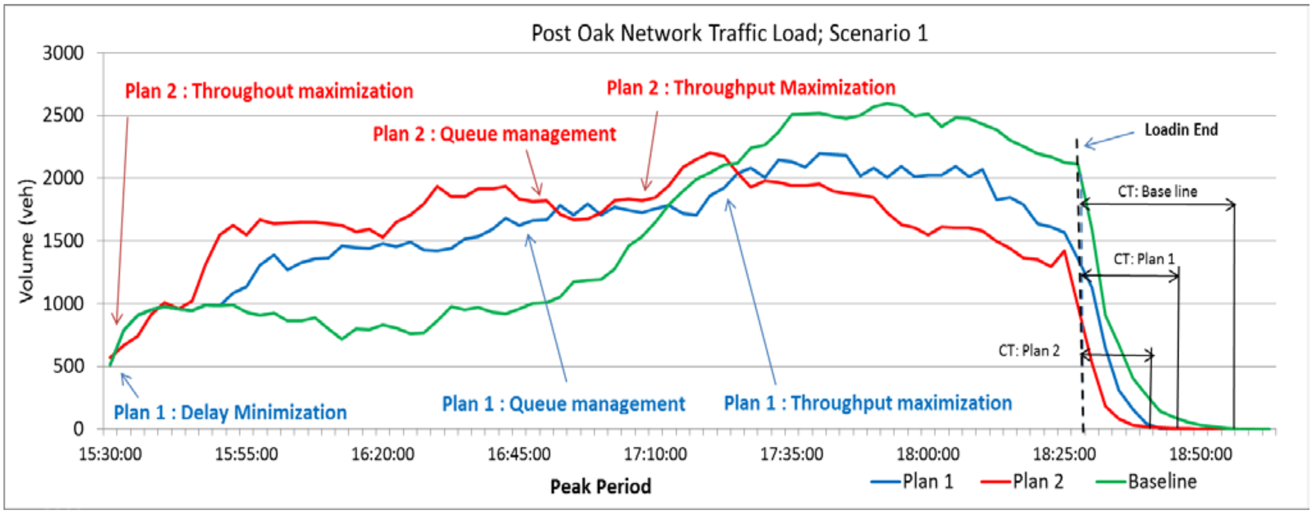


Figure 114. Number of vehicles in the system for each strategy for Scenario 1

Table 29 summarizes system-wide performance measures for the three strategies. As shown in the table, Strategy 1 is most effective in reducing total delay and total stops. Strategy 2 provides the fastest recovery time, but at the expense of higher delay and stops. Both strategies were able to reduce the recovery time by more than 10 minutes. This theme has been found to be recurrent in other test cases presented later in this report. Mitigation strategies for oversaturated conditions are particularly effective during the recovery regime.

Table 29. System-wide Summary performance measures

Metric	**Baseline**	**Strategy 1**	**Strategy 2**
Clearance Time (min)	27.5	17.5	14.5
Total Delay (hour)	3777.5	3327.3	3626.7
Total Stops (vehicle)	694.942	632,772	734,998
Max load in the system (vehicle)	2,596	2,202	2,196

Scenario 2: Critical Route Flows from Origins Inside the Network

Control strategies for Scenario 2 were developed using the same methodology applied to Scenario 1, but considered different assumptions on the critical routes. In Scenario 1 we assumed that the critical routes began outside of the network and progressed through the network to external destinations. In Scenario 2 we consider that the critical routes have origins inside of the network and are progressing to destinations outside the network. Similar to Scenario 1, we developed two mitigation strategies explicitly taking into account the three regimes of the scenario. The objectives of the Strategy 1 were to minimize delay during the loading regime, manage queues during the processing regime, and then maximize throughput during the recovery regime (denoted as min delay-manage queues-max throughput). The objectives of the Strategy 2 were to maximize throughput during both the loading and recovery regimes and manage queues in the processing regime (denoted as max throughput-manage queues-max throughput). Table 30 presents the control strategies selected for Scenario 2 with their corresponding operational objectives and expected outcomes. Figure 115 illustrates spatially where those strategies were planned to be tested in Scenario 2.

Table 30. Attributes of control strategies selected for testing on Scenario 2

Control Strategy	Location	Expected Improvement	Operational Objective
Left-turn phase Reservice	WBL Richmond Ave. & Rice Ave.	Prevent spillback	Queue Management
	SBL Sage Rd. & Richmond Ave		
	SBL Post Oak & Richmond Ave.		
	SBL Chimney Rock & Richmond Ave.		
	SBL Post Oak Ave. & Richmond Ave.		
	NBL Yorktown & Westheimer Rd.		
“Flare the green”	WB on Westheimer Rd.	Clear downstream queues	Queue Management
	WB on Richmond Ave.	Improve progression	&
	WB on Richmond Ave.	Enhance traffic throughput	Throughput maximization

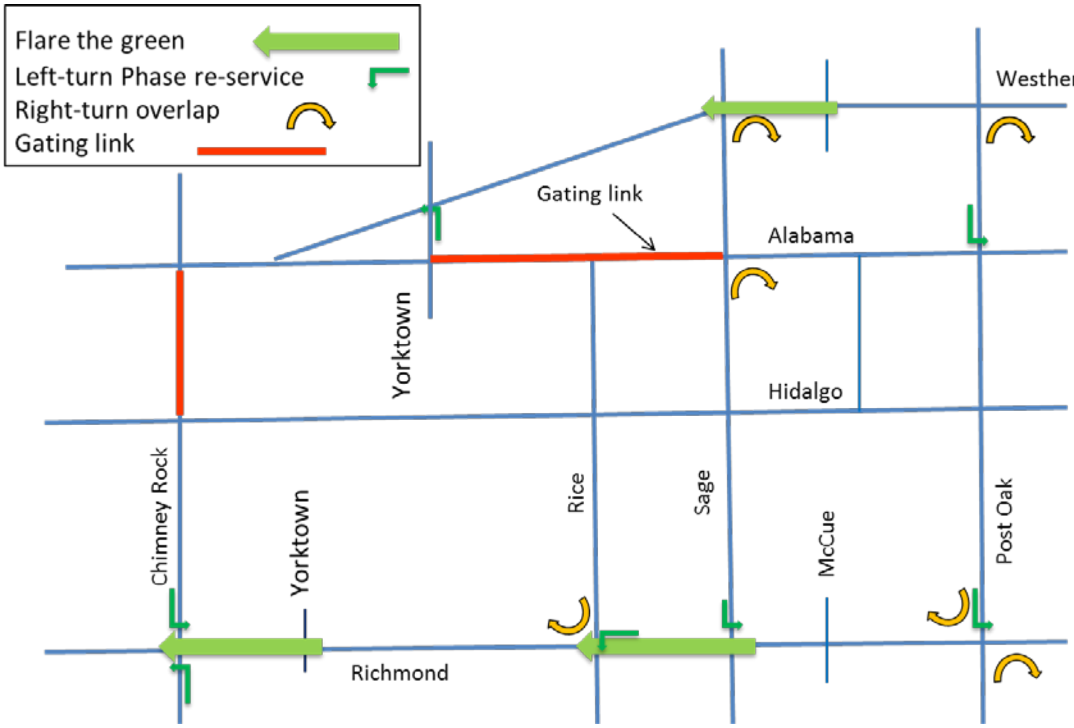


Figure 115. Spatial illustration of control strategies for Scenario 2

Scenario 2: Development of Timing Plans

The timing plans and TOD schedule for implementing the plans were generated using the same methodology followed for Scenario 1. Fifty-one plans were generated using the iterative procedure described previously. Timing plans 1-17 were developed to minimize total delay, timing plans 18-34 were developed to manage queues, and timing plans 35-51 were developed to maximize throughput. An optimization process was then applied to identify the sequence of three timing plans that produces the best performance for the objectives in each regime of operation.

The resulting schedule of timing plans corresponding to Scenario 2 are shown in Figure 116 and Figure 117. Start times for each plan in the schedule are shown in Table 31. Cycle lengths for each plan in the two strategies are listed in Table 32.

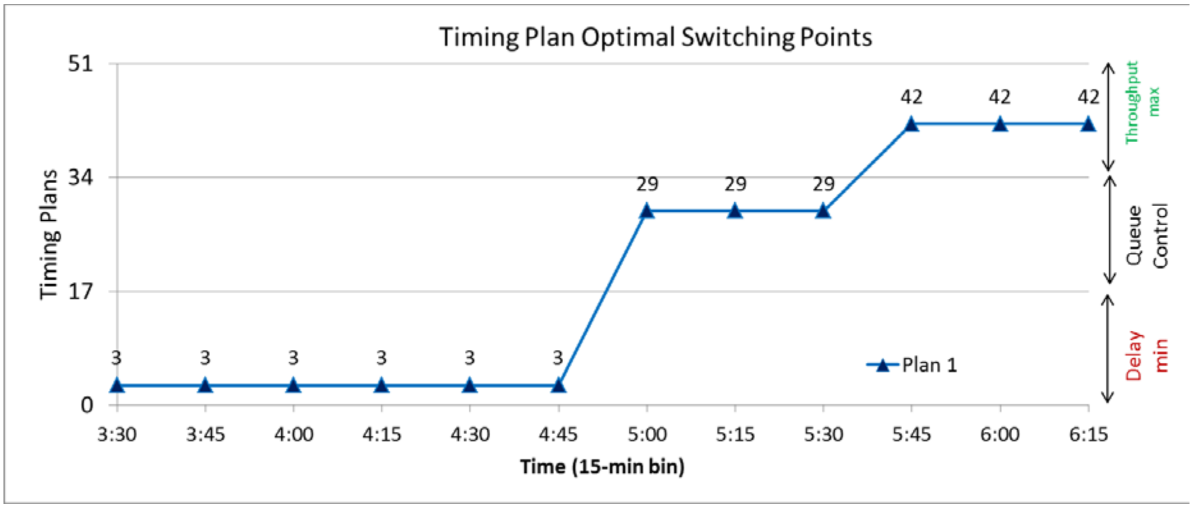


Figure 116. Min Delay-Queue Management-Max Throughput timing plan schedule

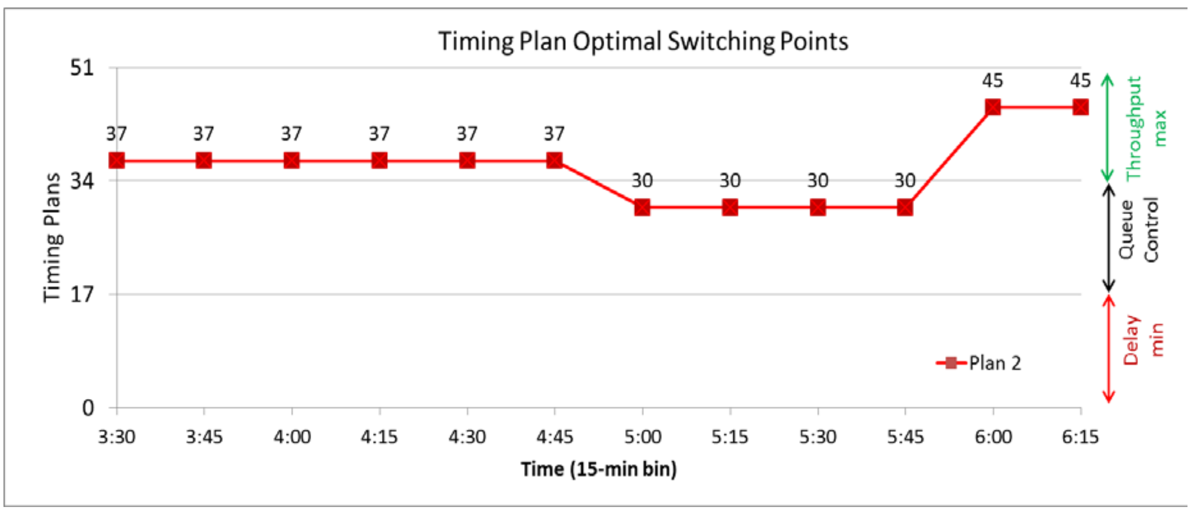


Figure 117. Max Throughput-Queue Management-Max Throughput timing plan schedule

Table 31. Plan start times for each strategy

Strategy No.	Loading Regime			Processing Regime			Recovery Regime		
	Operation Objective	Plan #	Starting time	Operation Objective	Plan #	Starting time	Operation Objective	Plan #	Starting time
1	Delay min	3	3:30	Queue management	29	5:00	Throughput max	42	5:30
2	Throughput max	37	3:30	Queue management	30	5:00	Throughput max	45	5:45

Table 32. Cycle lengths used in each plan for Scenario 2

	Timing Plan #	Optimal Cycle length (sec)
Strategy 1	11	80
	22	90
	43	150
Strategy 2	34	120
	20	150
	37	100

Performance measure comparisons are presented in Figure 118 through Figure 121.Traffic loading profiles are shown in Figure 122. Summary network performance statistics are shown in Table 34. The results show significant improvements in throughput and delay compared to the "bad" plan as well as comparing the mitigations to the baseline plan. This can be explained by the fact that the mitigation strategies were specifically designed to address Scenario 2, which might not be the case for the baseline plan. This is a confirmation of the importance of determining the critical routes in oversaturated networks.

Table 33 indicates that Strategy 1dominates Strategy 2 on all of the summary performance measures related to delay and stops. This is not surprising since Strategy 1 focuses on the minimization of delay during the loading regime and Strategy 2 focuses on maximizing throughput. Maximizing throughput at the beginning of a scenario that evolves slowly may apply the preferential treatment to the critical routes a bit too soon, allowing more vehicles in the system, but storing more of those vehicles on side streets.

Table 33. System-level comparison of performance of the two strategies for Scenario 2

Performance Measure	Strategy 1		Strategy 2	
	Mean	St. Dev.	Mean	St. Dev.
Total delay time [hr]	3172.96	4.97%	3327.30	1.67%
Average delay time per vehicle [s]	187.02	4.82%	198.07	1.81%
Total stopped delay [hr]	2156.24	5.57%	2242.04	1.72%
Average stopped delay per vehicle [s]	127.09	5.43%	133.47	1.85%
Average number of stops per vehicles	9.34	6.70%	11.48	2.58%
Average speed [mph]	9.77	5.19%	9.39	1.14%
Total number of stops	570,284	6.86%	693,982	2.43%

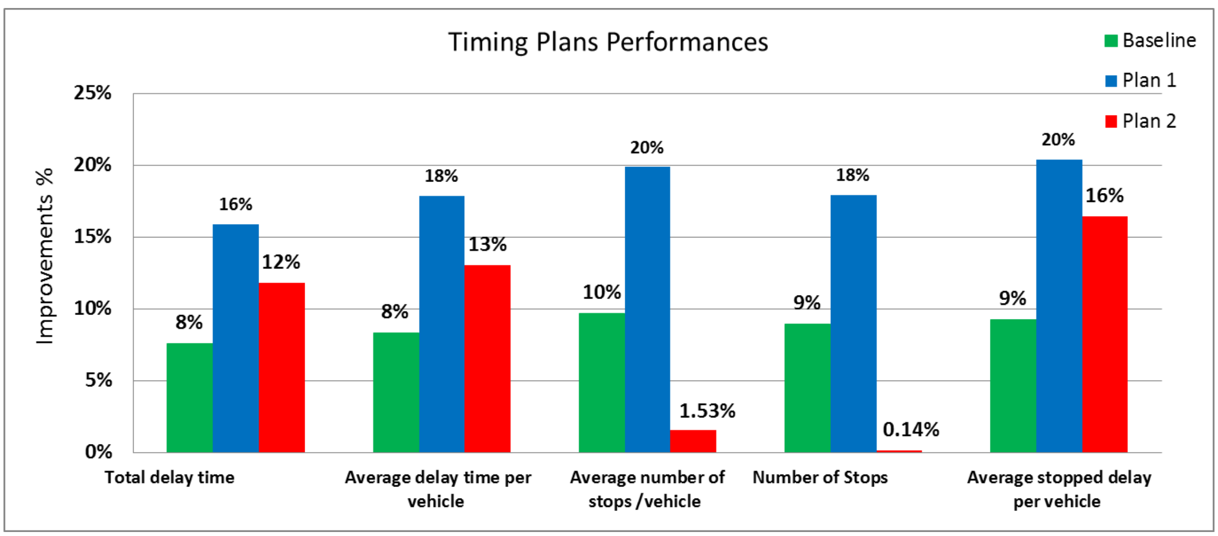


Figure 118. Comparison of performance improvements of the two strategies with the baseline for Scenario 2

Figure 119 and Figure 120 below illustrate the throughput performance of the strategies compared to the baseline plan. The two figures are then combined in Figure 121. The comparative performance is rather striking. Strategy 2, where throughput is maximized in both the loading and recovery regimes, improves the throughput for 14 of the 18 critical links where Strategy 1 which focuses on minimizing delay improves the throughput for only four critical links. This again illustrates the importance of selecting the appropriate optimization objectives during each regime of the scenario.

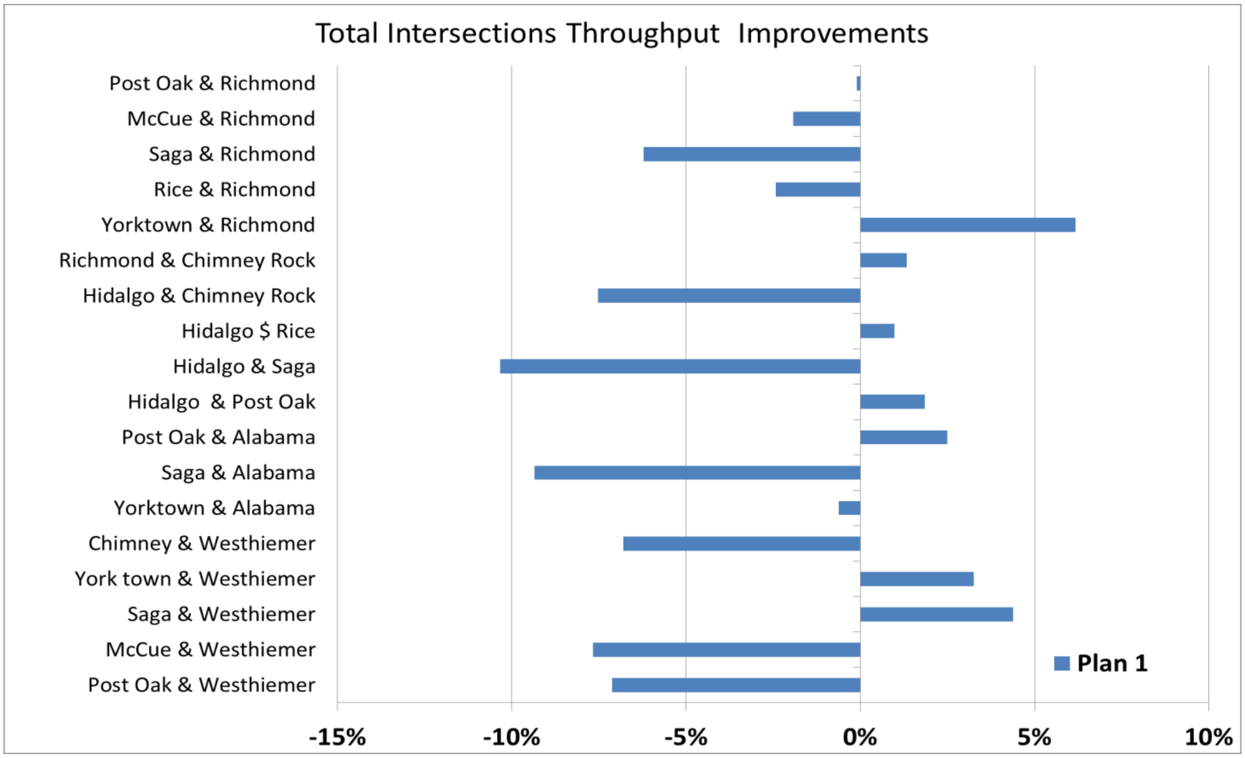


Figure 119. Intersection throughput improvements for Strategy 1 on Scenario 2

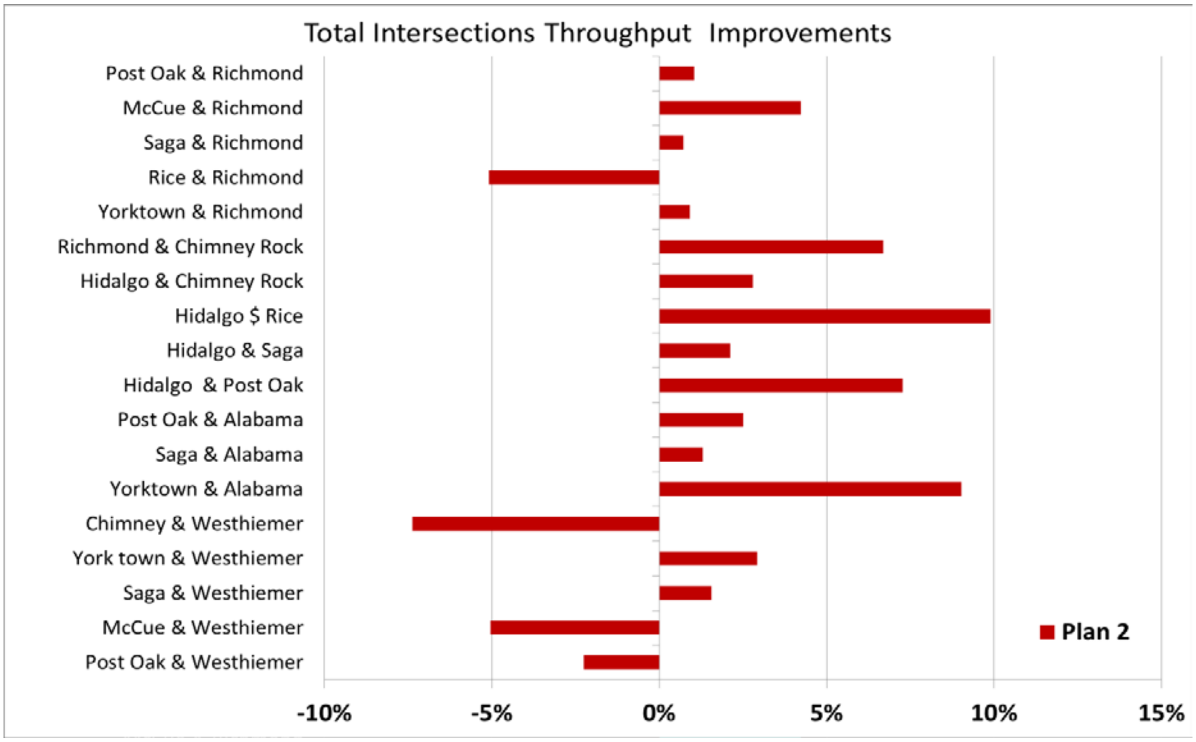


Figure 120. Intersection throughput improvements for Strategy 2 on Scenario 2

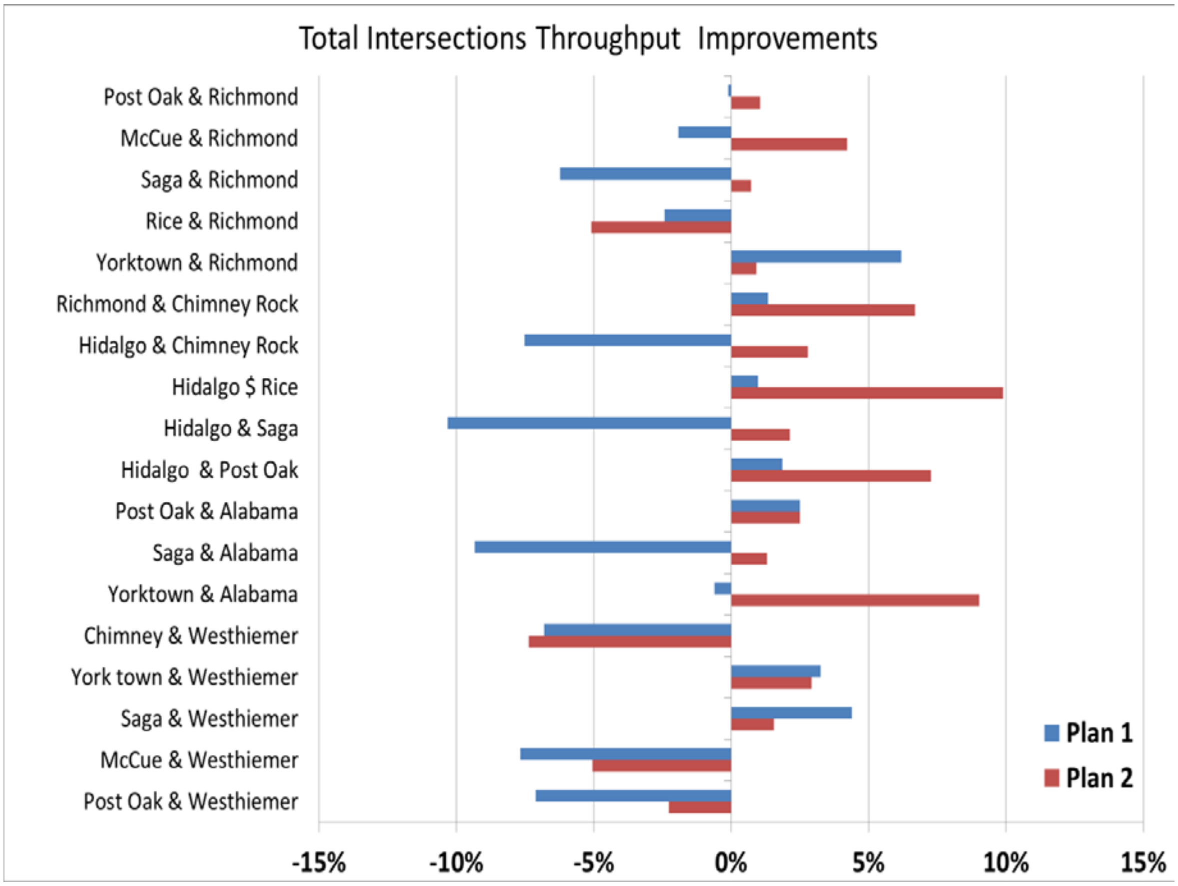


Figure 121. Comparison of throughput improvements between the two strategies

Figure 122 compares the traffic loading over time during the scenario for the baseline and the two strategies. In the loading regime, Strategy 1, which minimizes delay, follows the baseline plan initially, but then begins to store more vehicles in the system than the baseline. When the queue management plan is started at 5 P.M., the number of vehicles in the system stabilizes and remains more or less constant until the end of the simulation time. In comparison, Strategy 2 initially stores more vehicles in the system as it focuses on maximizing throughput on the critical routes. The performance during the processing regime is more peaked than Strategy 1, but still does not result in the spiking that occurs in the baseline operation. These results mimic what was found in other test cases that will be presented later.

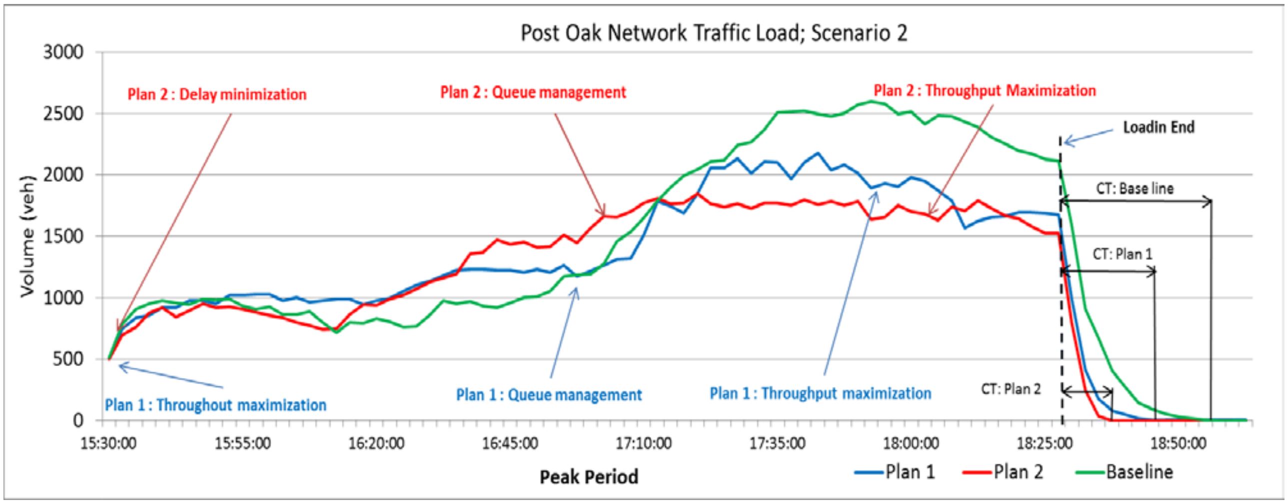


Figure 122. Number of vehicles in the system for each strategy for Scenario 2

Table 34. System-wide Performance Results for Scenario 2

	Baseline	**Strategy 1**	**Strategy 2**
Clearance Time (min)	27.5	17.5	14.5
Total Delay (hr)	3777.5	3327.3	3626.7
Total Stops	694.942	632,772	734,998
Max load in the system (vehicles)	2,596	2,202	2,196

Conclusions from the Post Oak Test Case

The simulation results from the Post Oak network test case showed that significant improvements are achievable by applying the methodology developed in this research. This test case extended the methodology applied to the Reston Parkway test case by explicitly considering the three regimes of operation during a scenario. In the Reston Parkway case, we considered application of only a single timing plan during the entire scenario. In this test, we proposed two canonical combinations of optimization objectives (1) min delay-manage queues-max throughput, and (2) max throughput -manage queues-max throughput. One of the goals in this test case was to determine some indication of the superiority or at least characterize the differences between the two canonical strategies.

Two feasible critical route scenarios were identified from system detector data logs obtained from the City of Houston and from discussions with City and County traffic engineers. Mitigation strategies were developed for each critical scenario as a combination of ad-hoc inspection of the network's critical locations and the optimization procedure developed for determining feasible combinations of cycle, splits, and offsets described in Chapter 2. The optimization process was extended further to consider the selection of a sequence of timing plans and the associated switching times between the timing plans. Each timing plan in the sequence in each regime of operation was selected to achieve an optimization objective that represented the system objective for that regime (minimize total delay, maximize total throughput, or manage queues). These two canonical strategies were then tested on the critical routing scenario using simulation.

Both routing scenarios showed that significant performance improvements were possible by applying mitigation strategies. The first critical routing scenario showed less impressive improvements for the two strategies over the baseline performance for measures based on delay and stops but more substantial improvements in intersection throughput. The second scenario showed more impressive improvements over the baseline for delay and stops, but less substantial improvements to throughput.

In both scenarios, the strategy that minimized delay during the loading period significantly outperformed the strategy that maximized throughput during the loading period on measures related to stops and delay. The strategy that maximized throughput during both loading and recovery regimes outperformed the other strategy on throughput. Neither strategy would be considered to completely dominate the other, so it is still an open question whether or not the system objective should be to maximize throughput or minimize delay during the loading regime. We did not test a strategy which minimized delay during the recovery regime, but it was clear from both scenarios that maximizing throughput during the recovery regime is an effective transition in optimization goals.

This methodology and approach focuses on a top down method to identify and mitigate oversaturated conditions with fixed-parameter timing plans. In the next section, we revisit our "bottom up" approach using real-time estimates of TOSI and SOSI to directly calculate modifications to green time on an oversaturated route. The theory guiding this heuristic procedure is first described and then two test cases are presented. The first test case describes the application of the process on the TH55 test network in Minneapolis, MN. The second test case describes the application of the process on a grid network in downtown Pasadena, CA. Following this section, we demonstrate the online application of mitigation strategies using real-time detector and TOSI / SOSI oversaturation estimates.

Using TOSI and SOSI Measures to Directly Calculate Green Time Adjustments

In the previous section, we discussed a top-down approach for generating mitigation timing plans using an optimization methodology. This methodology is experimental and relatively complicated and could not be applied in practice without additional research and development effort. This optimization methodology also does not directly integrate the measurement of TOSI and SOSI metrics developed earlier because it was developed in parallel to those measures.

In this section we address the challenge of oversaturated conditions from the bottom-up by developing a heuristic process that can compute green time adjustments for oversaturated routes. This process uses the TOSI and SOSI measures to determine the amount of green time to add and subtract, respectively, from a phase. The theory and foundation of the heuristic is first presented and then followed by two test cases. The first is a simple situation on the TH55 arterial network in Minneapolis, MN. This test case validates the basic concept. The second test case is a grid network in Pasadena, CA. This test case illustrates the extension of the methodology for two oversaturated routes that cross at a critical location. The Pasadena, CA network was selected for testing as the city has current interest in deploying the experimental hardware and software for calculating queue lengths and TOSI/SOSI measures.

Forward-Backward Procedure

In this section, we introduce a heuristic procedure to mitigate traffic congestion along an oversaturated route by directly applying the TOSI and SOSI theory described in Chapter 2. The procedure, denoted as the forward-backward procedure (FBP), can be applied for both online and offline signal timing adjustments. In these test cases we apply the procedure in an offline manner. In the FBP, the forward process follows the direction of traffic on the oversaturated route and aims to remove the oversaturation by changing the green and red times for all phases along the route. The backward process follows the opposing direction of traffic on the oversaturated route and evaluates the feasibility of the changes to green and red times by considering the constraints such as minimum green times and queue storage space. In this procedure, our intent is not to find the optimal solution by solving a complicated optimization program; instead, we present a logical heuristic to address oversaturation by adjusting green times and offsets. The effectiveness of the proposed FBP is demonstrated in simulation tests.

For the sake of simplicity we keep the cycle length of each intersection unchanged. The control variables at each intersection of interest are the green and red times for the oversaturated phase. We should note that, although the cycle length is unchanged, the changes to the green and red times do not necessarily have the same absolute value and hold the opposite sign. As we will explain further, the changes to the green and red times can subsequently be transformed into the adjustments of offsets and green splits. We first introduce the control variables and basic theory for applying TOSI and SOSI measures in order to mitigate oversaturation.

Control Variables

In the FBP, two sets of control variables $\Delta r_{n,i}$ and $\Delta g_{n,i}$, namely red time changes and green time changes for phase i at intersection n, are introduced for each oversaturated phase. The two control variables have direct association with specific oversaturation mitigation strategies. Whether to change red or green time is determined by the cause of the oversaturation. Changing red times (i.e. $\Delta r_{n,i}$) aims to eliminate spillover and changing green times (i.e. $\Delta g_{n,i}$) aims to clear an overflow queue. A positive red time change (**red extension**) means that extra red time is added. Since the cycle length is kept unchanged, the green start would be postponed with the red extension (see Figure 123a) and the total green time is reduced. A negative red time change (**red reduction**) means a portion of red time is cut from the end of red, therefore, green start will be advanced (see Figure 123b) and the total green time is increased. Similarly, a positive green time change (**green extension**) indicates that additional green time is added to the original end of the green time (see Figure 123c), and a negative green time change (**green reduction**) represents that some green time is cut from the end of green (see Figure 123d). Depending on the offset reference point used for the intersection (start of yellow, start of green, barrier crossing, etc.), each case of adjusting green or red may require a corresponding change to the offset value.

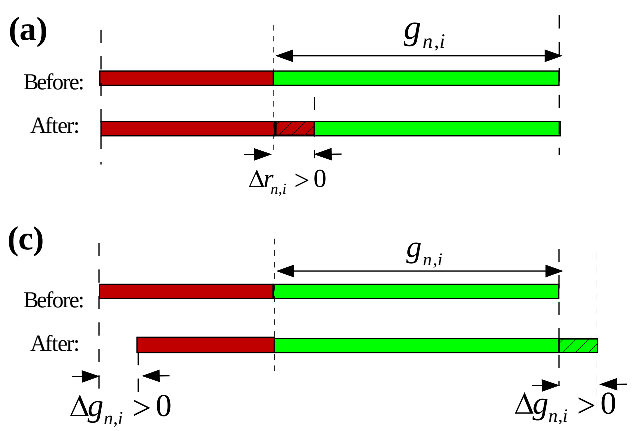


Figure 123. Red time changes and green time changes

The values $\Delta r_{n,i}$ and $\Delta g_{n,i}$ can be easily transformed into the values of new offset and green duration, which can easily be modified in the signal timing plan. If we assume that the oversaturated approach is the coordinated direction and the green start time of the coordinated phase is the offset reference point, Eq. can be used to calculate the new offset $\tilde{o}_n$, green duration $\tilde{g}_{n,i}$ and red duration $\tilde{r}_{n,i}$. Here c_n is the cycle length for intersection n. Figure 124 demonstrates an example of signal timing changes at one intersection with $\Delta r_{n,i} < 0$ and $\Delta g_{n,i} > 0$.

$$\begin{cases} \tilde{o}_n = o_n + \Delta r_{n,i} \\ \tilde{g}_{n,i} = g_{n,i} - \Delta r_{n,i} + \Delta g_{n,i} \\ \tilde{r}_{n,i} = c_n - \tilde{g}_{n,i} \end{cases} \quad \text{Eq. 51}$$

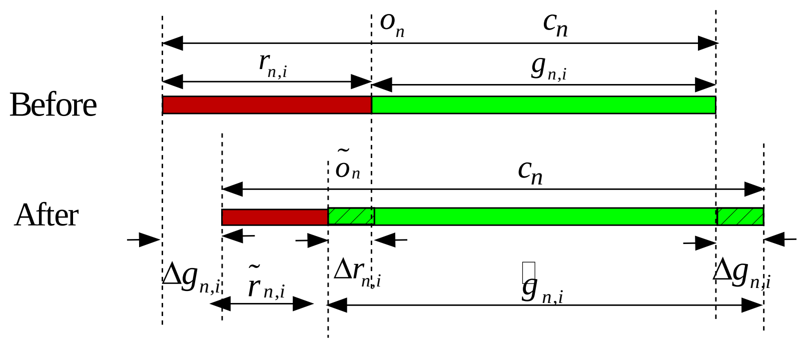


Figure 124. Signal timing changes ($\Delta r_{n,i} < 0, \Delta g_{n,i} > 0$)

Basic Strategies Using TOSI and SOSI Measures

Simply speaking, there are two ways to deal with oversaturation: one is to increase the downstream output rate and the other is to constrain the upstream input rate. These two basic actions result in three mitigation strategies for an oversaturated phase when considering the values of TOSI and SOSI. These three basic cases will be used in the FBP to compute the changes to green and red times.

Green Extension for Scenario 1: TOSI > 0 & SOSI = 0

Since a positive TOSI value indicates an overflow queue at the end of a cycle and zero SOSI value indicates that there is still spare capacity to store vehicles in the downstream link, the strategy to deal with this situation is to extend the green time for the oversaturated phase. Figure 125 illustrates this case by presenting the shockwave profiles for two intersections. After extending the green in Figure (b), the overflow queue disappears and TOSI becomes zero. The green extension can be calculated as the following Eq. 52.

$$\Delta g_{n,i} = TOSI_{n,i} * g_{n,i} \quad \text{Eq. 52}$$

where $\Delta g_{n,i}$ is the adjustment to the green time at intersection n for phase i; $TOSI_{n,i}$ is the TOSI value at intersection n for phase i; and $g_{n,i}$ is the green time at intersection n for phase i. Note that positive $\Delta g_{n,i}$ means green extension; and a negative value means green reduction. By extending or reducing green, the start time of the following red signal will be shifted.

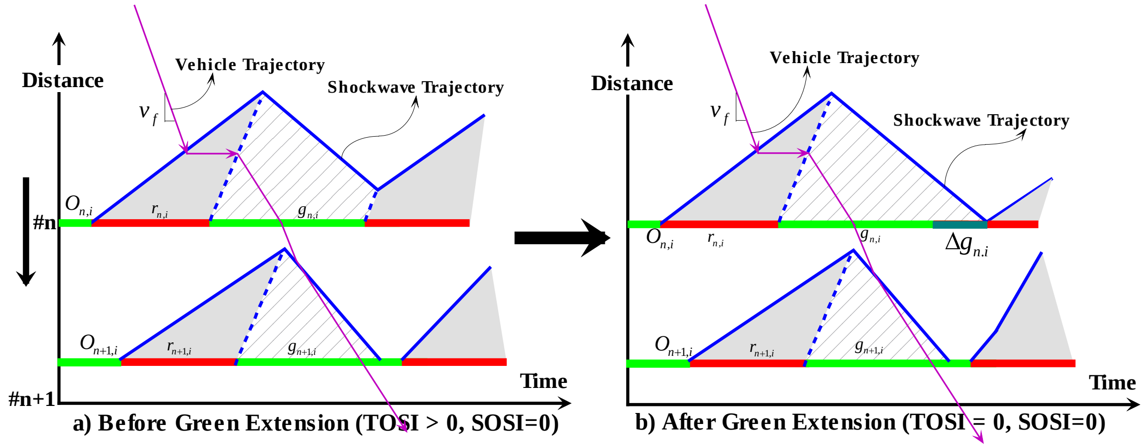


Figure 125. Green extension for Scenario 1

We should also note that by extending the green time at the current intersection, the queue length at the downstream approach may increase and result in spillover. This problem will be addressed in the FBP.

Red Extension for Scenario 2: TOSI = 0 & SOSI > 0.

If SOSI is larger than zero, it indicates that the downstream queue spills back to the upstream intersection and results in unusable green time as shown in Figure 126. But since TOSI is zero, all queued vehicles can be discharged even with reduced green time. One way to remove downstream spillover is to gate the upstream flow by extending the red time. The red extension can be calculated as the following Eq. 53.

$$\Delta r_{n,i} = SOSI_{n,i} * g_{n,i} \quad \text{Eq. 53}$$

where $\Delta \mathrm{r}_{\mathrm{n,i}}$ is the adjustment to the red time at intersection *n* for phase *i*; and $SOSI_{n,i}$ is the measured or estimated SOSI value at intersection *n* for phase *i*. The positive $\Delta \mathrm{r}_{\mathrm{n,i}}$ means red extension and a negative value means red reduction. Note that by adjusting the red time, the start of the following green will be shifted.

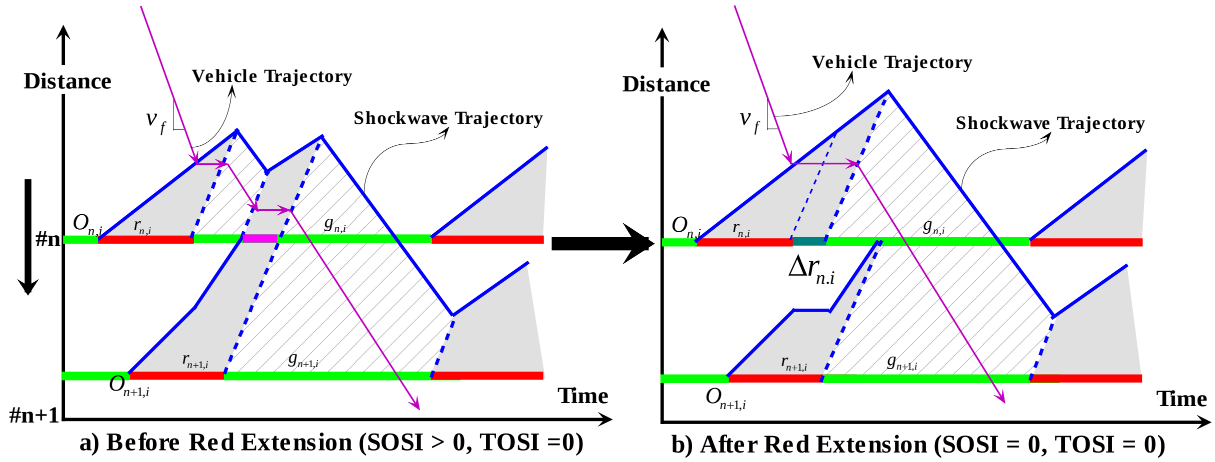


Figure 126. Red extension for Scenario 2

Downstream Red Reduction for Scenario 3: TOSI > 0 & SOSI > 0.

A more serious situation exists when both TOSI and SOSI are larger than zero, as shown in Figure 127a. In this case, at the upstream intersection a portion of the green time is unused because of the downstream spillover. At the same time, the useable green time at the upstream intersection is not sufficient to discharge queued vehicles, i.e., an overflow queue exists. One way to address this scenario is to increase downstream capacity by reducing the red time at the downstream intersection. As shown in

Figure **127**b, by reducing the downstream red, positive TOSI and SOSI values for the upstream intersection will be reduced. Once the downstream spillover is removed or reduced, the unusable green time at the upstream intersection may become available and can be used to discharge the overflow queue. If TOSI < SOSI, the overflow queue can be cleared by using this strategy. The reduction of downstream red can be calculated as the following Eq. 54.

$$\Delta r_{n+1,i} = SOSI_{n,i} * g_{n,i} \qquad \text{Eq. 54}$$

As an alternative to reducing the downstream red, we can also deal with this situation by combining the methods for Scenarios 1 and 2 together, i.e., extending both the red and green times. This would mean an increase to the total cycle time at this intersection. In this heuristic procedure, we are attempting to avoid the re-calculation of the cycle time since the calculations for each intersection may result in different cycle values; thus requiring an optimization procedure to select the best value.

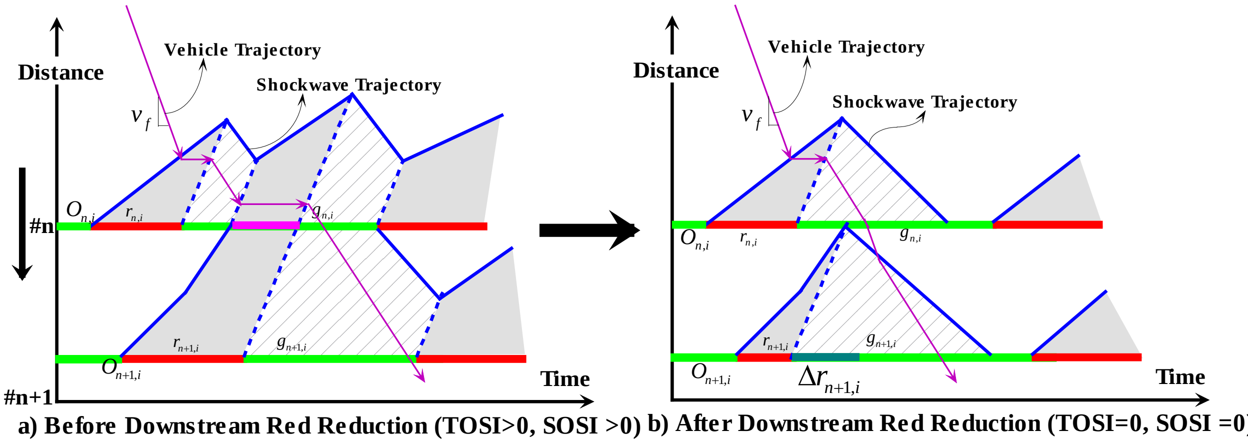


Figure 127. Red reduction at downstream intersection for Scenario 3

Among the three strategies, extending green (Strategy 1) is to increase the discharge capacity for the oversaturated phase; extending red (Strategy 2) is to gate traffic arrivals at the upstream intersection; and reducing downstream red (Strategy 3) is to remove the downstream bottleneck by discharging the queue earlier at the downstream intersection. By considering maximum/minimum green and storage space limitations on side streets, these strategies may be directly applied for an isolated intersection or two intersections in tandem. For a longer oversaturated route, the FBP needs to be applied for two reasons. First, the increase of green time of an upstream approach may create oversaturation on the downstream link and secondly, capacity constraints at a downstream phase may limit the possible signal timing adjustments for the upstream phase.

The Forward-Backward Procedure (FBP)

The FBP systematically evaluates the three strategies by considering the impacts of the green and red time modifications on upstream and downstream intersections. As shown in Figure 128, the FBP consists of two processes: a forward pass and a backward pass. We will first describe the calculations in the forward process, followed by the calculations in backward process.

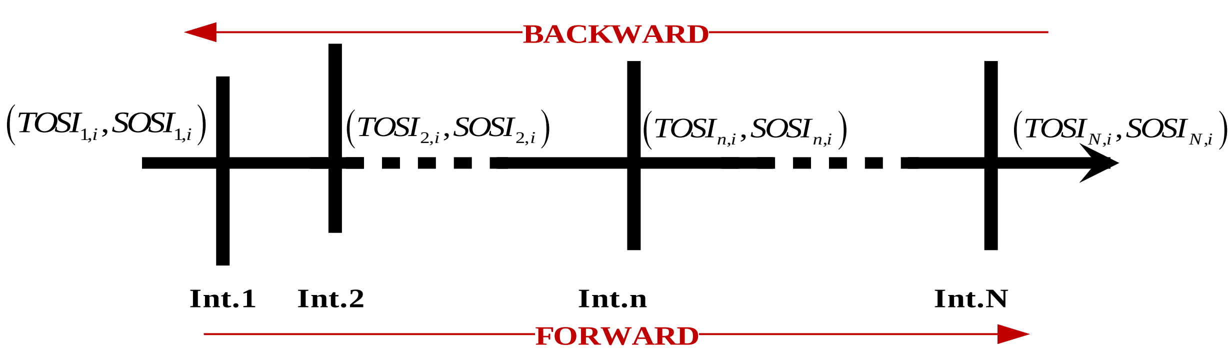


Figure 128. FBP for an oversaturated route

Forward Process

The forward process aims to eliminate both spillovers and overflow queues by reducing red time or increasing green time of oversaturated phases without considering the constraints from other phases at the intersection. The process is applied along the direction of flow and calculates the red and green time changes for each oversaturated phase on the route. Only basic Strategy 1 and 3 (described in the last section) are considered in this pass.

For any intersection (except the first and last ones, which need slightly different treatments) on an oversaturated route (see Figure 128), the first step is to reduce red time according to the upstream *SOSI* value to remove spillover (see Eq. 55). The amount of red reduction at any intersection should accommodate not only the removal of the spillover to the upstream intersection ($SOSI_{n-1,i} \times g_{n-1,i}$), but also the increase in the arrival flow due to the red reduction made at the upstream intersection ($\Delta r^F_{n-1,i}$). The second step is to extend green times to discharge overflow queues. Similarly, we need to account for the green change of the upstream intersection ($\Delta g^F_{n-1,i}$). Since in the forward process we assume that downstream spillover will be removed by reducing downstream red, the unusable green time caused by spillover will become available and can be used to discharge overflow queues. The backward process will consider the situation if capacity constraints are violated.

Therefore, in order to remove an overflow queue, additional green time in the amount of $\left(TOSI_{n,i} - SOSI_{n,i}\right) \times g_{n,i}$ is needed. The additional green time might be negative if the *SOSI* value is greater than the *TOSI* value. This situation would mean that the elimination of spillover has already given enough extra green time to discharge the overflow queue and we need to reduce green time to prevent green starvation. The signal timing adjustment for any intersection except the first and last ones can be calculated by the following equation (Note that the superscript "F" denotes "Forward" process):

$$\begin{cases} \Delta r^F_{n,i} = \Delta r^F_{n-1,i} - SOSI_{n-1,i} \times g_{n-1,i} \\ \Delta g^F_{n,i} = \Delta g^F_{n-1,i} + \left(TOSI_{n,i} - SOSI_{n,i}\right) \times g_{n,i} \end{cases} \quad \text{Eq. 55}$$

For the first intersection, we only need to consider the situation when TOSI is greater than zero (positive SOSI will be handled at the downstream intersection), so Eq. 55 becomes:

$$\begin{cases} \Delta r^F_{1,i} = 0 \\ \Delta g^F_{1,i}(t) = \left(TOSI_{1,i} - SOSI_{1,i}\right) \times g_{1,i} \end{cases} \quad \text{Eq. 56}$$

For the last intersection (intersection N), since SOSI is zero (otherwise, this intersection should not be the last one), Eq. 55 becomes:

$$\begin{cases} \Delta r_{N,i}^{F} = \Delta r_{N-1,i}^{F} - SOSI_{N-1,i} \times g_{N-1,i} \\ \Delta g_{N,i}^{F} = \Delta g_{N-1,i}^{F} + TOSI_{N,i} \times g_{N,i} \end{cases}$$ Eq. 57

Backward Process

The forward process follows the traffic direction and adds extra green time ($\Delta g_{n,i}^{F} - \Delta r_{n,i}^{F}$) for each phase to discharge the overflow queue and to remove spillover. However, available green time increases for some intersections may not be achievable due to the other constraints, *i.e.*, minimum green requirement for conflicting phases and queue storage space limitations. To solve this problem, the backward pass is designed to gate traffic when the green time changes calculated in the forward pass are not achievable. The backward pass starts from the last intersection and follows the opposing direction of traffic on the oversaturated route to determine how much green time needs to be reduced for each phase.

To compute the backward adjustment, we first need to calculate the available green time $g_{n,i}^{a}$ for each oversaturated phase. If only the minimum green time requirement for conflicting phases is considered, $g_{n,i}^{a}$ for intersection n and phase i can be calculated by Eq. 58, where c_n is the cycle length for intersection n; $Z_{n,i}$ is the set of conflict phases (other phases in the same ring if dual-ring control is utilized) to phase i at intersection n; and $g_{n,j}^{\min}$ is the minimum green time for phase j at intersection n. Note that $g_{n,i}^{a}$ may also be constrained by other conditions, such as the storage space on side streets (*i.e.*, $g_{n,i}^{a}$ can be computed by comparing the maximum queue length with the link length for side streets).

$$g_{n,i}^{a} = c_n - \sum_{j \in Z_{n,i}} g_{n,j}^{\min} - g_{n,i}$$ Eq. 58

Next we need to follow the direction of the opposing flow and calculate the residual capacity $R_{n,i}$ for each phase on the route. Positive $R_{n,i}$ means the available green time can accommodate the required green time increase $\left(\Delta g_{n,i}^{F} - \Delta r_{n,i}^{F}\right)$; negative $R_{n,i}$ means available green time is insufficient.

$$R_{n,i} = g_{n,i}^{a} - \left(\Delta g_{n,i}^{F} - \Delta r_{n,i}^{F}\right), n = N, \ldots, 1$$ Eq. 59

After calculating the residual capacity for each phase on the route, the backward green time adjustment term Δg_{i}^{B} is equal to the minimum residual capacity among all phases on the route, see Eq. 60. Note that Δg_{i}^{B} has no subscript "n", which means it is the same for every intersection along the route. If $\min_{n\in\{1,\ldots,N\}}\left(R_{n,i}\right)\geq 0$, the requested green time increase $\left(\Delta g_{n,i}^{F}-\Delta r_{n,i}^{F}\right)$ from the forward process will be satisfied at all intersections and no further adjustment is needed in the backward process (i.e., $\Delta g_{i}^{B}=0$). However, if $\min_{n\in\{1,\ldots,N\}}\left(R_{n,i}\right)<0$, there is at least one phase on the route where the available green time constraint is violated and the adjustment term Δg_{i}^{B} is utilized to make sure the available green time constraints are satisfied at all phases on the route.

$$\Delta g_{i}^{B}=\min\left\{\min_{n\in\{1,\ldots,N\}}\left(R_{n,i}\right),0\right\} \qquad \text{Eq. 60}$$

The final changes to every phase on the route are calculated by Eq. 61, where the red time change is equal to the value calculated in the forward process and the green time change is equal to the summation of the calculated value in the forward process $\Delta g_{n,i}^{F}$ and the backward adjustment term Δg_{i}^{B}.

$$\begin{cases}\Delta r_{n,i}=\Delta r_{n,i}^{F}\\ \Delta g_{n,i}=\Delta g_{n,i}^{F}+\Delta g_{i}^{B}\end{cases} \qquad \text{Eq. 61}$$

FBP for an Oversaturated Network

When extending the FBP to an oversaturated network, the first step is to identify oversaturated routes. For routes that cross each other, an algorithm is designed to allocate the available green time to the conflicting phases.

If we have the *TOSI* and *SOSI* values for each movement, it is easy to identify oversaturated routes, since the oversaturated movements on the route will have positive *TOSI* and/or *SOSI* values. This route need not be a straight line (see Figure 129), since oversaturation at some intersections may be caused by turning movements. When two oversaturated routes intersect with each other, we call the crossing intersection the “critical intersection”. For each oversaturated route, the FBP will determine the changes to the green/red times along the route. There will be a conflict between the two routes at the critical intersection since both directions will be fighting for the available green time. The available green for both phase *i* and *j* ($g_{I,\,i\&j}^{a}$) at intersection *I* can be calculated by Eq. 62, where $Z_{I,\,i\&j}$ is the set of conflict phases to phase *i* and *j* at intersection *I*.

$$g_{I,\,i\&j}^{a}=c_{I}-\sum_{k\in Z_{I,\,i\&j}}g_{I,k}^{\min}-g_{I,i}-g_{I,j} \qquad \text{Eq. 62}$$

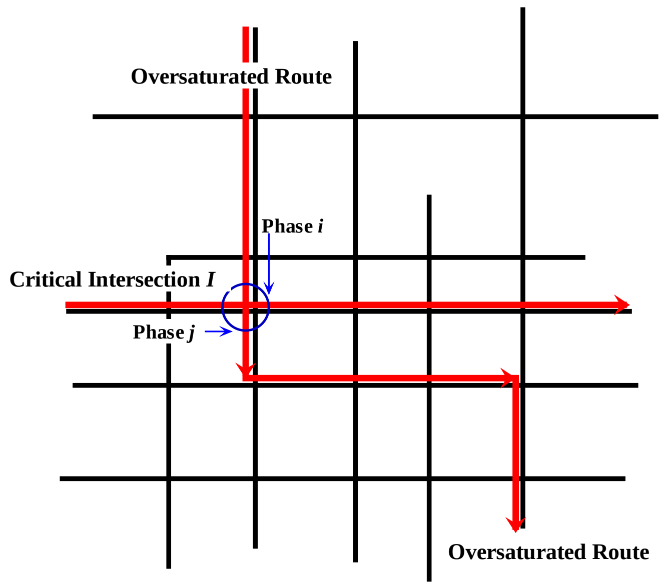


Figure 129: Oversaturated route and critical intersection

For two oversaturated routes that intersect, the available green time at intersection I ($g^a_{I,\,i\&j}$) needs to be divided between phase *i* and *j*. We can split the available green time proportionally according to the requested green times from the forward process, i.e.,

$$\begin{cases} g^a_{I,i} = g^a_{I,i\&j} \times \dfrac{g^F_{I,i}}{g^F_{I,i} + g^F_{I,j}} \\ g^a_{I,j} = g^a_{I,i\&j} \times \dfrac{g^F_{I,j}}{g^F_{I,i} + g^F_{I,j}} \end{cases} \qquad \text{Eq. 63}$$

However, such a method of splitting the green time may not be efficient because the binding constraints for available green times on one or both oversaturated routes may not come from the critical intersection. To overcome this deficiency, we first compute the residual capacity for all intersections except the critical intersection I and the backward adjustment term for both directions, $\Delta\hat{g}^B_i$ and $\Delta\hat{g}^B_j$, using Eq. 60. The "^" sign is used because here we did not consider intersection *I*.

We can then calculate the requested green time increase for phase *i* and *j* of intersection I, denoted by $g^R_{I,i}$ and $g^R_{I,j}$.

$$\begin{cases} g^R_{I,i} = \Delta g^F_{I,i} + \Delta\hat{g}^B_i - \Delta r^F_{I,i} \\ g^R_{I,j} = \Delta g^F_{I,j} + \Delta\hat{g}^B_j - \Delta r^F_{I,j} \end{cases} \qquad \text{Eq. 64}$$

If $g^{a}_{I,\,i\&j} \geq g^{R}_{I,i} + g^{R}_{I,j}$, then the available green time constraint at intersection I is satisfied. The backward process adjustment terms Δg^{B}_{i} and Δg^{B}_{j} are equal to $\Delta \tilde{g}^{B}_{i}$ and $\Delta \tilde{g}^{B}_{j}$ respectively, see Eq. 65.

$$\begin{cases} \Delta g^{B}_{i} = \Delta \tilde{g}^{B}_{i} \\ \Delta g^{B}_{j} = \Delta \tilde{g}^{B}_{j} \end{cases} \qquad \text{Eq. 65}$$

Otherwise, the available green time at intersection I cannot satisfy the total requested green time increase for both directions i and j. The total available green time $g^{a}_{I,\,i\&j}$ is split proportionally to two directions, $g^{a}_{I,\,i}$ and $g^{a}_{I,\,j}$ according to the requested green time increase, see Eq. 66. The total backward process adjustment terms for the two directions is then determined by Eq. 67.

$$\begin{cases} g^{a}_{I,i} = g^{a}_{I,i\&j} \times \dfrac{g^{R}_{I,i}}{g^{R}_{I,i} + g^{R}_{I,j}} \\ g^{a}_{I,j} = g^{a}_{I,i\&j} \times \dfrac{g^{R}_{I,j}}{g^{R}_{I,i} + g^{R}_{I,j}} \end{cases} \qquad \text{Eq. 66}$$

$$\begin{cases} \Delta g^{B}_{i} = g^{a}_{I,i} - \left(\Delta g^{F}_{I,i} - \Delta r^{F}_{I,i}\right) \\ \Delta g^{B}_{j} = g^{a}_{I,j} - \left(\Delta g^{F}_{I,j} - \Delta r^{F}_{I,j}\right) \end{cases} \qquad \text{Eq. 67}$$

The final changes to the red/green time can then be calculated using Eq. 61 for both conflicting routes.

A Simple Illustrative Example

To illustrate the FBP, we consider the following hypothetical scenario involving just two intersections (A and B). Calculations are illustrated in Table 35. Both intersections are in two-phase single-ring operation with a 90s cycle time. Only the route from A→B is oversaturated. The operation at A is affected by the queuing at B because of the limited link length between A and B (about 25 vehicles storage available or about 400 ft). Assume the speed limit is 35mph so the unimpeded travel time from A to B is about 8s.

Table 35. Illustration of calculation procedure

	Intersection A	Intersection B
Phase Green Time	45s	40s
Lost Time	8s	8s
Phase Red Time	37s	42s
Minimum Green	15s	15s
Minimum Green – Conflicting phase	15s	15s
Offset	0s	3s
Upstream storage space	100 vehicles	25 vehicles

Intersection A

Cycle	Green Time	TOSI	SOSI	Overflow Queue Length (per lane)
1	55s	0	0	0
2	52s	0.1	0	2
3	50s	0.2	0	3
4	45s	0.3	0	6
5	45s	0.5	0.1	15
6	45s	0.6	0.2	18
7	45s	0.7	0.2	22
8	45s	0.6	0.2	18
9	50s	0.2	0.1	7
10	55s	0.1	0	2

Intersection B

Cycle	Green Time	TOSI	SOSI	Overflow Queue Length (per lane)
1	40s	0	0	0
2	40s	0	0	3
3	40s	0	0	6
4	40s	0.1	0	6
5	40s	0.2	0	8
6	40s	0.3	0	9
7	40s	0.2	0	6
8	45s	0.1	0	3
9	50s	0	0	0
10	55s	0	0	0

To compute the recommended changes to the green times at A and B, first calculate the average TOSI and SOSI values over the study duration (neglecting zeros in determining the average). This is illustrated in Table 36 below.

Table 36. Illustration of calculations for green time modifications

	Intersection A	Intersection B
Average TOSI	0.33	0.18
Average SOSI	0.15	0
Average green time	49s	43s

	Intersection A	Intersection B
Forward Pass		
Delta-R	0s	-(0.15)*49s = -7.3s
Delta-G	(0.33 – 0.15)*49s = 8.8s	8.8s – 0s + 0.18*43s = 16.5s
Backward Pass		
Delta-R	0	0
Delta-G	Min(90s –15s – 8s –45s –8.8s, 90s –15s – 8s –40s –16.5s–7.3s, 0s) = 0s	Min(90s –15s – 8s –45s –8.8s, 90s –15s – 8s –40s –16.5s–7.3s, 0s) = 0s
Resulting adjustments		
Final Delta-R	0s + 0s = 0s	-7.3s + 0s = -7.3s (round to -7s)
Final Delta-G	8.8s + 0s = 8.8s (round to 9s)	16.5s + 0s = 16.5s (round to 17s)
Phase Red	37s - 9s = 28s	42s –(17s- (-7s)) = 18s
Phase Green	45s + 9s = 54s	40s + (17s- (-7s)) = 64s
Cycle Time	28 + 54s + 8s = 90s	18s + 64s + 8s = 90s
Offset	0s + 0s =0s	3s – 7s = -4s (or 86s)
Effective changes	Reduce side street phase 9s	Reduce side street phase 24s
	Increase phase green 9s	Increase phase green 24s
		Modify relative offset from 3s to 86s

Real-World Examples

In this section, two examples are provided for applying the FBP to real-world scenarios. In both cases, average values for TOSI and SOSI over the peak hour were used to re-calculate green times on the route. Simulation tests were then conducted with the same random seed to estimate the performance differences with and without applying the FBP.

An Oversaturated Intersection

The first test site is an oversaturated intersection located on TH55, in Minneapolis, MN (see Figure 130). Congestion usually occurs on the westbound traffic during the A.M. peak period because traffic signals are coordinated for the eastbound direction. Because of the short link between Winnetka and Rhode Island, queues spill back to the Rhode Island intersection creating large values of SOSI and TOSI. Signal timing plans for the intersections of Winnetka and Rhode Island are shown in Figure 131.

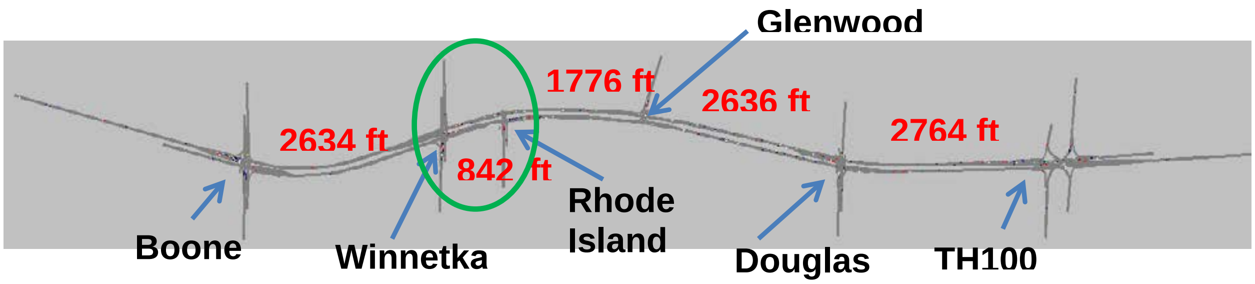


Figure 130. Test arterial on TH55, Minneapolis, MN

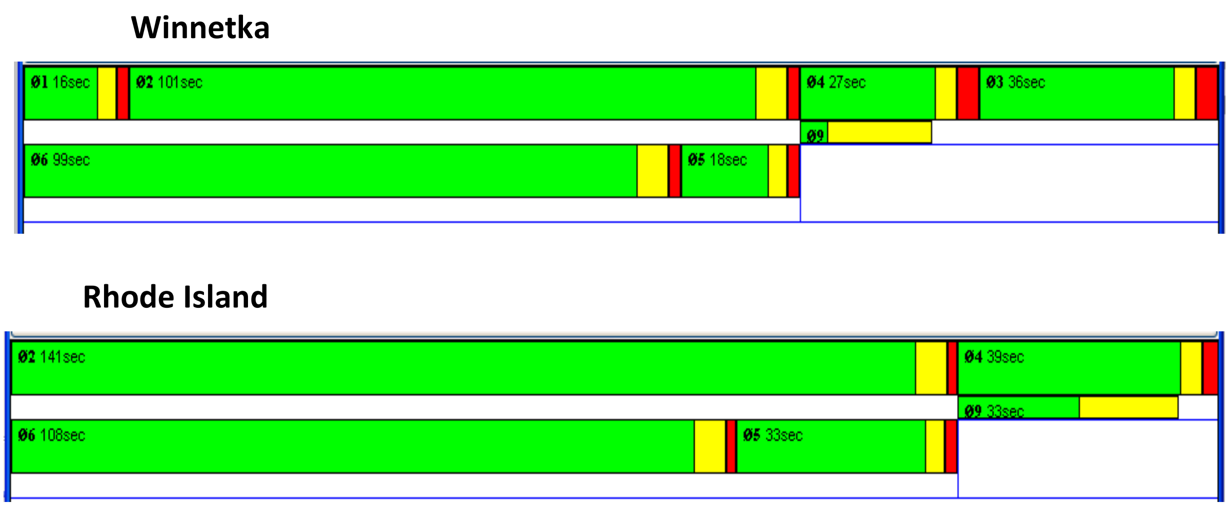


Figure 131. Signal timing plan (A.M. peak) for Winnetka and Rhode Island

To test whether the FBP can deal with this situation, we increased the traffic demands on all approaches by 50% in order to create severe oversaturation on the corridor. Figure 132 shows the SOSI and TOSI values at the Rhode Island intersection from a two-hour simulation period.

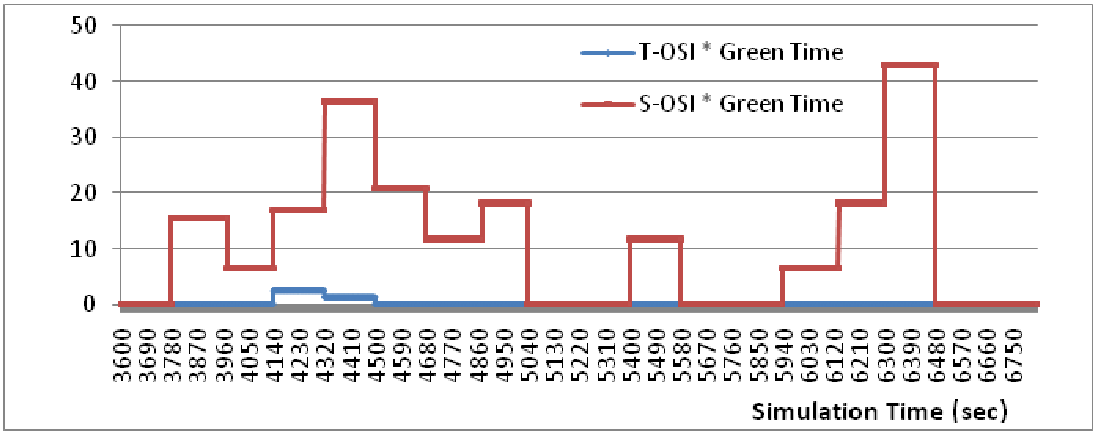


Figure 132. SOSI and TOSI values of Rhode Island westbound

Based on the average SOSI and TOSI values, the FBP suggests increasing the green time of the downstream intersection at Winnetka by 10s. Note that the FBP becomes a simple combination of the three strategies described earlier when applied to an individual phase. Figure 133 and Figure 134 show the SOSI and TOSI values before and after the change, respectively. From both figures, it is clear the TOSI and SOSI values are reduced. Figure 135 also shows the queue length at the downstream Winnetka intersection. After applying the FBP strategy, the queue length is significantly reduced. But as expected, the queue lengths on the side streets were increased due to the loss of green time (Figure 136).

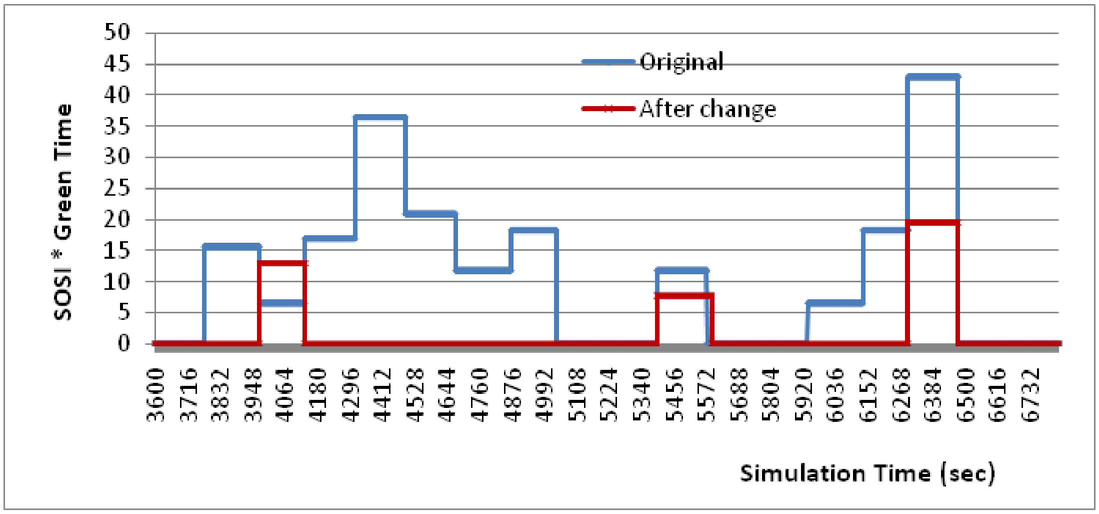


Figure 133. SOSI values of Rhode Island westbound before and after the FBP

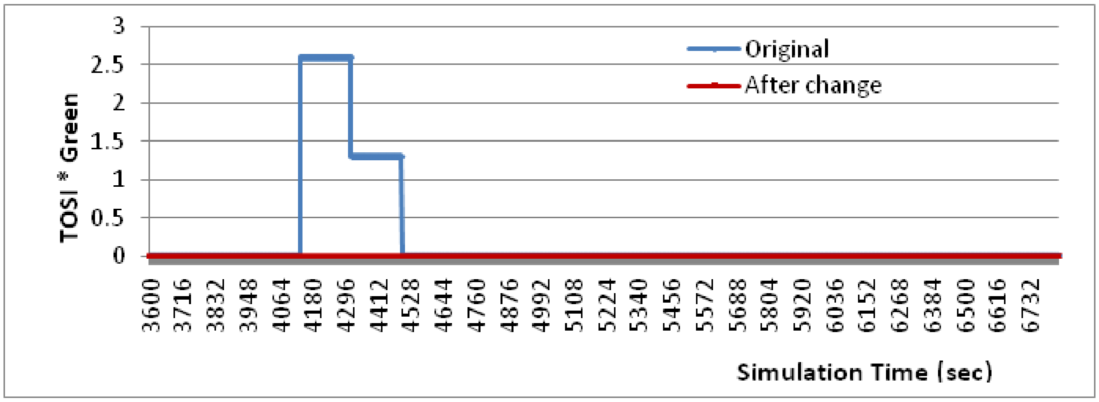


Figure 134. TOSI values of Rhode Island westbound before and after the FBP

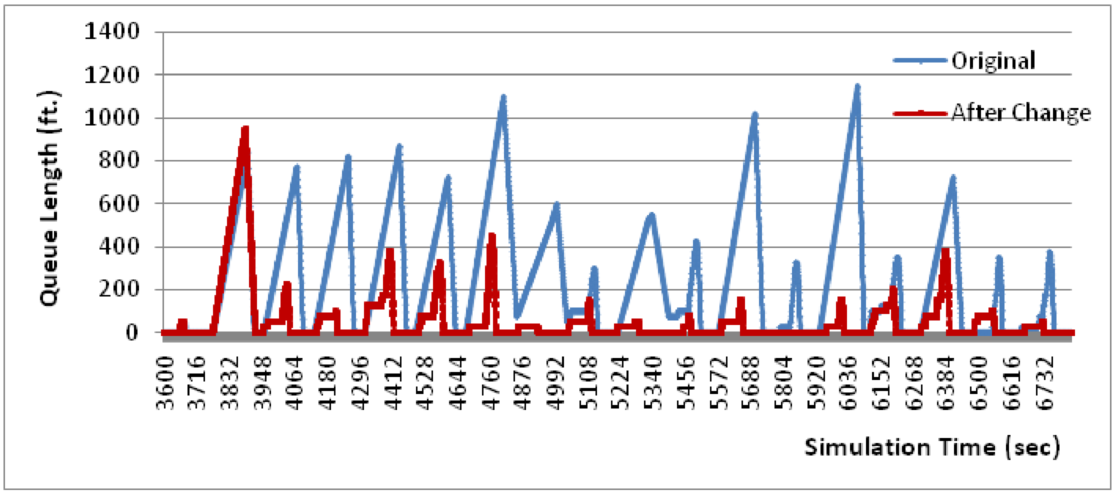


Figure 135. Estimated queue lengths at Winnetka westbound before and after the FBP

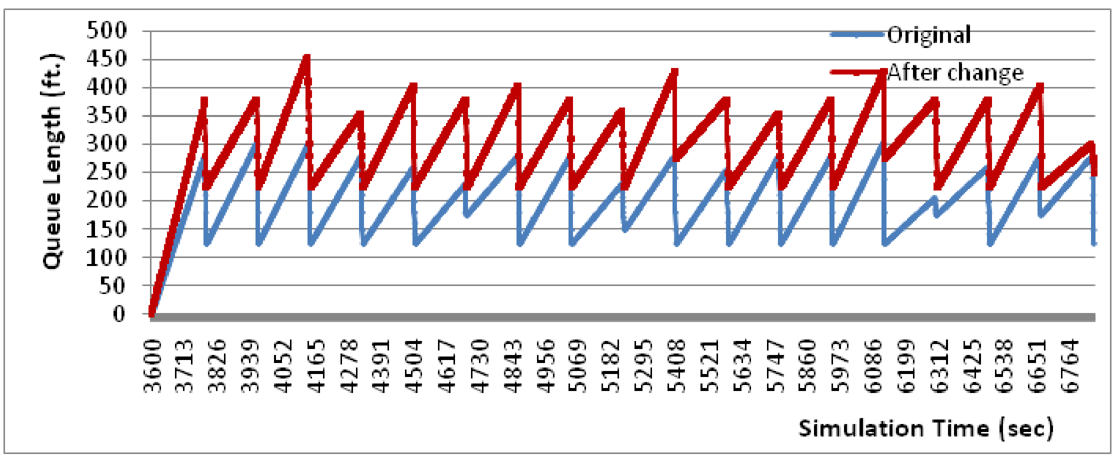


Figure 136. Queue lengths on the side street (southbound) at Winnetka intersection before and after the FBP

An Oversaturated Arterial

The second test site, as shown in Figure 137, is an oversaturated arterial corridor of five intersections on Fair Oaks Avenue in the City of Pasadena, CA. The length of the corridor is 0.4 mile and the speed limit is 30 mph. To create oversaturation, we assume that there is a large directional flow (3000 vph) from north to south. This may cause serious oversaturated conditions (spillovers and overflow queues) if the signal timings are not properly designed.

First, Synchro was used to optimize the signal timings according to the traffic volumes shown in Figure 137. The optimized cycle length is 140s and the north-south phase is the coordinated phase. The simulation was run ten times using different random seeds and each run was for two hours. The first half hour of the simulation was considered as the warm up time, and thus the average values of TOSI and SOSI in the following one and half hours were used to represent the oversaturation condition on the corridor. Since Synchro does not describe the queue interactions between intersections, it cannot deal with the spillover situation and always assumes that the discharging capacity at downstream is available. Therefore, severe oversaturation occurs under optimized timing plans provided by Synchro due to the large directional flow from north to south. Table 37 shows the southbound average SOSI and TOSI values over ten simulation runs and the green time duration of coordinated phase (i.e., the southbound direction). Note that spillover (i.e., SOSI >0) mainly happens at the southbound of Intersection 2 and on average 12.18% of the green time is wasted. This also causes a minor spillover at Intersection 1 with a 2.52% average SOSI value. At the same time, overflow queues (i.e., TOSI >0) exist at every intersection, which indicates insufficient discharging capacities.

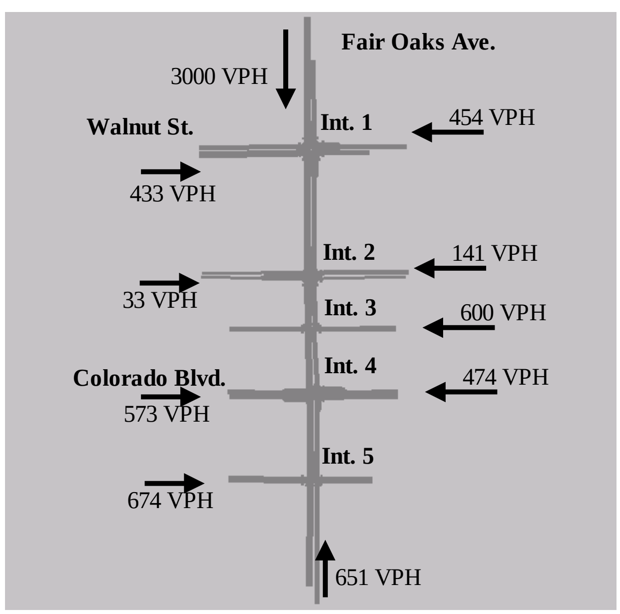


Figure 137.Vissim simulation network

Table 37. Southbound average SOSI and TOSI values under original signal timings

Inter. No.	Avg. *SOSI* (%)	Avg. *TOSI* (%)	Green Time (Sec)
1	2.52	13.59	83
2	12.18	5.45	110
3	0.99	4.46	81
4	0.00	0.85	94
5	0.00	2.93	98

Since oversaturation exists, the FBP is then applied and the calculation process is shown in Table 38. In the forward process (FP), Eq. (49), (50) and (51) are applied to determine $\Delta r_{n,i}^{F}$ and $\Delta g_{n,i}^{F}$ for each intersection. In the backward process (BP), residual capacity $R_{n,i}$ is calculated by Eq. 59 and the backward adjustment term Δg_{i}^{B} can be determined using Eq. (54). The final red time changes $\Delta r_{n,i}$ and green time changes $\Delta g_{n,i}$ are calculated by Eq. (55) and shown in Table 38. Table 39 compares the offset and green time duration (coordinated phase) of the plan provided by Synchro and the plan after the change of the FBP. One can see that the plan from the FBP modifies the offset value of each intersection and increases the green time duration of the coordinated phase for each intersection except Intersection 2.

By comparing the signal timing change before and after applying the FBP, we can clearly see the working "logic" of the FBP. The original TOSI/SOSI values indicate that the congestion begins from the link between Intersections 3 and 4 as indicated by the positive TOSI and SOSI values at Intersection 3 and significantly grows at Intersections 2 and 1. So the first step in the FBP is to increase the downstream capacity by reducing red time and increasing green time (see the FB column in Table 38). This explains why the green times at Intersections 3, 4, and 5 were significantly increased (see Table 39) even though they had very small TOSI/SOSI values (see Table 1). However, since the available green time is not enough for green expansion, the second step in the FBP is to "gate" the upstream intersections so the input demand can be reduced. This is why for Intersection 2 the green time was slightly reduced from 110sec to 109sec. With the green and red changes, the offset has been adjusted accordingly. Interestingly, the offset changes for Intersections 3, 4, and 5 generated by the FBP are identical (16sec, see Table 3). This indicates that the FBP results in a "green flare" solution.

Table 38. FBP calculation process

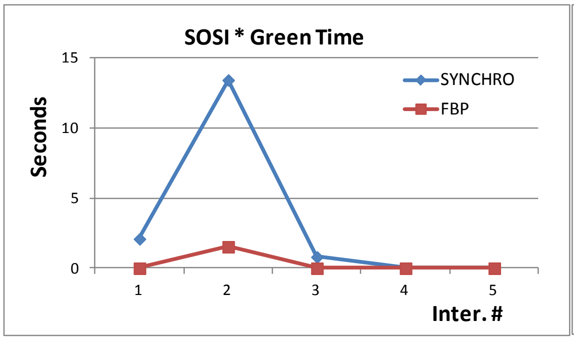
Inter. No.	$SOSI_{n,i} \times g_{n,i}$	$TOSI_{n,i} \times g_{n,i}$	$g^a_{n,i}$	FP		BP		$\Delta r_{n,i}$	$\Delta g_{n,i}$
				$\Delta r^F_{n,i}$	$\Delta g^F_{n,i}$	$R_{n,i}$	Δg^B_i		
1	2.1	11.3	14	0	9.2	4.8	-5.1	0	4.1
2	13.4	6.0	11	-2.1	1.8	7.1		-2.1	-3.3
3	0.8	3.6	15	-15.5	4.6	-5.1		-15.5	-0.5
4	0.0	0.8	22	-16.3	5.4	0.3		-16.3	0.3
5	0.0	2.9	22	-16.3	8.3	-2.6		-16.3	3.2

Table 39. Offset and green time of two plans

Inter. No.	Synchro		FBP		Change	
	Offset	Green Time	Offset	Green time	Offset	Green time
1	0	83	0	87	0	4
2	138	110	136	109	-2	-1
3	25	81	9	96	16	15
4	23	94	7	111	16	17
5	28	98	12	117	16	91

The average results for the TOSI and SOSI measures for southbound traffic under the two plans are compared in Figure 138. Overall, it is clear that the FBP plan improves the performance of the oversaturated route above and beyond what was calculated by Synchro. The FBP plan reduces the spillover time of the Intersection 2 (the major bottleneck on the route) from 13.4s to 1.5s and clears the spillover at Intersection 1. Because of the increase in discharge capacity, overflow queues at each link on the route have been greatly reduced as well. The average delay per vehicle of the FBP plan is 64s, which is a 12% decrease from the 73s average delay of the Synchro plan.

Table 40 summarizes the network performance under the two plans. With the FBP, the average number of stops has been reduced from 1.66 per vehicle to 1.31 per vehicle and the FBP plan increases the average speed 8.65%.



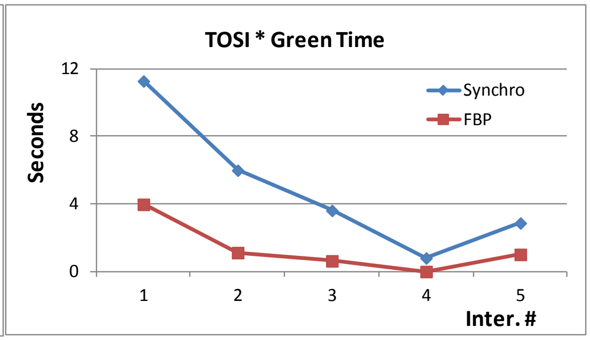


Figure 138.: Comparison of spillover time and overflow queue discharge time

Table 40. Network performance comparison

	SYNCHRO	FBP	CHANGE (%)
Average Delay (Seconds/per veh.)	73.14	64.07	-12.39
Average # of stops (per veh.)	1.66	1.31	-20.90
Average Speed (MPH)	11.13	12.09	+8.65

Figure 139 and Table 41 compare the throughputs of different exits of the network in the two-hour simulation period. The first group represents the throughput of the southbound exit, the second indicates the throughput of northbound exit and the remaining five groups represent the respective throughputs of the side streets at each intersection on the route. One can see that the throughput of the southbound exit, where the large directional goes to, is increased by 4.42% during the simulation time. Due to the decrease of green time on side streets, some side street throughputs are slightly decreased. Overall, the total throughput of the network is increased by 1.58%.

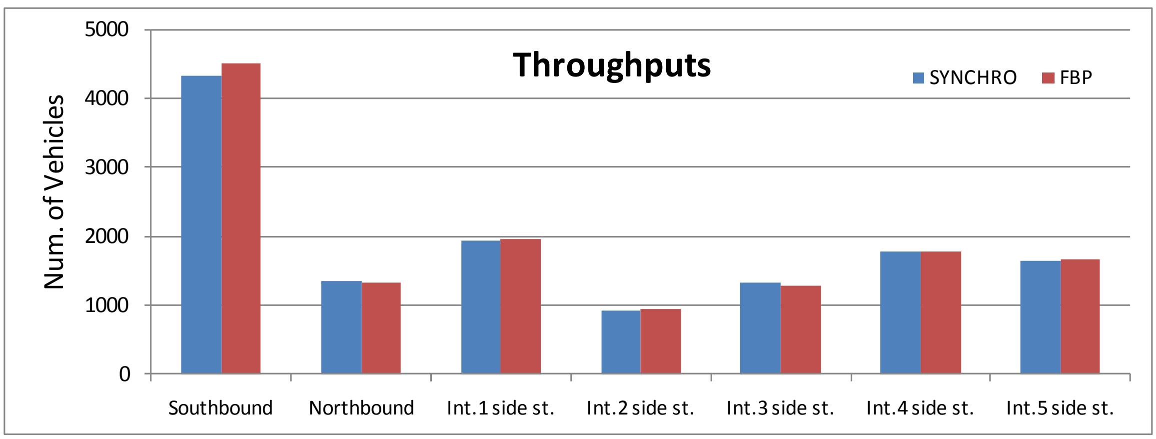


Figure 139. Comparison of throughput by route

Table 41. Comparison of throughput by route

	SYNCHRO	FBP	CHANGE (%)
Southbound	4315	4505.9	+4.42
Northbound	1342.2	1318.8	-1.74
Int. 1 side streets	1945	1967	+1.13
Int. 2 side streets	916.1	953.8	+4.12
Int. 3 side street	1319.6	1289.6	-2.27
Int. 4 side streets	1779.8	1771.9	-0.44
Int. 5 side street	1646.7	1666.7	+1.21
TOTAL	13264.4	13473.7	+1.58

Similarly, because of the decrease in the side street green time, the side street queue lengths under the FBP plan are longer than in the baseline plan. This is illustrated in Figure 140. Notably the increases to the side street queues are not significant since the green time adjustments were fairly modest at each intersection.

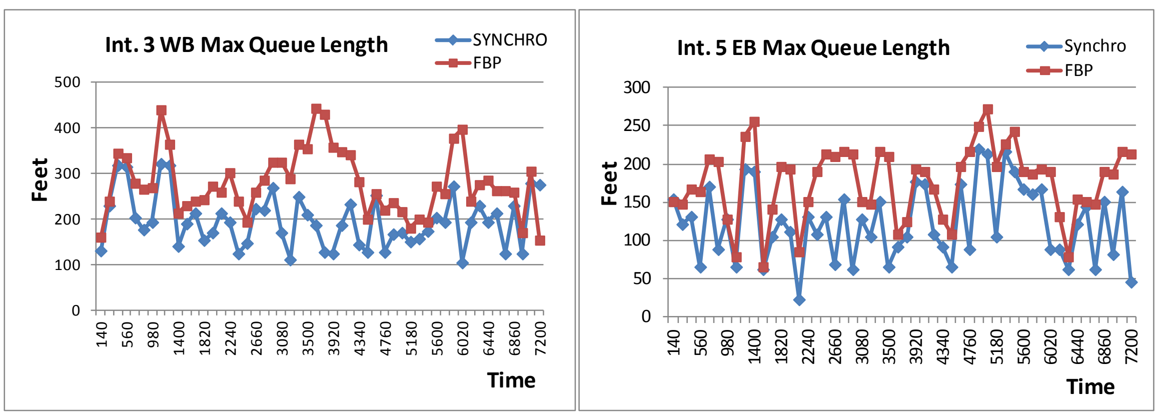


Figure 140. Comparison of side streets' maximum queue length in each cycle

Summary

A forward-backward procedure (FBP) was developed to adjust traffic signal timing to mitigate oversaturation on a route. The procedure is based on measured or estimated TOSI and SOSI values. The forward process aims to increase green time by searching for available green time which can be taken from side streets or conflicting phases to improve throughput for the oversaturated route. The backward process is used to gate some intersections to prevent overflow queues and downstream queue spillback when green time increases are insufficient. In the test cases, we calculated the average TOSI and SOSI values and applied a fixed adjustment to a new timing plan for the entire test duration. These tests indicated that the procedure can improve total delay and alleviate oversaturation along the oversaturated route. As expected, side street delays were increased to gain the system-wide improvements.

The proposed procedure was applied for offline signal timing adjustment, but we envision that a similar process could be applied in an online, real-time feedback manner with appropriate interface to the signal controller or a signal control system. The approach at this time would be considered experimental in nature. Further development in a number of directions would be necessary for more holistic application of the procedure in conjunction with the timing plan design principles and rules of thumb illustrated in other sections of this research.

In the next section, we will discuss a test case where the online tool was applied to a non-recurrent oversaturated scenario. Based on the status of several queue detection points, a variety of mitigations were evaluated that were envisioned to improve the network performance.

Online Application Test Case: Response to Incident at a Single Common Destination

In the previous section, we presented a bottom-up heuristic procedure for processing TOSI and SOSI measurements into green time adjustments for an oversaturated route. This method was shown to be effective in improving performance on the oversaturated route in two test cases by adjusting the green time of the timing plans in an offline manner.

The third component of our research methodology for addressing oversaturated conditions with signal timing plans is our development of an online tool that can use detector data including TOSI, SOSI, and queue length measures to select mitigation strategies in an online, feedback control system. This approach was described in Chapter 2. In this section, we describe a test case where this tool was evaluated. We first provided the background and motivation for the test followed by an outline of the methodology used to develop the mitigation strategies. Then we present the findings of the simulation tests and describe conclusions and future directions from the results.

The City of Windsor, ON is located southeast of Detroit, MI. Two border crossings in Windsor serve significant traffic flow between the U.S and Canada, the heaviest flows of any crossing at the northern border. In fact, over 65% of the truck traffic between the U.S. and Canada flows through this port of entry. The City of Windsor maintains approximately 300 signalized intersections and one limited access freeway facility. The City is a mixture of downtown grid and sub-urban arterials, with the majority of the signalized intersections in the downtown grid area. The approach to the Detroit-Windsor tunnel is in the heart of the downtown area just north of the intersection of Wyandotte and Goyeau Streets. These features are illustrated in Figure 141.

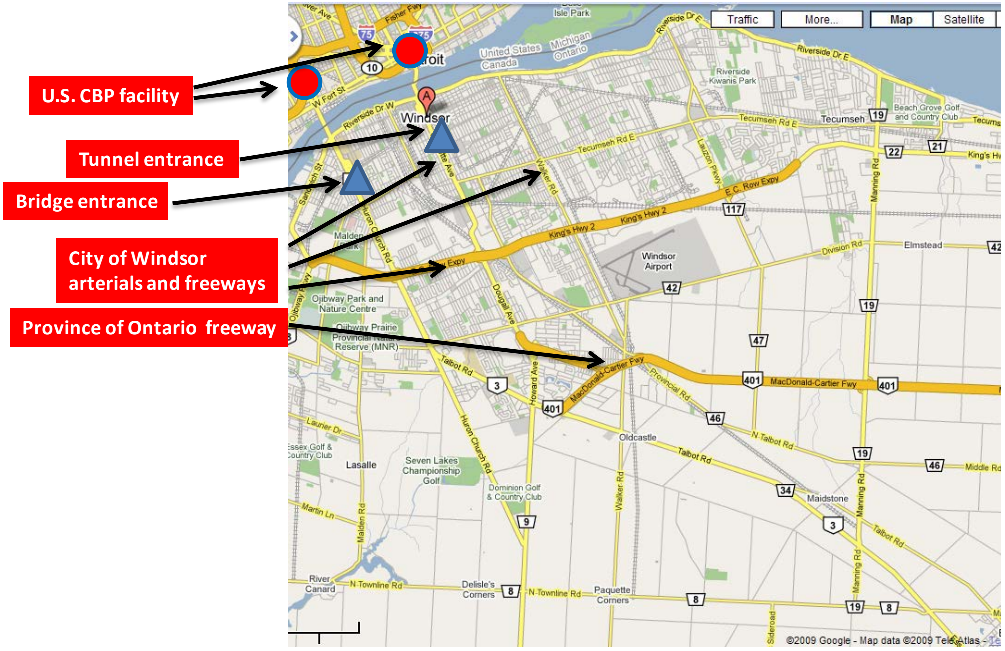


Figure 141. City of Windsor, ON arterial and freeway network

Tunnel operations can be significantly affected when the Customs and Border Protection (CBP) Agency of the U.S. changes policies. This occurs somewhat regularly during terrorist threats and other border protection events, as well as during inclement weather. In these situations, the traffic congestion at the approach to the tunnel can grow significantly and cause intersection blocking, inefficient signal timing, and major queuing and delays. At times, the City has deployed traffic police to the intersection to direct drivers and manages the backups. The intersection at the entrance to the Tunnel and surrounding intersections are illustrated in Figure 142.

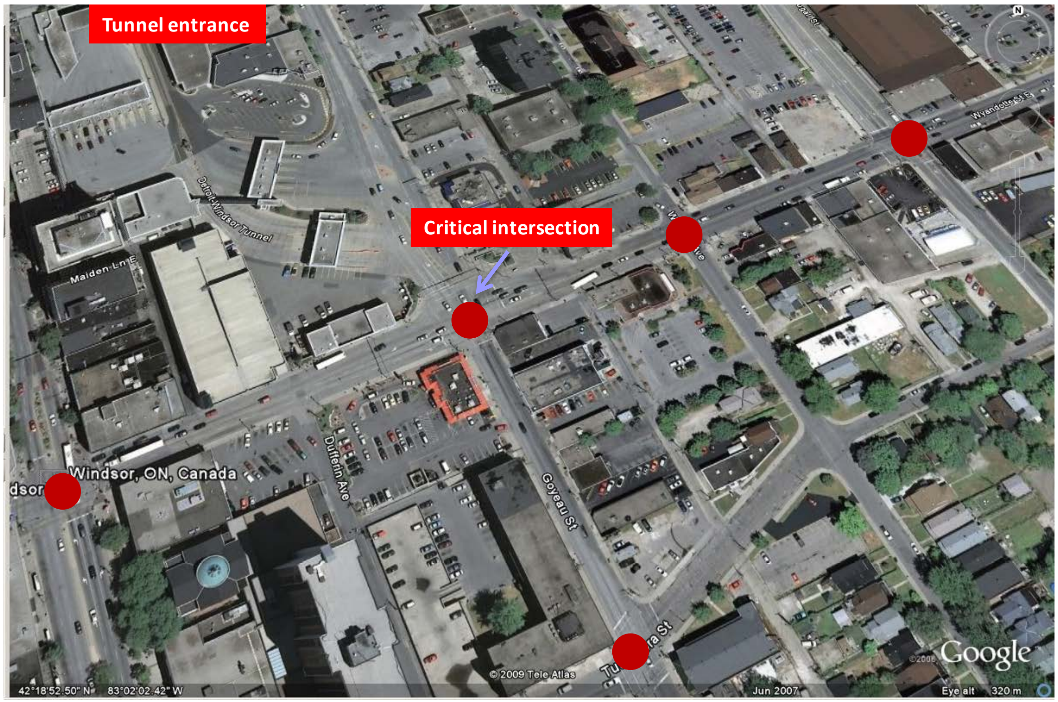


Figure 142. Intersections near the Detroit-Windsor Tunnel border crossing

Recently, the Province and the City have designed signal timing strategies to manage the operation of the intersection when the queues are persistent and oversaturation persists.   These strategies are proposed to be triggered by occupancy on queue detectors shown in Figure 143.

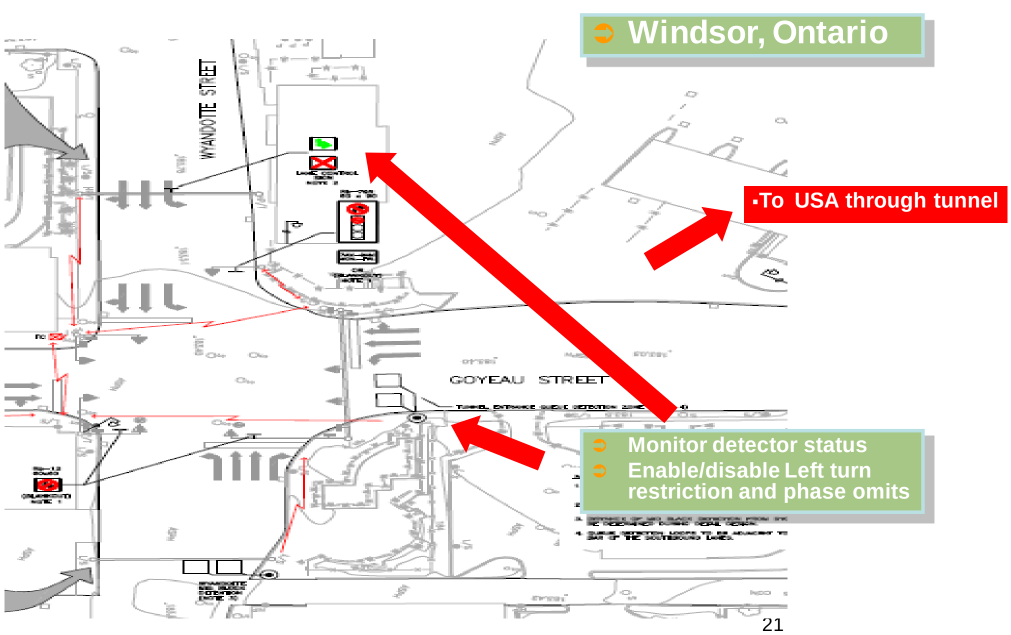


Figure 143. Detail of detector deployment and operational strategies at Tunnel entrance

Based on the status of the queue detectors, pedestrian indications, and other inputs, a number of mitigation actions are taken, as shown in the Table 42. This table is a truth table decision making device. The top part of the table lists the status of the inputs. Y indicates a condition is true, N indicates a condition is not true, and - indicates that it does not matter if the condition is true or false. In the top part of the table, each column of Y, N and - indications must be mutually exclusive to ensure that only one column of actions will be taken at a time. The bottom part of the table then shows which actions will be taken when a certain combination of conditions in the top part of the table is true.

Table 42. Proposed operational strategy at the Tunnel entrance

CONDITIONS	Queue detected on NB Goyeau Entry Link Detector	Y	Y	Y	Y	N	N	N
	Queue detected on NB Goyeau Mid-block Detector	-	-	-	-	Y	N	Y
	Queue detected on WB Wyandotte RT Turn Lane Mid-block Detector	-	-	-	-	N	Y	Y
	SB Vehicle Phase Call	-	Y	-	-	-	na	-
	N/S Ped Call on West Side	N	-	N	Y	-	na	-
	N/S Ped Call on East Side	N	N	Y	Y	-	na	-
ACTIONS	Omit NB Vehicle Phase	√	√					
	Extend NB Phase Green					√		
	Extend WB Phase Green						√	
	Display N/S Ped Walk on West Side		(√)		√	(√)		(√)
	Display N/S Ped Walk on East Side			√	√	(√)		(√)
	Display SB LT Turn Green Arrow		√					

Legend:

'Y'	Condition is present	'√'	Action is taken
'N'	Condition is not present	'(√)'	Action is taken if associated condition is present
'-'	Condition may or may not be present		
'na'	Condition is not applicable		

These combinations of conditions and actions are converted into if…then rules using the tool and procedure described in Chapter 2. In this specific case, the top three conditions describe three queue detection locations. Two are shown in Figure 143 (northbound Goyeau entry link and westbound Wyandotte right-turn lane) and the first location (northbound Goyeau entry link) is just to the left of the graphic in the northbound (to the right) direction. Various combinations of queues at each location in the top portion of the table prescribes certain actions to be taken to increase vehicle phase green time and omit turning movements. This strategy was developed to maximize throughput since each set of actions keeps traffic moving on movements that can still operate when queuing exists that blocks operations on other phases. Conceptually this logic seems very reasonable.

This logic only provides a mitigation strategy for the critical location. As part of the evaluation process we developed six other potential mitigation strategies and compared each strategy to the baseline operation to judge the performance and gain insight into the effectiveness.

Scenario Modeling

The baseline scenario for this test case consisted of creating an incident of one hour in duration caused by increased transaction time on the US side of the tunnel during the A.M. peak hour. During the incident, the average processing time for each customs transaction was increased from an average of two minutes to five minutes. This increase in processing time quickly results in traffic spilling back through the tunnel and out into the City streets.

The model was simulated in Vissim with all intersections under control of Virtual D4 controller firmware. The toll plaza and customs plaza were approximated in Vissim using the parking lot feature of the simulation system. Using this feature, vehicles that enter the toll plaza search for a "parking lot" with the most capacity which equates to the toll booth lane with the shortest line. Similarly at the customs plaza, each vehicle chooses the shortest line to be processed through customs. As is the case in the real-world, a subset of vehicles were designated as NEXUS vehicles and allowed to use their own special toll booth and customs plaza lanes. NEXUS is a pre-clearance system for border crossing that reduces the time to cross for very frequent travelers.

During the simulation, the volumes were increased according to the time of day schedule shown in Table 43.

Table 43. Schedule of volume changes during the scenario

Time of Day (Seconds into Simulation)	Volumes
6:30 (0000) - 7:00 (1800)	75% of Peak
7:00 (1800) - 8:30 (7200)	Peak Volumes
8:30 (7200) - 9:00 (9000)	75% of Peak
9:00 (9000) - 9:30 (10800)	25% of Peak

The one hour incident begins at 7:30 A.M. (3600s) and continues until 8:30 A.M. (7200s). The processing time for vehicles at the customs plaza is then returned to the original distribution with an average of two minutes. The processing time in the customs plaza equates to a dwell time in the "parking lot" that represents each inspection station along the vehicle's route.

The resulting input, output, and number of vehicles in the system are shown in Figure 144. Notice in the figure that the output curve is closely overlaid on the input curve up until the incident starts at 3600s. Then the output curve quickly begins to drop while the input rate remains constant, creating significant oversaturation in the network. After the incident is over, the input and output curves come back to closely matching their trajectories when the input volumes drop to 75% of the peak hour volumes. The overflow queues during this time do not dissipate quickly. It is not until the input rates are dropped to 25% of the peak hour flows do the queues dissipate and the system returns to steady state operation.

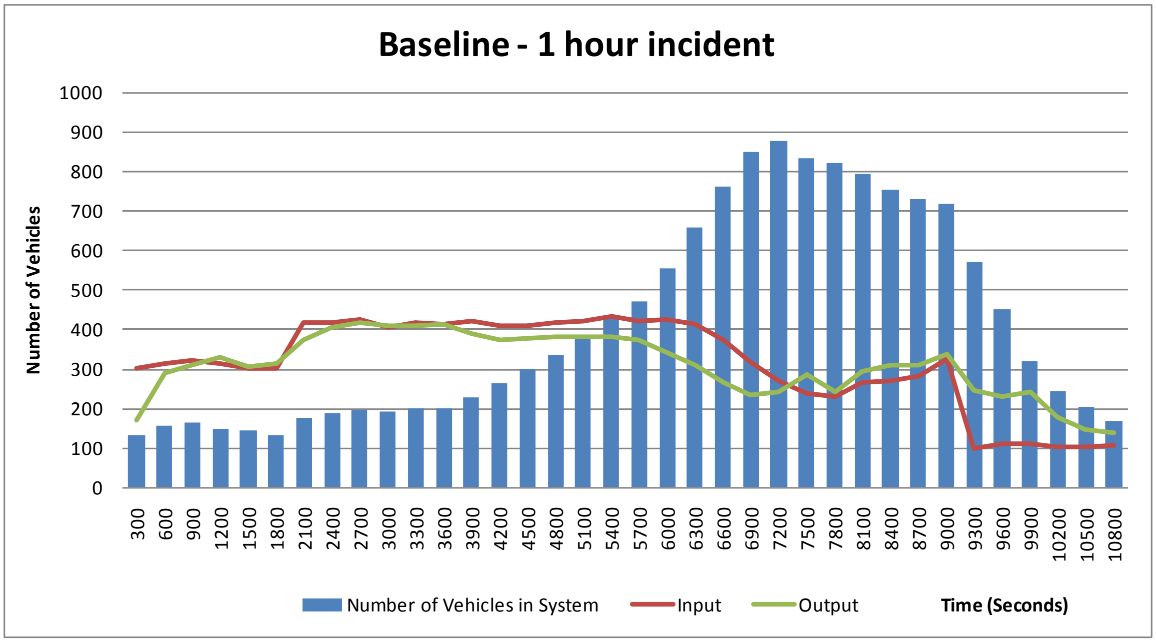


Figure 144. Baseline oversaturated scenario

A dynamic map was created to illustrate the spatial and temporal aspects of the oversaturated scenario over time. The resulting map is consistent with field observations at this location. This map was utilized to determine critical routes approaching the oversaturated intersection. Approximately 50% of the traffic entering the tunnel approaches from the west and turns north on to Goyeau. The remaining traffic is split, 25% each, between traffic approaching from the south on Goyeau or turning right from westbound Wyandotte. A small amount of NEXUS traffic approaches the tunnel from the north and turns right into the tunnel. During the A.M. and P.M. peaks, only NEXUS traffic is allowed to travel on southbound Goyeau.

The eastbound to northbound route is the most critical because it carries the most volume. During the baseline scenario the left-turn bay fills quickly and spills back to block through traffic on westbound Goyeau and intersections upstream as well. Northbound and eastbound queues at Wyandotte extend two intersections upstream as well.

The operational objectives that were identified for this test case include throughput maximization and queue management. Because the cause of the oversaturated condition blocks the receiving lane of the north leg of the critical intersection, increased green time for any of the oversaturated movements will not result in increased throughput. For this reason, mitigation strategies were chosen which had the potential for increasing the throughput of phases that do not enter the plaza (eastbound and westbound through movements). For queue management, the focus was on the northbound and eastbound queues which extend to intersections adjacent to the critical intersection. Additional mitigation strategies were envisioned in addition to the logic developed by the Province. Table 44 lists the mitigation strategies that were tested in this scenario.

Table 44. Summary of mitigation strategies

Mitigation Number	Name	Description
1	Original Mitigation Logic	Omit phases and adjust green times at critical intersection
2	Expanded Mitigation Logic	Omit phases and adjust green times at critical intersection
3	Dynamic Lane Assignment	Allow double left-turns on critical eastbound route
4	Westbound and Northbound Metering	Meter traffic upstream of critical intersection on non-critical routes
5	Eastbound Metering	Meter traffic upstream of critical intersection on critical route
6	Westbound, Northbound and Eastbound Metering	Meter traffic everywhere
7	Re-routing	Re-route traffic around critical intersection to alternate entrance to store vehicles on longer route

These mitigation strategies are summarized in the next section.

Intersection Strategies

Two strategies were developed that adjust the operation only at the critical intersection. The first strategy was envisioned by consultants for the City and Province several years before this project (March 1990) began. This logic was implemented and tested on the incident scenario described here. In addition, we developed a more comprehensive set of logical conditions that would take additional actions based on the inputs from additional queue detectors. These two intersection logic approaches are described in the next sections.

Original Windsor Incident Mitigation Logic

This mitigation strategy used the Congestion Manager in conjunction with the Virtual D4 controller to implement different strategies based on detection of queues at three locations. Other detection points as shown in Figure 145were envisioned to be used for more extensive feedback strategies.

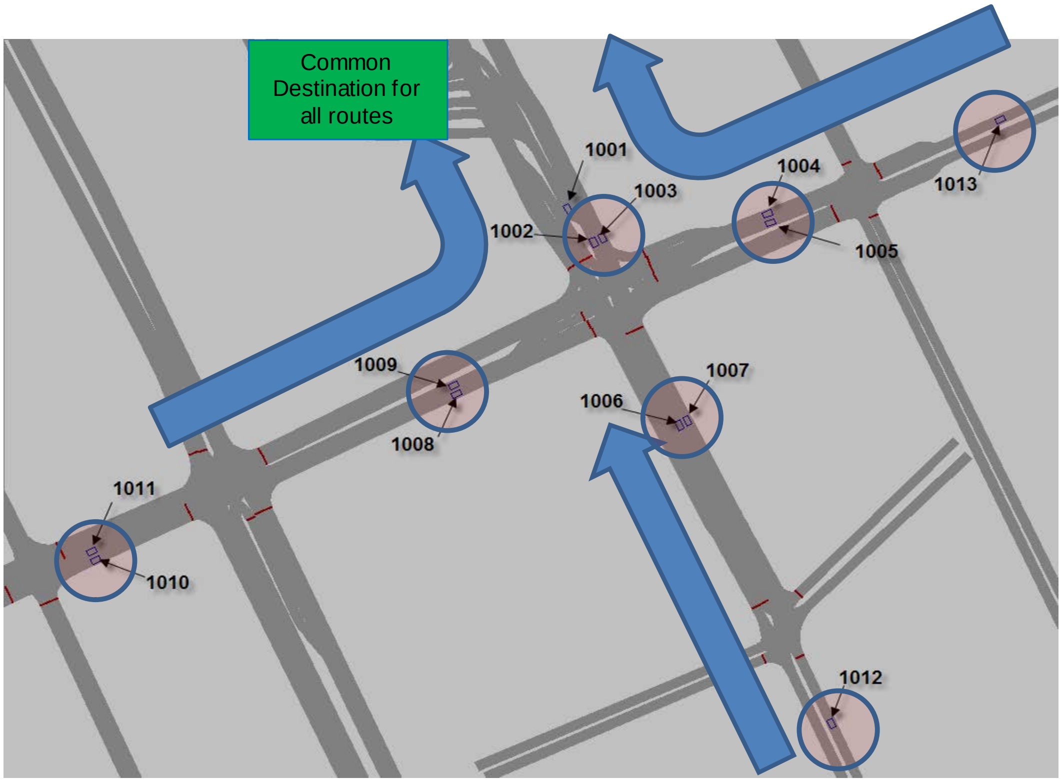


Figure 145. Detection points in the Windsor, ON traffic network

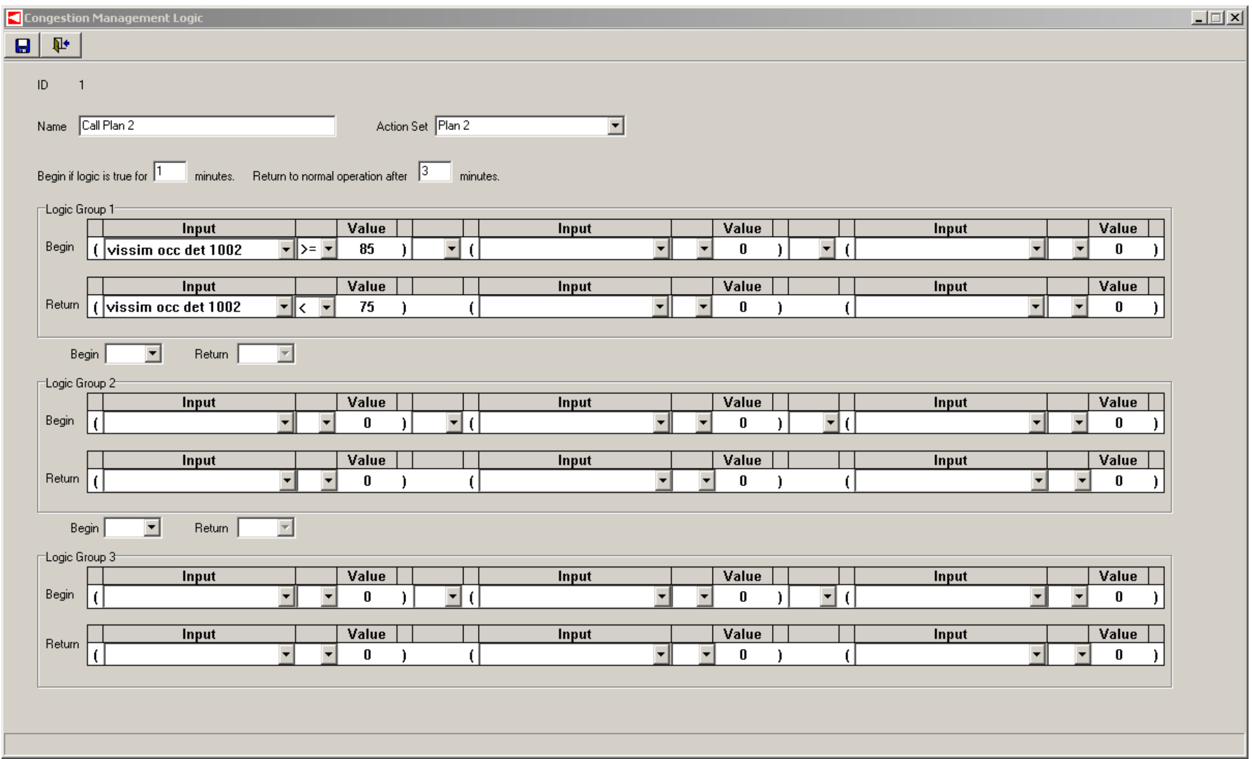
From the original responsive logic presented in Table 44, we removed the consideration of pedestrian concerns and developed the following logic table based just on the status of the queue detectors. See Table 45 below.

Table 45. Windsor queue-responsive logic

Condition	Queue detected on northbound Plaza Entry Link (Vissim Detector 1002)	Y	N	N
	Queue detected on northbound Goyeau Link (Vissim Detectors 1006, 1007)	-	Y	N
	Queue detected on westbound Right Turn at Goyeau (Vissim Detector 1004)	-	N	Y
Action	Increase northbound Through Phase		√	
	Increase westbound Through Phase			√
	Omit northbound Through Phase	√		
	Action Set Plan	**2**	**3**	**4**

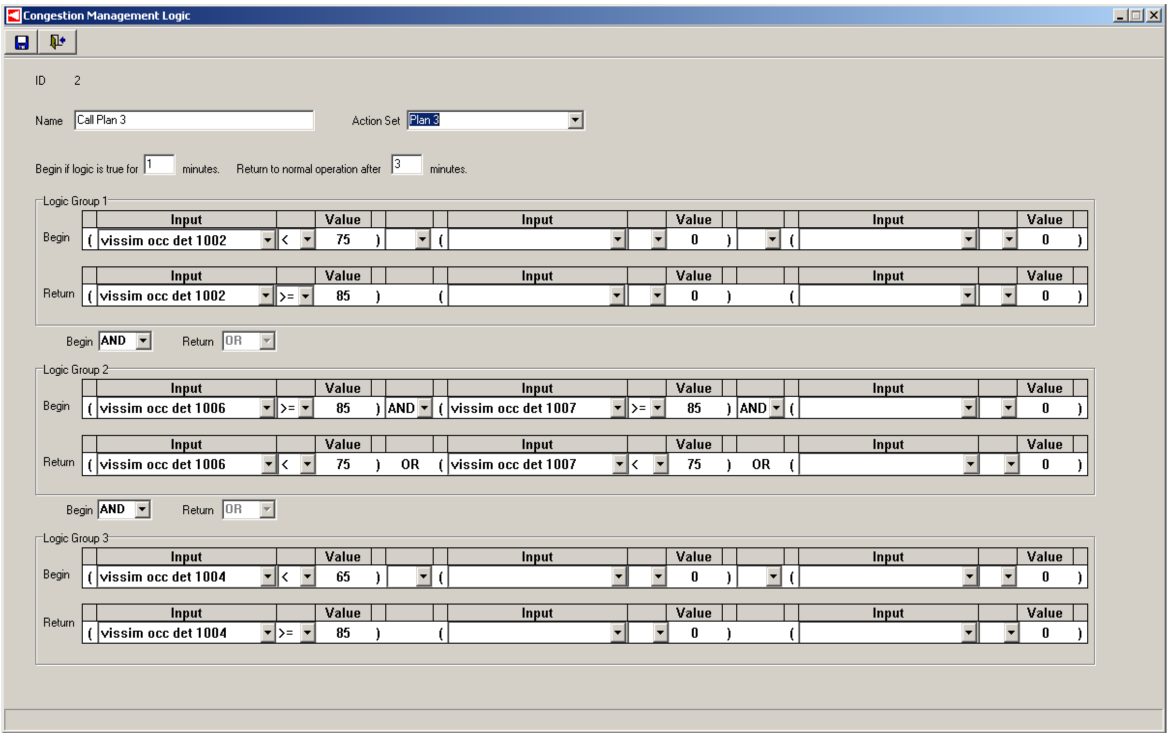
A detection point was created in the Vissim model for each of the locations to be monitored for conditions as shown in Figure 145.

For each action included in the table above, a coordination timing plan (Action Set Plan) was created in the Virtual D4 controller to be called by the oversaturation management logic when the triggering condition is met. The logic statements used for this strategy are depicted in Figure 146 a, b, and c. For each logic statement, the condition must be detected as true for one minute before the Action Set Plan is initiated. The Action Set Plan remains in use until the return condition is met for three minutes. We set the begin threshold for each queue detector to 85%. This correlates to a TOSI value of 100% or more. We set the return condition for each queue detector to 75%. This correlates to a TOSI value of approximately 50%. Since the queues grow back quickly and saturate the entire set of eastbound links, once the incident begins, the value of TOSI is 100% for the majority of the hour. SOSI is similarly greater than zero for most of the time at almost all of the detection locations during the simulation.

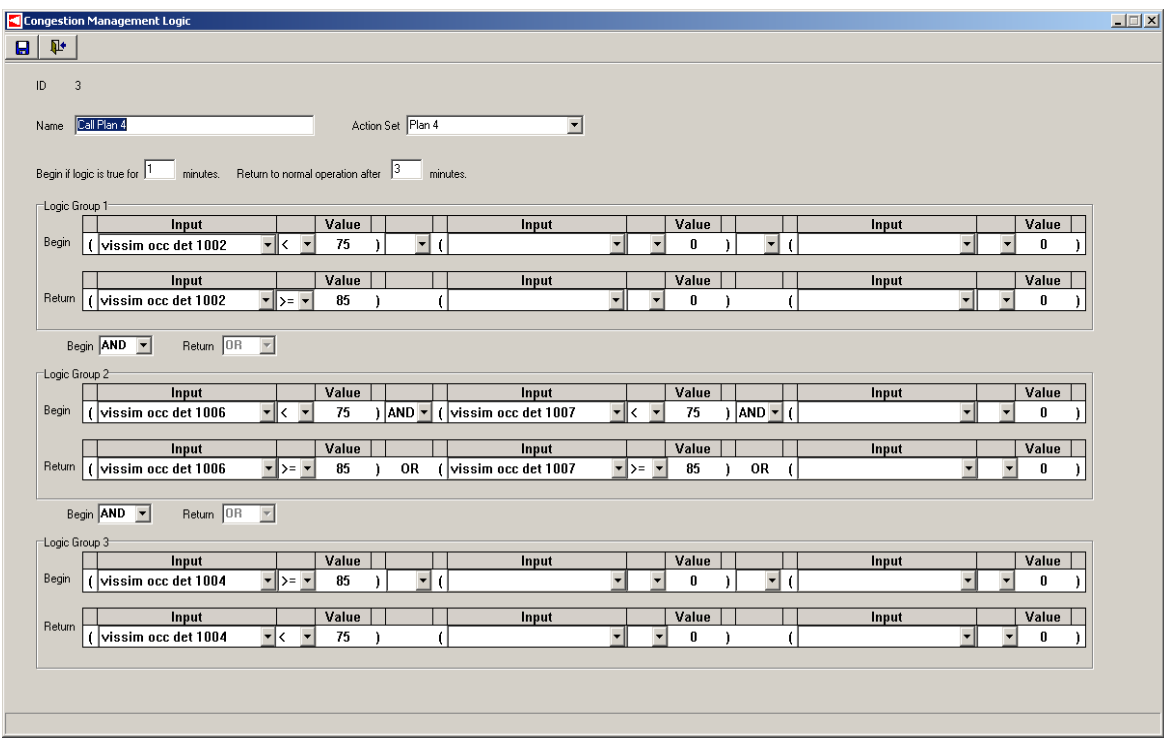


(a)

Figure 146. Logic engines for (a) Plan 2 [omit eastbound left turn] (b) Plan 3 [increase northbound through] (c) Plan 4 [increase westbound through]



(b)



(c)

Figure 146. ***continued***

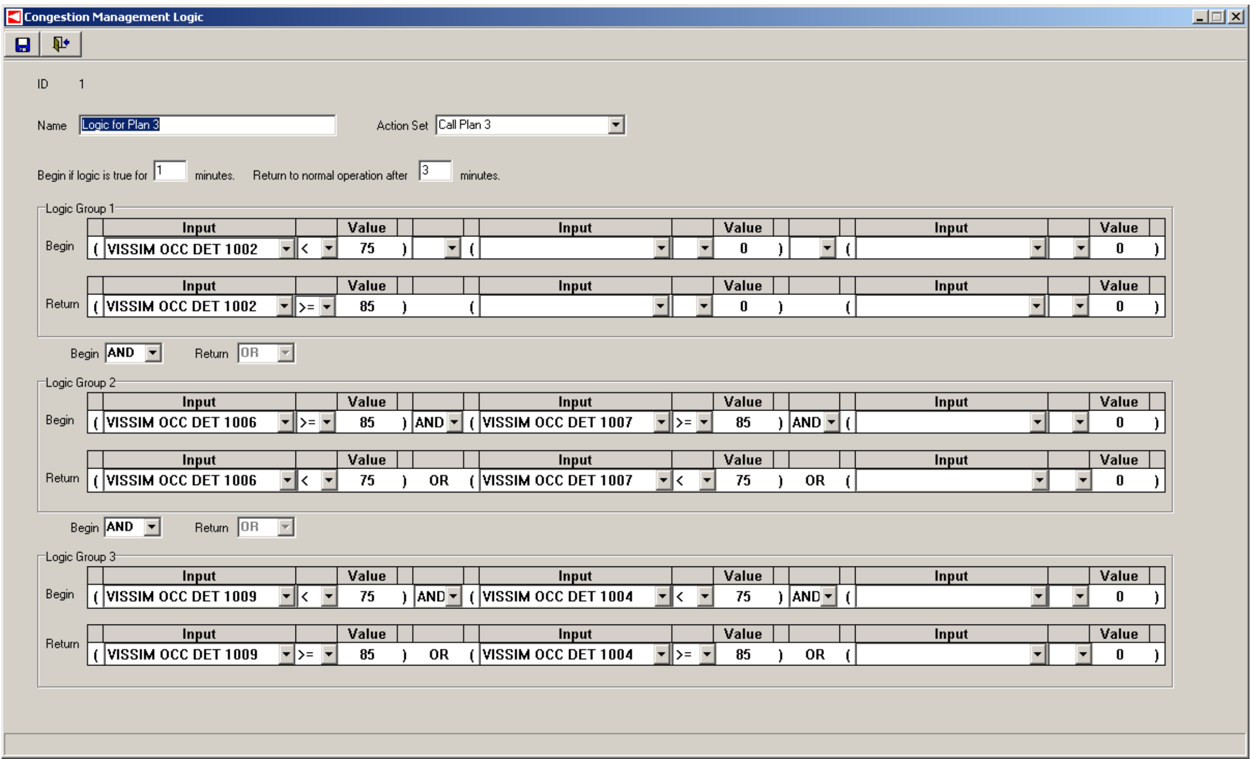
Expanded Incident Mitigation Logic

The original mitigation logic was expanded to include consideration of queue detection on the eastbound approach of the Goyeau/Wyandotte intersection. Additional actions were also designed to increase the throughput of phases that do not enter the border crossing tunnel. In addition, the eastbound right-turn-on-red was eliminated during conditions where the plaza entry link experienced queuing. This action was chosen based on observations of the baseline conditions. In the baseline scenario, the westbound right turning vehicles were consistently turning RTOR during small gaps. This resulted in the entry link for the plaza filling up and then creating starvation for the critical eastbound left-turn movement. Table 46 summarizes the Expanded Windsor Logic and identifies the Vissim detectors associated with each condition. The actions were included in each response plan, and the plan number assigned to the associated group of actions.

Table 46. Expanded Windsor Logic

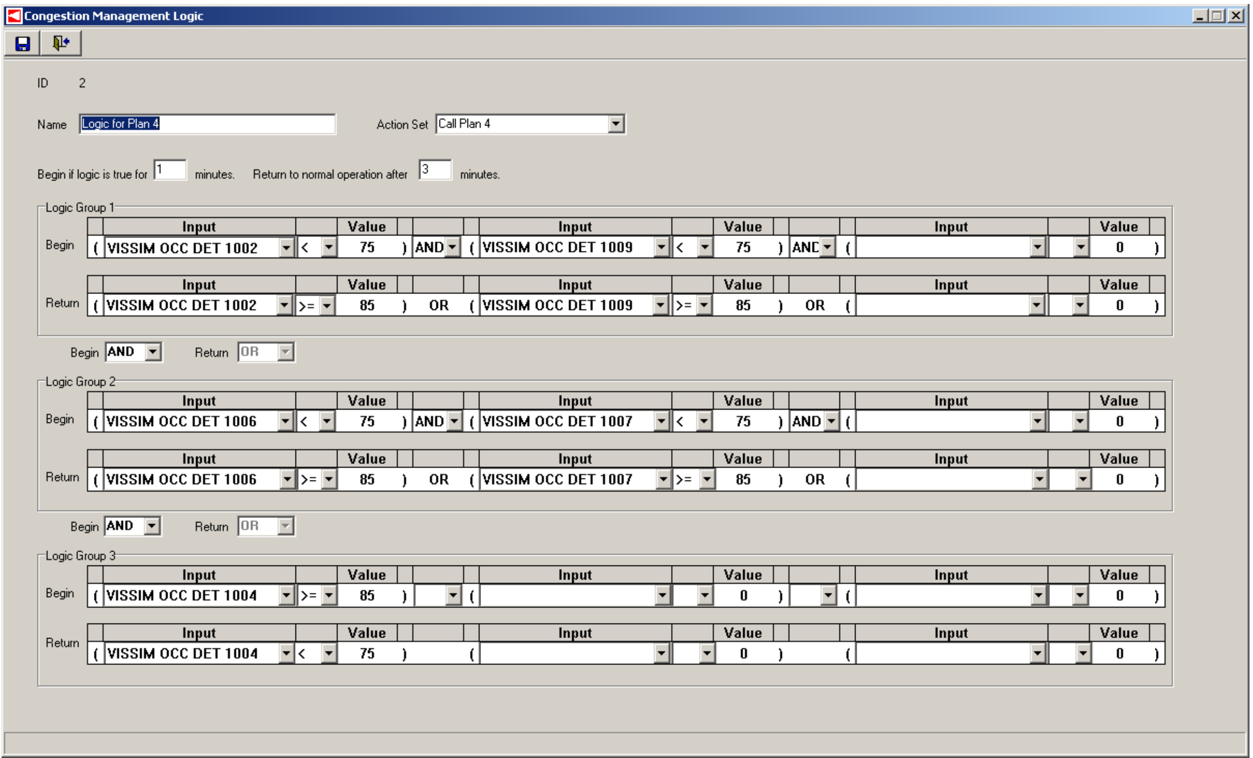
Condition	Queue detected on northbound Plaza Entry Link (Vissim Detector 1002)	Y	Y	Y	Y	N	N	N	N	N	N
	Queue detected on northbound Goyeau Link (Vissim Detector 1006, 1007)	Y	Y	N	N	Y	N	N	Y	Y	N
	Queue detected on eastbound Left Turn at Goyeau (Vissim Detector 1009)	Y	N	Y	N	N	Y	N	Y	N	Y
	Queue detected on westbound Right Turn at Goyeau (Vissim Detector 1004)	-	-	-	-	Y	Y	Y	N	N	N
Action	Increase eastbound Left Turn Phase						√		√		√
	Increase northbound Through Phase					√				√	
	Increase westbound Through Phase					√	√	√	√		
	Eliminate right-turn-on-red for westbound Right Turn	√	√	√	√						
	Omit northbound Through Phase				√						
	Omit eastbound Left Turn Phase		√	√	√						
	Action Set Plan	**5**	**6**	**7**	**8**	**9**	**10**	**4**	**11**	**3**	**12**

Action Set Plans 3 and 4, developed for the previous responsive strategy, were also used for the expanded logic as the resulting actions were the same. The logic statements used for this strategy are depicted in Figure 147a-j below. Similarly to the logic statements used in the original responsive strategy, the condition must be detected as true for one minute before the Action Set Plan is initiated. The Action Set Plan remains in use until the return condition is met for three consecutive minutes. We set the "begin" threshold for each queue detector to 85%. This correlates to a TOSI value of 100% or more. We set the return condition for each queue detector to 75%. This correlates to a TOSI value of approximately 50%. Since the queues grow back quickly and saturate the entire set of eastbound links, once the incident begins, the value of TOSI is 100% for the majority of the hour. SOSI is similarly greater than zero for most of the time at almost all of the detection locations during the simulation. We could not justify using TOSI or SOSI values for triggering different actions for this particular test case.

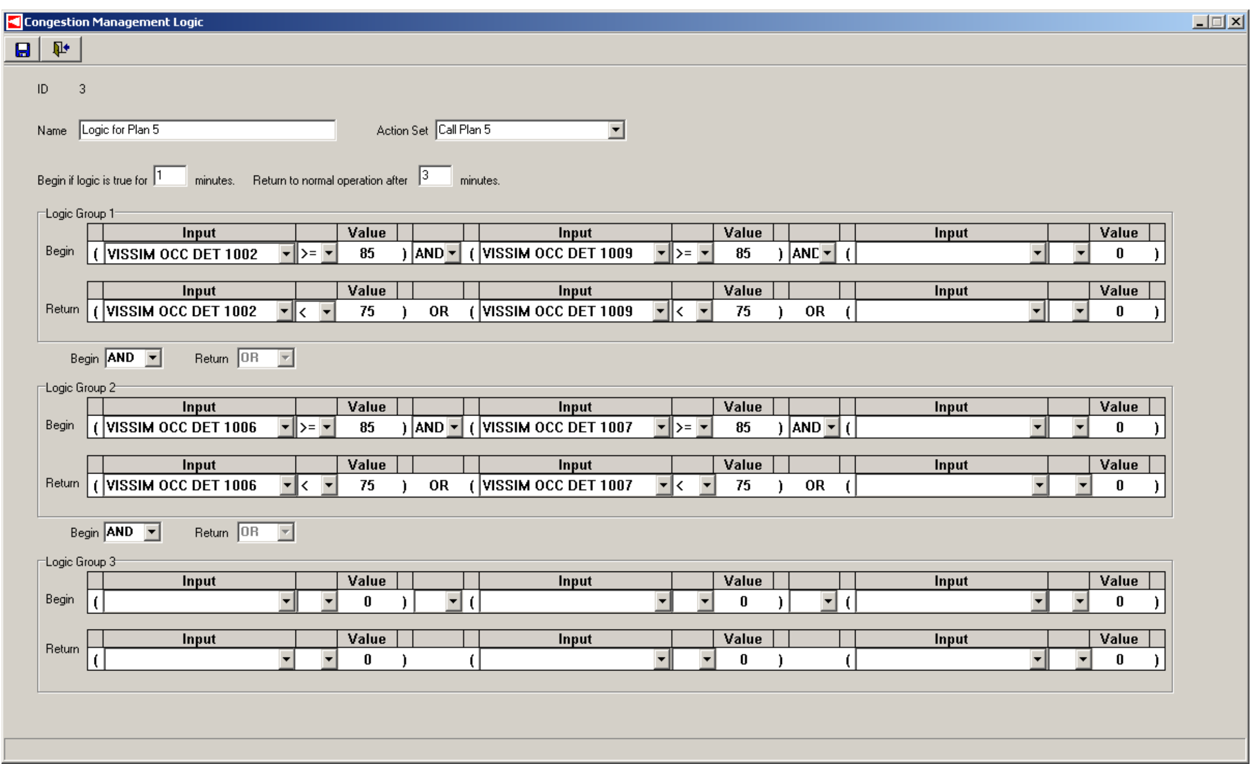


(a)

Figure 147. Logic engines for expanded mitigation logic

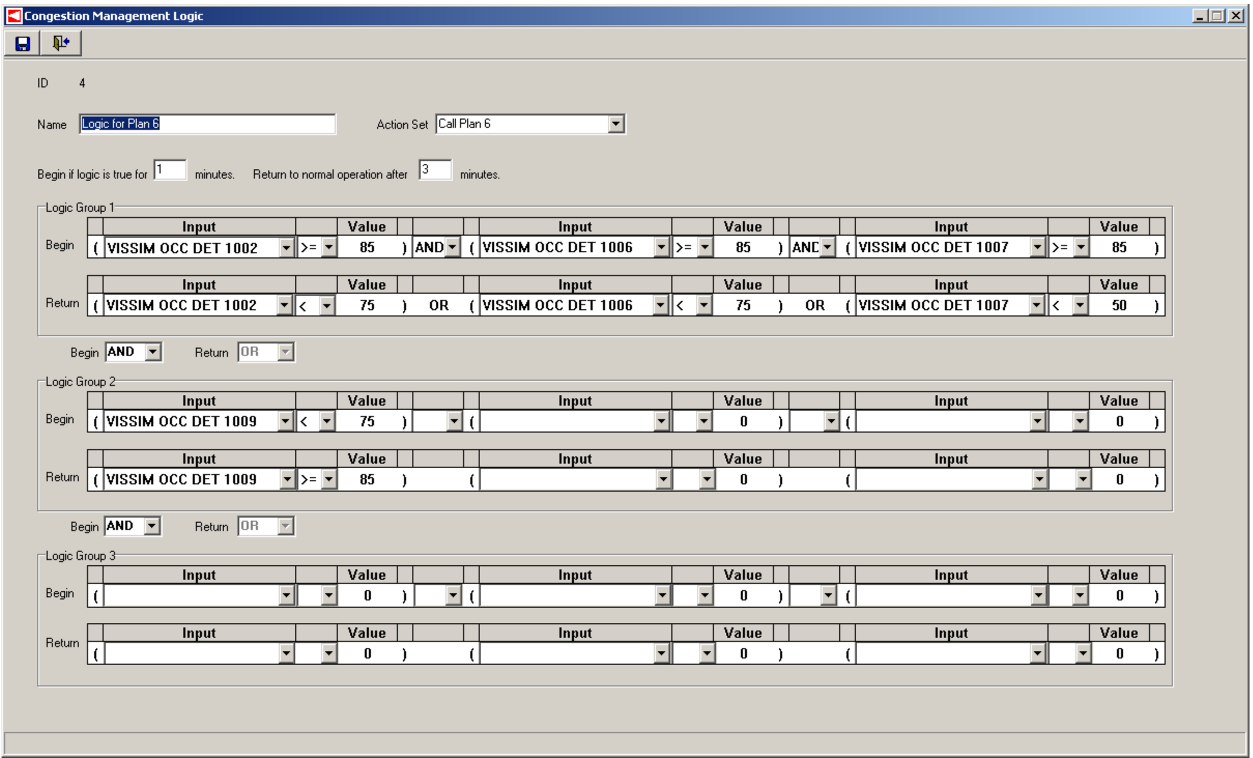


(b)

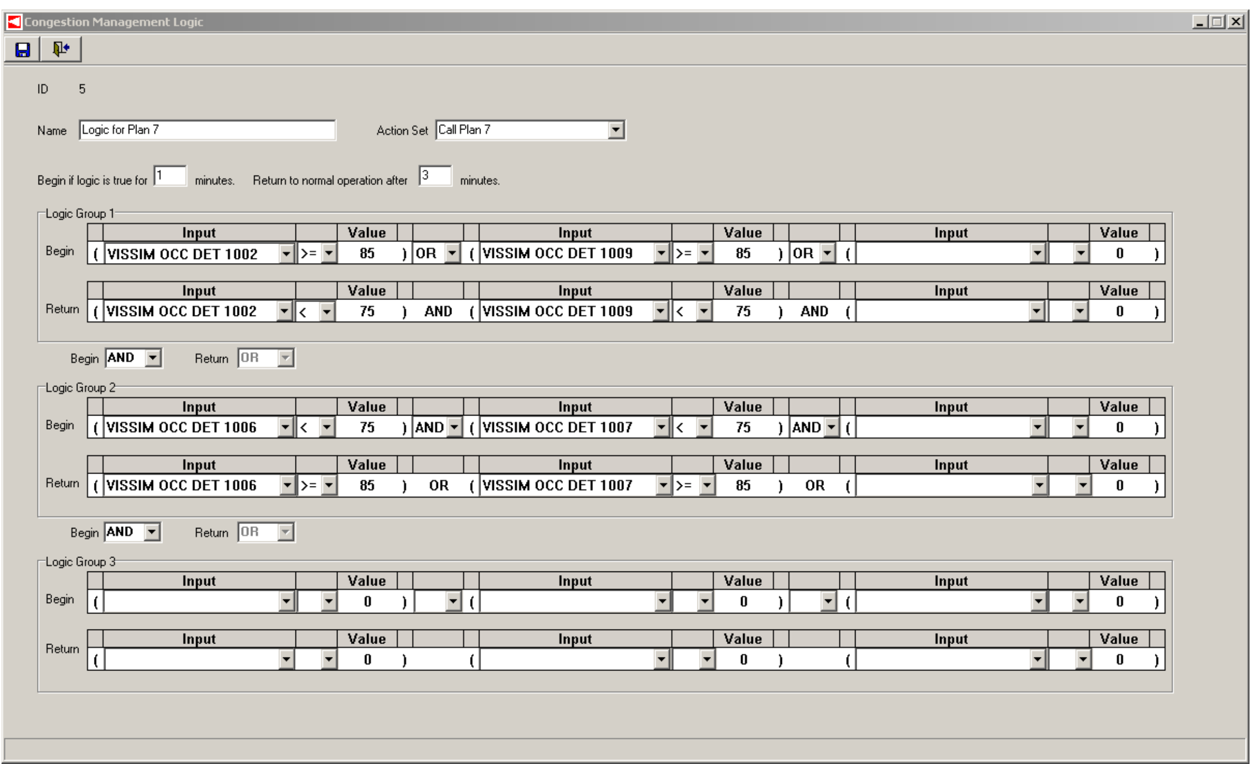


(c)

Figure 147. *continued*

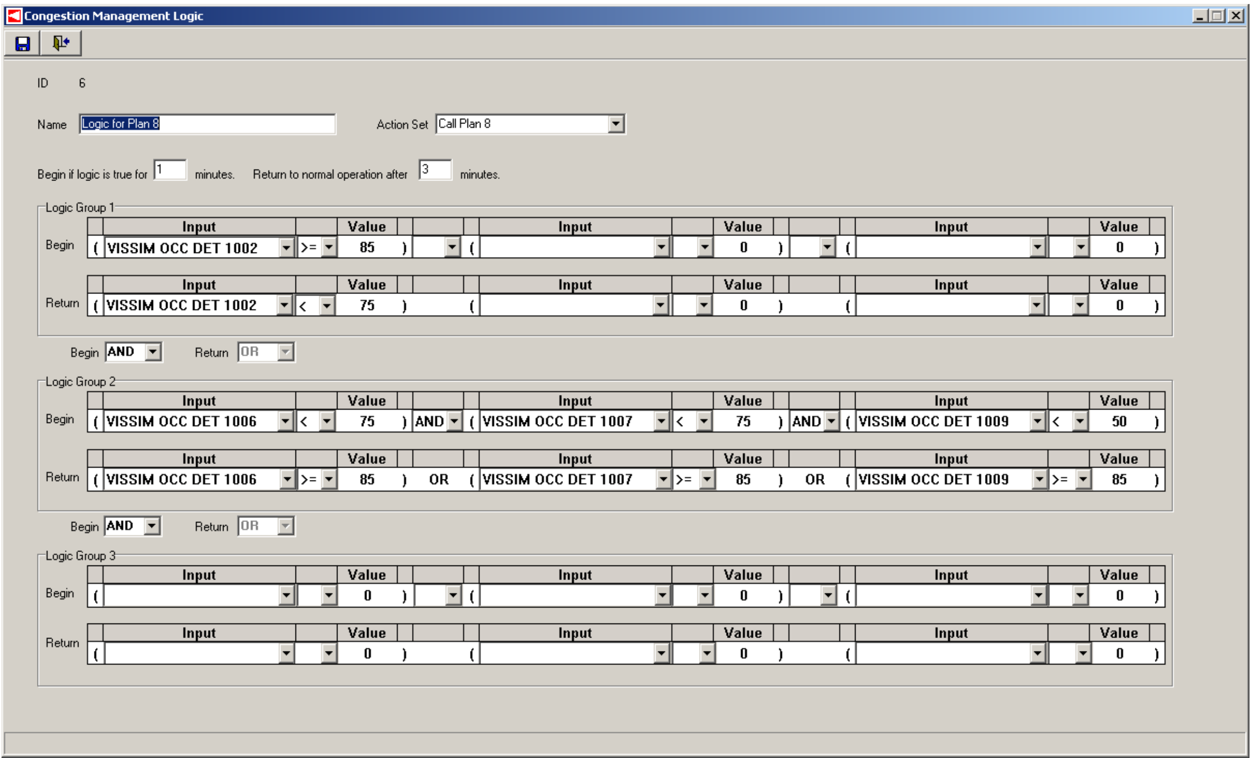


(d)

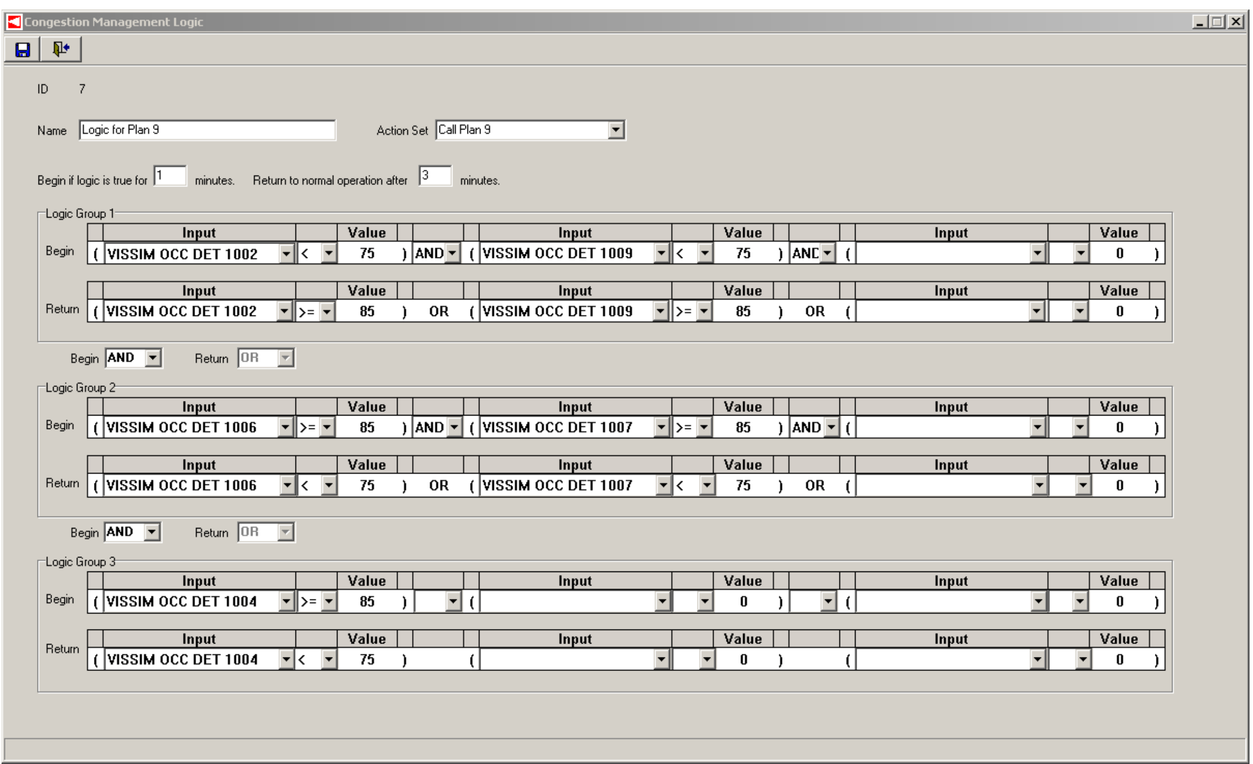


(e)

Figure 147. ***continued***

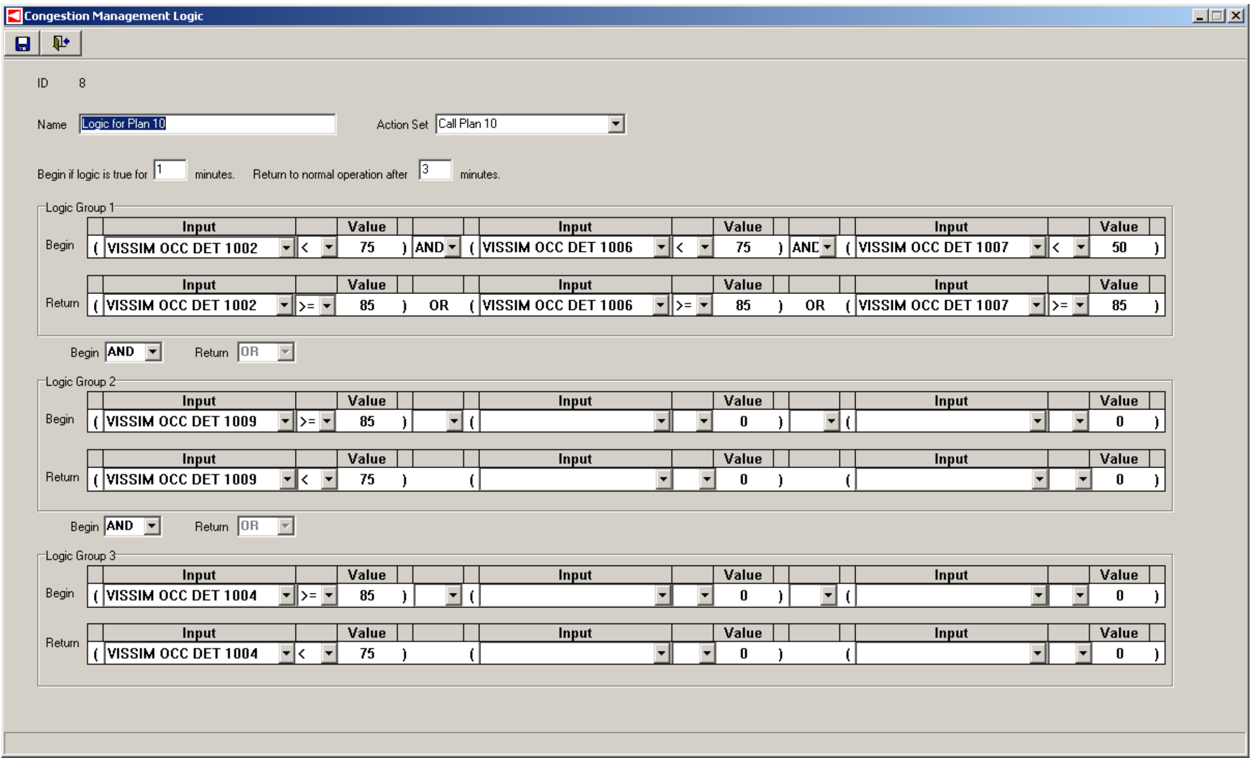


(f)

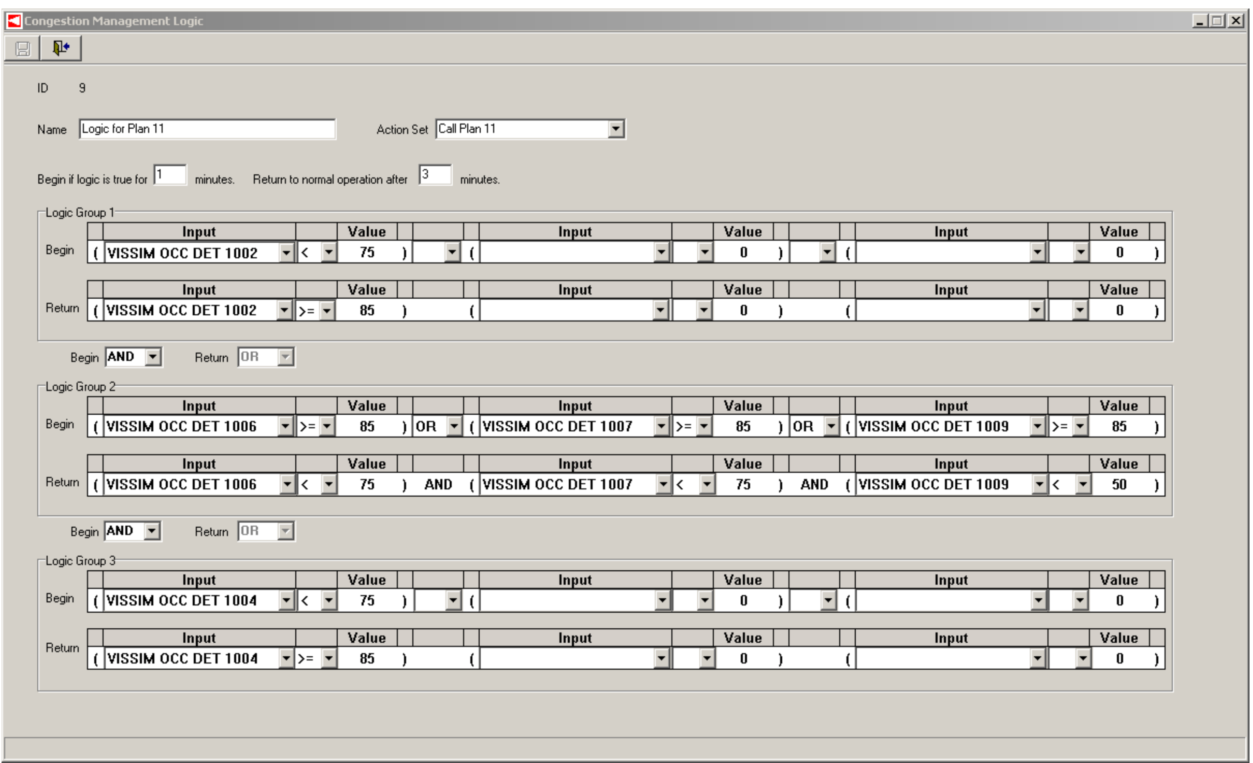


(g)

Figure 147. ***continued***

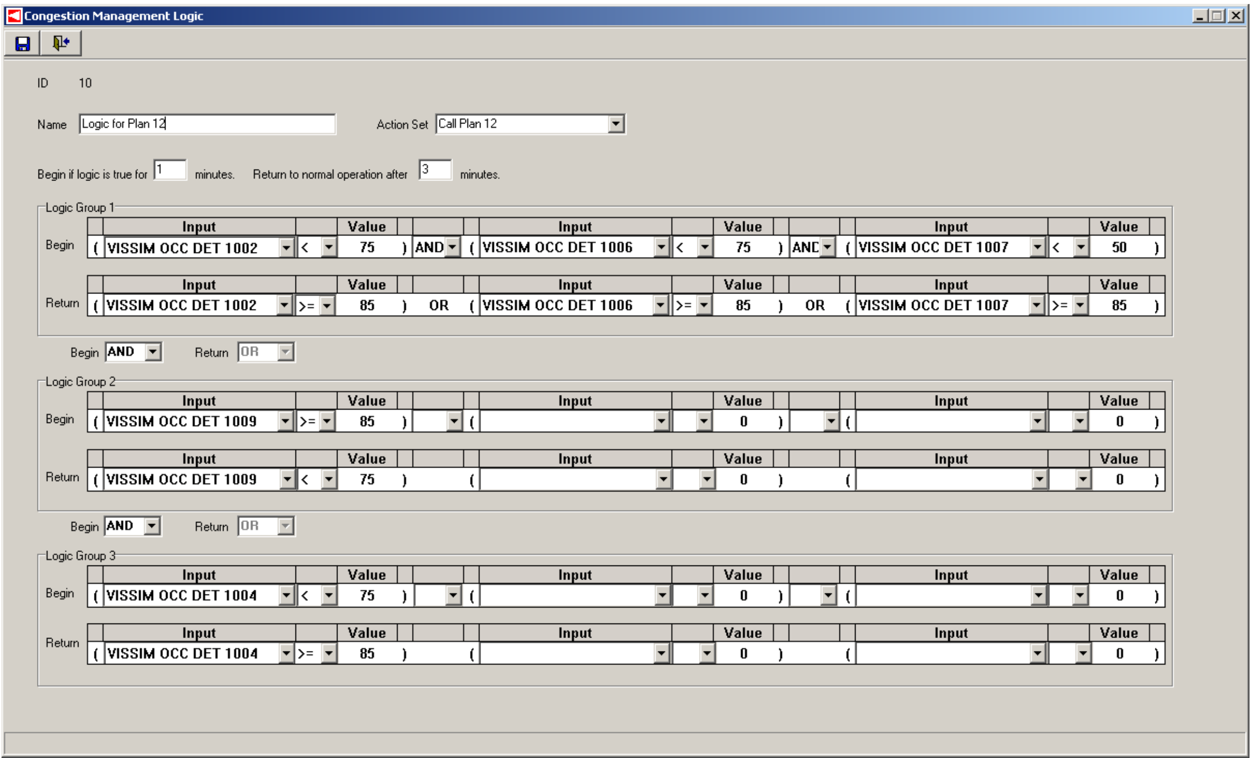


(h)



(i)

Figure 147. ***continued***



(j)

Figure 147. ***continued***

Dynamic Lane Assignment

In this mitigation strategy, the eastbound left turn at Goyeau/Wyandotte was modified to a dual left turn by converting an eastbound through lane to an additional left turn lane. Similarly, the left turn off of Wyandotte into the plaza was also converted to a dual lane movement. This was envisioned to allow additional physical storage for vehicles on the critical route, which might be possible to allow other blocked movements to proceed through the network. The lane assignments were modified by time of day to correspond to the times that the congestion management logic rules would be firing for two queue-responsive strategies. Lane control signs and indications would be necessary to implement such a strategy.

Route/Arterial Strategies

In addition to the three strategies that were applied at just the critical intersection, four other methods were tested. Since this scenario has a common destination for all routes, it was envisioned that metering is perhaps the only way to allow other vehicles not on the critical routes to be able to move through the network.

Westbound and Northbound Metering

For this mitigation strategy, the offsets at the intersections east and south of the critical intersection were modified to create a metering effect at the critical intersection. In addition, the green times on side street movements at these intersections were set to max recall so that they would be serviced regardless of vehicle demand, to provide consistent green time on the main line as well as to attempt to manage queues by minimizing the amount of SOSI > 0 that occurs when the downstream queue has not begun to move, but the upstream light has turned green.

Eastbound Metering

For this mitigation strategy, offsets at the intersections west of the critical intersection were modified to create a metering effect at the critical intersection. The side street movements at these intersections were set to max recall to provide consistent green time on the main arterial as well as to attempt to manage queues by minimizing the amount of SOSI > 0 that occurs when the downstream queue has not begun to move, but the upstream light turned green.

Eastbound, Westbound and Northbound Metering

This mitigation strategy was a combination of the previous two metering strategies.

Re-Routing

To alleviate congestion at the critical intersection and to use storage across a wider area, the original modeled network was expanded to allow re-routing of vehicles. Vehicles that normally enter the border crossing from the west were redirected to enter the plaza from the north. We envisioned that with lane control signals and blank out signs, we could assume that the lane assignments could be changed in a responsive manner after the incident occurs at the tunnel plaza and the queues are detected. By re-routing those vehicles to the southbound approach to the plaza (normally reserved only for vehicles with NEXUS pre-approved v clearance) it was envisioned that this better use of the available network space would “buy time” while the situation on the other side of the river was being resolved (i.e. assigning additional customs agents, capturing the terrorist, etc.).

Performance Analysis

Five runs for each scenario were conducted with varying common random number seeds. A wide array of performance metrics were retrieved from Vissim and post-processed. The results for the analysis of average delay are presented in Table 47 through Table 52 and Figure 148 through Figure 153. Each movement affected by the incident conditions is listed in the table. Each table reports a comparison of the performance of the mitigation strategy for 15 minute periods of the simulation. Finally, Table 53 and Figure 154 illustrate the results for the total average delay for each link for the entire three hour simulation. Because a significant difference in operation was not observed during the ‘loading’ portion of this test case, only the 15 minute periods between 8:00 A.M. and 9:30 A.M. are presented. Each cell of the table is color coded to illustrate the degree of difference between the performance of the baseline scenario and the mitigation strategy. In the Figures and Tables in the remainder of this section, green indicates that the average delay experienced during the mitigation is significantly better (t-test value > 4.4) as compared to the no mitigation scenario. Yellow indicates slightly better (t-test value > 2), white indicated little or no change, orange indicated slightly worse (t-test value < -2), and red indicates significantly worse (t-test value < -4.4) average delay results.

We also present a summary graph indicating the number of links that are represented in each performance category for each mitigation strategy. For each of the figures below, the following number identifiers were used:

1 – Original Mitigation Logic

2 – Expanded Mitigation Logic

3 – Dynamic Lane Assignment

4 – Westbound and Northbound Metering

5 – Eastbound Metering

6 – Westbound, Northbound and Eastbound Metering

7 – Re-routing

Table 47. Average delay per link 8:00 – 8:15

Segment	1	2	3	4	5	6	7
EB TH at PEL-WYAN	86.3	557.6	1714.3	94.2	119.8	635.4	62.3
WB TH at PEL-WYAN	41.2	42.8	36.6	40.2	33.7	38.0	12.8
NB TH at PEL-WYAN	114.4	115.8	384.1	92.4	116.0	199.8	80.3
EB TH at OUEL-WYAN	195.6	567.1	783.6	237.8	259.8	511.6	27.6
SB TH at OUEL-WYAN	100.2	46.2	26.7	48.2	40.1	27.6	16.7
WB TH at OUEL-WYAN	11.0	7.7	14.0	11.4	14.1	28.6	2.5
NB TH at OUEL-WYAN	121.1	129.1	634.6	273.6	157.6	142.5	32.2
NB TH at OUEL-WYAN upstream of LT bay	54.6	11.4	682.3	257.5	85.2	126.8	35.6
NB LT at OUEL-WYAN	60.5	65.8	50.6	57.1	53.1	41.8	20.7
WB TH at GOY-WYAN	18.4	18.1	93.5	12.7	16.5	19.2	3.6
WB TH at GOY-WYAN upstream of LT bay	596.9	1079.6	2038.1	1008.5	859.8	1478.3	64.0
WB LT at GOY-WYAN	470.3	741.6	1181.4	835.9	740.7	1088.5	69.8
SB TH at GOY-WYAN	78.5	3.3	8.6	3.3	3.5	5.4	4.8
SB TH at GOY-WYAN upstream of LT bay	102.5	286.0	750.0	208.4	119.4	789.2	248.2
SB LT at GOY-WYAN	0.0	0.0	0.0	0.0	0.0	0.3	0.0
Tunnel entrance RT lane	294.1	318.1	371.5	234.7	240.7	240.3	13.4
Tunnle entrance upstream from RT bay	21.2	46.5	152.5	8.5	6.0	32.8	1.1
Just past tunnel entrance on GOY	0.5	0.6	259.1	0.6	0.6	0.5	0.1
WB RT at GOY-WYAN	183.8	200.5	292.5	286.6	232.7	301.9	14.3
WB TH at GOY-WYAN	45.3	68.1	100.0	24.5	31.7	92.0	11.8
WB TH at GOY-WYAN upstream of LT bay	124.1	180.9	303.5	262.4	161.8	275.8	21.8
WB LT at GOY-WYAN	37.3	67.9	27.2	28.6	32.6	46.6	10.1
NB TH at GOY-WYAN	3710.5	1107.5	2023.5	1260.1	1125.3	1393.7	142.0
EB TH at WIND-WYAN	2.0	1.4	2.0	2.5	0.9	6.9	0.5
SB TH at WIND-WYAN	82.7	104.1	105.7	117.2	103.8	127.1	46.5
WB RT at WIND-WYAN	0.8	1.7	3.5	2.8	0.9	2.6	0.4
WB TH at WIND-WYAN	105.7	87.7	117.6	155.2	103.1	130.6	7.5
WB TH at WIND-WYAN upstream of RT bay	151.5	213.4	359.5	266.2	194.5	292.9	22.9
NB TH at WIND-WYAN	85.2	99.4	95.7	143.8	162.1	154.6	48.8
EB TH at MCD-WYAN	9.5	8.7	9.2	7.0	9.5	6.3	2.8
SB TH at MCD-WYAN	48.5	51.2	64.2	48.9	50.7	60.0	19.2
SB TH at MCD-WYAN upstream of RT bay	0.1	0.1	0.1	0.1	0.1	0.1	0.0
WB TH at MCD-WYAN	227.8	662.6	2383.7	913.7	910.8	1198.6	153.3
NB TH at MCD-WYAN	30.7	29.5	33.2	29.8	28.9	30.4	9.8
NB TH at MCD-WYAN upstream of LT bay	69.6	59.7	124.9	44.4	47.6	48.0	40.5
NB LT at MCD-WYAN	139.8	104.7	182.9	109.2	84.3	143.8	41.5
EB TH at GOY-TUC	733.6	110.5	385.3	106.1	90.7	81.8	61.3
SB TH at GOY-TUC	12.1	13.0	11.2	13.1	16.2	13.8	4.0
WB TH at GOY-TUC	1767.6	234.4	393.7	248.7	135.6	143.7	113.9
NB TH at GOY-TUC	1693.7	400.9	1149.5	447.5	391.6	630.4	91.8
EB TH at GOY-PARK	60.9	61.1	65.8	61.1	61.1	60.9	19.1
EB TH at GOY-PARK upstream of LT bay	3.4	3.4	3.4	3.5	3.5	3.4	1.3
EB RT at GOY-PARK	82.0	82.1	80.7	80.7	80.7	80.7	28.3
SB TH at GOY-PARK	13.9	14.3	15.7	14.8	14.7	13.9	21.3
WB TH at GOY-PARK	30.4	30.4	30.4	30.3	30.4	30.4	12.1
NB TH at GOY-PARK	7.7	9.2	7.2	10.0	7.4	7.6	2.7
NB TH at OUEL-PARK	114.2	212.4	129.8	134.9	125.3	161.7	71.1
NB RT at OUEL-PARK	125.8	118.5	129.8	119.6	81.2	81.1	29.2
EB RT at OUEL-PARK	85.6	85.8	85.1	85.6	85.9	86.7	30.9
SB TH at OUEL-PARK	125.7	121.8	126.0	128.4	128.1	126.1	42.5
WB RT at OUEL-PARK	32.6	34.1	31.6	32.4	32.5	32.6	12.2
WB TH at OUEL-PARK	44.9	45.6	43.6	45.2	45.0	45.5	11.1
WB LT at OUEL-PARK	33.4	33.5	32.6	33.4	33.4	33.4	11.1
NB TH at OUEL-UNIV	29.0	27.3	33.6	32.5	30.9	23.2	9.5
NB TH at OUEL-UNIV upstream of LT bay	14.8	13.7	13.0	17.0	16.8	14.6	6.0
NB LT at OUEL-UNIV	36.4	32.3	35.1	33.6	36.3	33.4	11.8
EB TH at OUEL-UNIV	25.6	25.8	22.5	23.5	23.3	26.9	8.7
SB TH at OUEL-UNIV	49.6	50.4	58.5	53.2	51.2	45.5	17.6
SB TH at OUEL-UNIV upstream of LT bay	7.9	7.5	58.5	9.2	8.0	6.7	2.9
SB LT at OUEL-UNIV	53.6	58.8	59.4	54.6	64.2	57.8	20.9
EB TH at GOY-UNIV	11.0	12.2	11.9	12.8	14.3	12.5	11.9
SB TH at GOY-UNIV	68.0	66.9	70.7	68.2	67.5	68.8	35.9
WB TH at GOY-UNIV	11.0	11.9	19.6	11.1	11.0	10.8	6.2
NB TH at GOY-UNIV	14.4	13.8	17.0	17.5	16.6	15.5	5.7
To tunnel From WEST	3425.1	5381.2	7341.6	5009.4	4887.4	5982.1	531.2
Through tunnel	6206.3	6171.4	6127.8	6125.3	6132.5	6170.2	854.5
To tunnel from East	2669.5	3797.0	5536.4	3970.8	3703.5	4188.5	348.8
To tunnel from South	5425.6	3697.6	5044.2	3877.1	3869.5	4129.7	421.3

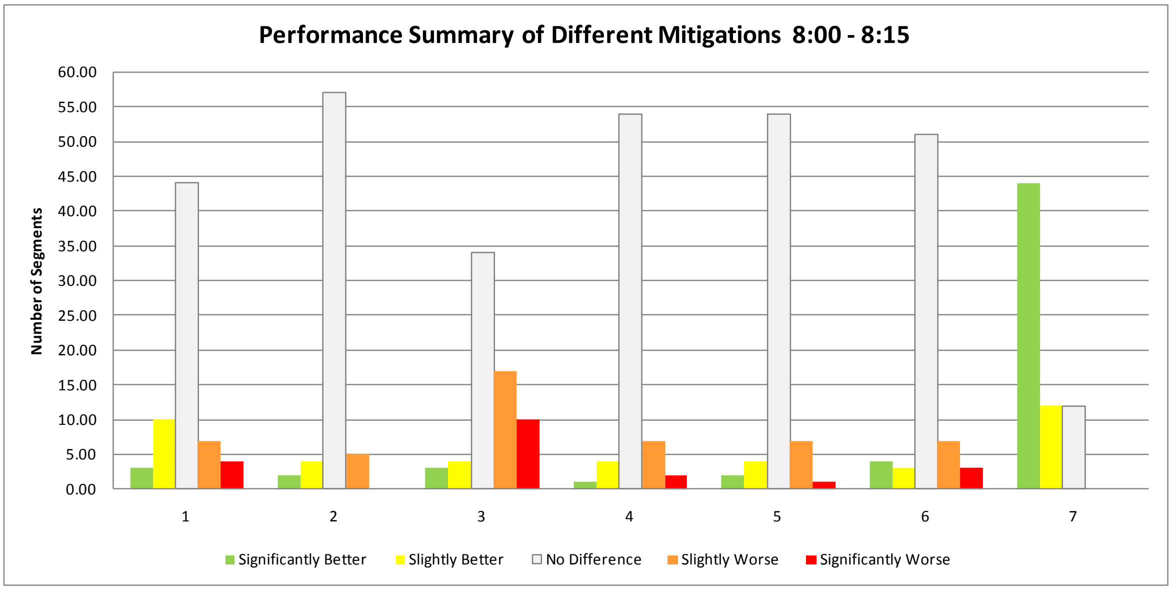


Figure 148. Performance summary 8:00 – 8:15

Table 48. Average delay per link 8:15 – 8:30

Segment	1	2	3	4	5	6	7
EB TH at PEL-WYAN	1192.3	5199.3	2528.8	5513.5	4123.4	4924.9	1417.5
WB TH at PEL-WYAN	36.1	33.6	27.3	29.1	20.1	18.1	32.3
NB TH at PEL-WYAN	207.9	1411.3	1435.1	515.8	1147.3	1602.3	862.1
EB TH at OUEL-WYAN	692.9	2213.3	1536.1	1988.7	1993.7	2052.1	665.5
SB TH at OUEL-WYAN	41.6	38.0	72.6	141.9	25.8	27.2	29.5
WB TH at OUEL-WYAN	18.1	14.9	7.7	25.0	17.2	15.1	9.9
NB TH at OUEL-WYAN	281.7	865.4	833.1	522.9	733.4	909.3	170.5
NB TH at OUEL-WYAN upstrem of LT bay	100.4	914.5	1661.2	802.2	1172.6	997.4	691.2
NB LT at OUEL-WYAN	38.8	34.7	17.6	25.6	33.7	33.2	36.2
WB TH at GOY-WYAN	2.1	1.0	143.9	2.8	15.9	0.9	11.3
WB TH at GOY-WYAN upstream of LT bay	990.9	2756.7	1925.2	2553.4	2156.4	3215.2	630.0
WB LT at GOY-WYAN	413.0	1002.3	1449.0	1155.6	1084.4	1383.4	677.7
SB TH at GOY-WYAN	95.2	8.7	9.1	10.3	7.6	9.9	8.8
SB TH at GOY-WYAN upstream of LT bay	1100.1	1081.4	568.1	984.2	1207.0	898.8	1087.0
SB LT at GOY-WYAN	318.8	5.6	0.0	0.0	0.7	0.0	0.0
Tunnel entrance RT lane	243.0	315.3	558.8	411.1	289.0	332.3	302.2
Tunnle entrance upstream from RT bay	26.7	8.4	124.0	7.3	45.4	36.8	5.8
Just past tunnel entrance on GOY	0.7	0.5	201.5	0.6	0.9	0.5	0.5
WB RT at GOY-WYAN	155.8	447.3	270.2	499.1	586.1	762.9	301.9
WB TH at GOY-WYAN	16.0	48.9	30.1	96.6	38.4	42.1	94.6
WB TH at GOY-WYAN upstream of LT bay	160.6	710.1	453.5	540.9	810.0	653.3	413.2
WB LT at GOY-WYAN	13.4	30.3	30.1	45.1	22.5	11.9	76.3
NB TH at GOY-WYAN	808.8	1637.6	1785.9	2035.7	2101.0	1723.9	1664.9
EB TH at WIND-WYAN	3.4	5.5	0.4	3.4	2.1	2.2	2.2
SB TH at WIND-WYAN	104.4	609.5	463.9	500.7	382.9	652.3	458.1
WB RT at WIND-WYAN	3.0	2.0	0.3	2.7	1.8	2.0	0.4
WB TH at WIND-WYAN	74.8	408.2	248.9	339.9	429.0	425.6	172.4
WB TH at WIND-WYAN upstream of RT bay	252.7	1256.1	652.8	1121.0	1197.1	816.2	490.1
NB TH at WIND-WYAN	102.4	929.2	425.6	771.3	1333.1	520.5	560.7
EB TH at MCD-WYAN	9.0	8.4	8.8	0.6	8.5	0.4	7.9
SB TH at MCD-WYAN	51.1	119.8	96.9	81.4	71.8	89.9	119.6
SB TH at MCD-WYAN upstream of RT bay	0.1	0.2	0.2	0.2	0.2	0.2	0.4
WB TH at MCD-WYAN	922.1	4042.9	3479.5	3533.9	3804.2	4136.2	3087.7
NB TH at MCD-WYAN	30.3	23.1	22.6	21.0	29.3	19.5	26.8
NB TH at MCD-WYAN upstream of LT bay	56.2	639.7	1099.5	748.6	858.5	619.5	899.1
NB LT at MCD-WYAN	124.4	568.0	656.3	609.3	501.9	479.1	587.6
EB TH at GOY-TUC	2577.3	475.3	2215.5	454.0	456.2	522.0	602.2
SB TH at GOY-TUC	19.7	12.2	18.2	18.2	15.1	23.5	14.2
WB TH at GOY-TUC	3095.3	1510.2	1586.0	1186.4	1029.3	948.1	1261.7
NB TH at GOY-TUC	1567.0	1682.9	2262.9	2340.9	2627.1	1881.6	1803.1
EB TH at GOY-PARK	49.0	52.1	62.3	50.6	51.6	53.0	54.4
EB TH at GOY-PARK upstream of LT bay	82.9	3.9	3.0	72.8	15.3	3.0	3.2
EB RT at GOY-PARK	124.2	96.7	108.1	165.3	107.7	81.9	91.8
SB TH at GOY-PARK	84.1	162.0	183.1	131.5	119.7	336.4	240.2
WB TH at GOY-PARK	61.6	32.5	38.8	32.7	32.6	37.6	38.1
NB TH at GOY-PARK	19.4	9.2	5.4	5.4	10.6	6.5	6.7
NB TH at OUEL-PARK	336.3	134.7	143.6	118.7	127.5	157.2	256.2
NB RT at OUEL-PARK	86.3	66.2	66.6	60.9	92.2	90.0	88.5
EB RT at OUEL-PARK	94.4	93.7	90.1	92.6	93.5	92.1	93.2
SB TH at OUEL-PARK	119.0	122.7	120.0	120.9	119.8	127.7	122.3
WB RT at OUEL-PARK	39.9	38.1	39.9	39.4	39.6	38.9	39.5
WB TH at OUEL-PARK	33.1	32.0	30.8	32.0	31.9	33.2	31.7
WB LT at OUEL-PARK	31.7	30.9	31.9	31.4	31.8	30.9	31.7
NB TH at OUEL-UNIV	19.7	32.3	47.2	33.6	32.0	47.4	34.7
NB TH at OUEL-UNIV upstream of LT bay	11.5	11.0	15.6	13.1	17.2	14.6	14.1
NB LT at OUEL-UNIV	22.4	51.4	51.3	50.5	41.2	46.9	37.6
EB TH at OUEL-UNIV	31.6	25.2	24.3	25.1	24.6	26.8	27.0
SB TH at OUEL-UNIV	49.6	59.0	61.3	57.1	57.3	59.5	59.8
SB TH at OUEL-UNIV upstream of LT bay	8.6	12.1	61.3	12.1	11.5	11.2	11.9
SB LT at OUEL-UNIV	47.3	57.2	52.1	54.7	53.9	67.4	54.1
EB TH at GOY-UNIV	44.3	24.9	89.6	40.2	10.2	123.2	31.8
SB TH at GOY-UNIV	141.2	147.5	314.7	104.2	138.8	280.5	147.3
WB TH at GOY-UNIV	77.5	87.1	116.0	65.1	33.9	127.2	43.6
NB TH at GOY-UNIV	17.0	16.9	20.3	15.7	16.1	11.5	13.6
To tunnel From WEST	3871.0	5692.8	4588.2	6193.8	7176.3	5751.1	3898.8
Through tunnel	4570.9	4522.4	4448.5	4581.6	4690.5	4546.4	4543.6
To tunnel from East	2951.4	5065.2	3478.4	4294.2	4079.5	3245.9	4229.3
To tunnel from South	1949.7	3888.0	4243.8	4823.9	4660.9	4274.8	4434.4

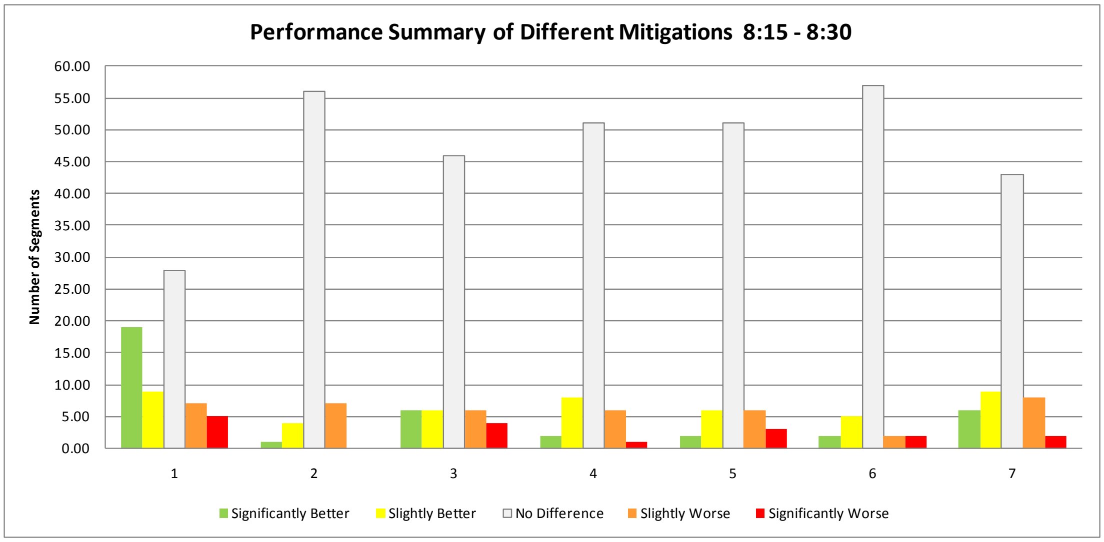


Figure 149. Performance summary 8:15 – 8:30

Table 49. Average delay per link 8:30 – 8:45

Segment	1	2	3	4	5	6	7
EB TH at PEL-WYAN	802.7	4041.5	2418.1	3739.9	3621.0	2969.6	488.6
WB TH at PEL-WYAN	34.8	25.2	22.3	23.3	13.9	17.7	32.4
NB TH at PEL-WYAN	171.1	1346.5	2538.7	1008.8	2067.9	1720.2	585.6
EB TH at OUEL-WYAN	352.3	2241.2	772.0	1815.5	1810.7	1607.7	432.1
SB TH at OUEL-WYAN	31.3	17.1	187.2	102.9	20.7	18.0	50.9
WB TH at OUEL-WYAN	16.4	14.3	16.4	20.5	26.2	20.7	12.1
NB TH at OUEL-WYAN	187.7	761.9	657.1	1235.3	1037.0	1093.9	250.8
NB TH at OUEL-WYAN upstrem of LT bay	72.1	2254.1	1773.6	2281.9	1904.5	2198.4	565.5
NB LT at OUEL-WYAN	38.1	27.7	17.8	26.7	24.9	16.9	52.5
WB TH at GOY-WYAN	0.6	7.3	46.3	1.5	3.5	7.0	16.7
WB TH at GOY-WYAN upstream of LT bay	570.5	2594.0	963.0	1947.9	1776.6	2093.6	473.3
WB LT at GOY-WYAN	238.7	967.0	692.3	810.2	874.6	867.4	549.0
SB TH at GOY-WYAN	93.3	9.2	4.3	9.7	10.1	11.6	15.8
SB TH at GOY-WYAN upstream of LT bay	809.0	658.0	789.8	666.6	558.5	773.4	1184.9
SB LT at GOY-WYAN	311.1	4.0	6.5	6.5	6.8	8.8	0.0
Tunnel entrance RT lane	150.1	164.5	269.8	169.0	232.5	129.9	240.2
Tunnle entrance upstream from RT bay	3.8	10.2	45.7	8.5	25.5	16.9	7.0
Just past tunnel entrance on GOY	0.5	0.7	59.6	0.6	0.6	0.8	0.5
WB RT at GOY-WYAN	55.8	241.5	161.0	228.0	278.5	285.3	274.9
WB TH at GOY-WYAN	25.7	51.8	92.1	48.0	43.6	33.8	60.3
WB TH at GOY-WYAN upstream of LT bay	77.8	367.7	303.2	334.2	484.5	417.5	544.2
WB LT at GOY-WYAN	15.9	38.8	36.1	50.8	22.8	38.9	50.6
NB TH at GOY-WYAN	197.2	1297.2	1145.3	1146.9	1296.6	1112.4	1711.1
EB TH at WIND-WYAN	1.8	5.6	2.0	1.4	2.4	3.0	0.6
SB TH at WIND-WYAN	104.4	246.5	141.0	312.1	468.8	284.7	142.9
WB RT at WIND-WYAN	0.6	0.5	1.0	2.0	3.6	2.4	1.5
WB TH at WIND-WYAN	30.4	188.9	117.8	123.6	250.3	267.5	185.6
WB TH at WIND-WYAN upstream of RT bay	85.1	496.2	378.5	452.6	566.1	564.0	588.2
NB TH at WIND-WYAN	61.5	651.7	382.6	414.8	752.2	417.1	212.3
EB TH at MCD-WYAN	3.2	5.5	8.1	0.1	5.0	0.2	7.8
SB TH at MCD-WYAN	48.5	127.0	90.6	49.3	94.4	45.3	58.0
SB TH at MCD-WYAN upstream of RT bay	0.0	0.2	0.1	0.08	0.1	0.1	0.1
WB TH at MCD-WYAN	368.5	2390.3	1839.5	2204.7	2554.6	2573.1	3075.2
NB TH at MCD-WYAN	35.7	30.9	16.8	17.4	27.4	21.5	29.7
NB TH at MCD-WYAN upstream of LT bay	33.1	892.5	874.1	494.1	431.5	394.3	395.5
NB LT at MCD-WYAN	87.0	417.8	302.5	286.5	448.1	520.0	265.8
EB TH at GOY-TUC	417.5	263.0	1290.0	335.9	336.4	202.0	840.4
SB TH at GOY-TUC	32.3	19.3	14.9	21.0	17.8	29.0	13.5
WB TH at GOY-TUC	470.5	617.5	1066.0	1138.3	919.2	850.9	1549.8
NB TH at GOY-TUC	454.1	1237.6	1120.6	1125.8	1269.2	1160.6	1799.8
EB TH at GOY-PARK	63.9	56.6	56.7	55.2	51.3	52.5	54.3
EB TH at GOY-PARK upstream of LT bay	23.5	3.1	6.2	3.1	9.8	2.4	2.4
EB RT at GOY-PARK	73.4	82.4	86.7	80.5	110.0	105.6	100.1
SB TH at GOY-PARK	50.7	67.6	124.9	74.5	165.3	258.0	295.3
WB TH at GOY-PARK	5.58	5.6	5.6	5.6	5.6	5.3	2.5
NB TH at GOY-PARK	8.1	4.8	6.3	10.6	7.9	8.4	11.3
NB TH at OUEL-PARK	264.5	150.2	173.2	129.4	122.7	160.2	450.3
NB RT at OUEL-PARK	86.2	89.8	78.1	51.3	87.8	57.5	84.4
EB RT at OUEL-PARK	98.2	99.4	98.8	100.1	98.2	96.9	99.4
SB TH at OUEL-PARK	121.8	125.2	126.9	125.4	126.1	128.8	121.4
WB RT at OUEL-PARK	22.0	20.7	22.8	20.8	20.5	20.5	21.7
WB TH at OUEL-PARK	39.1	36.8	35.0	34.6	33.9	34.5	38.7
WB LT at OUEL-PARK	26.4	24.3	25.1	24.3	23.4	23.5	22.3
NB TH at OUEL-UNIV	24.9	42.9	51.9	46.3	42.2	47.2	21.2
NB TH at OUEL-UNIV upstream of LT bay	20.2	12.7	13.0	15.6	11.8	22.8	14.9
NB LT at OUEL-UNIV	17.0	47.6	43.9	51.7	49.6	49.3	32.5
EB TH at OUEL-UNIV	22.4	15.8	15.8	14.8	15.6	26.0	25.1
SB TH at OUEL-UNIV	47.8	55.9	61.5	58.8	53.6	55.2	49.7
SB TH at OUEL-UNIV upstream of LT bay	6.0	5.7	61.5	6.7	7.0	7.3	6.1
SB LT at OUEL-UNIV	55.9	69.8	78.7	77.2	60.7	81.2	68.5
EB TH at GOY-UNIV	41.3	16.3	84.3	43.2	10.6	52.1	117.9
SB TH at GOY-UNIV	130.5	100.7	102.0	185.7	218.9	288.8	312.0
WB TH at GOY-UNIV	47.2	46.5	76.0	37.8	21.4	65.8	35.8
NB TH at GOY-UNIV	17.0	16.8	15.3	13.6	17.7	58.6	11.2
To tunnel From WEST	2763.3	2012.9	4066.2	3673.6	4076.2	3143.5	2189.4
Through tunnel	3187.6	3221.9	3080.7	3136.4	3112.8	3119.6	3085.8
To tunnel from East	1750.8	1860.9	1000.5	1929.4	2707.0	1417.0	3072.1
To tunnel from South	577.4	2623.1	2685.4	2791.7	2755.7	2655.9	3614.7

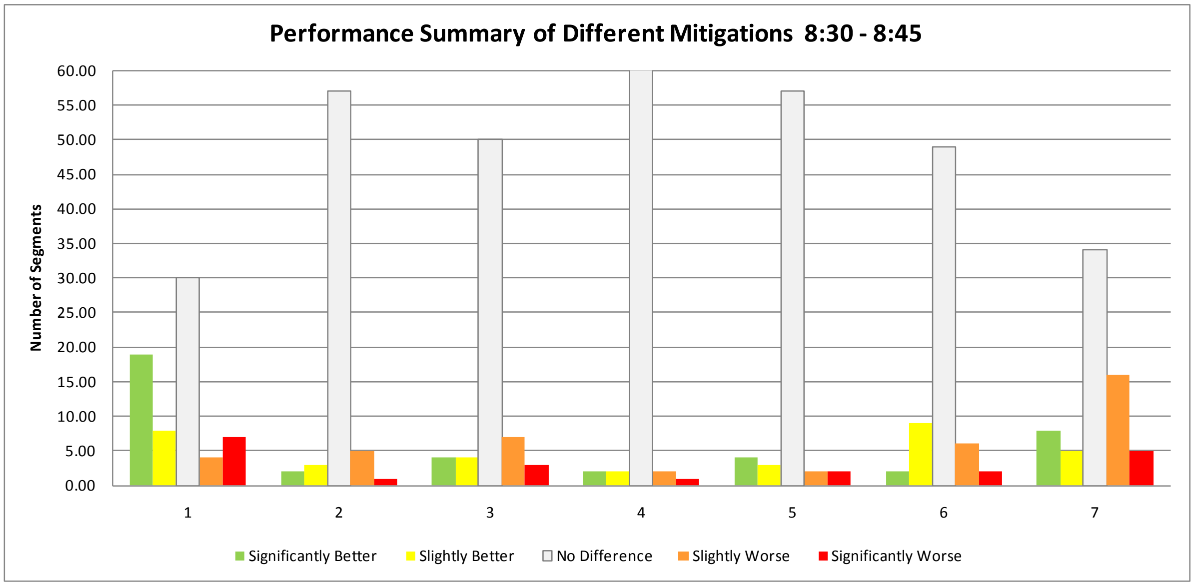


Figure 150. Performance summary 8:30 – 8:45

Table 50. Average delay per link 8:45 – 9:00

Segment	1	2	3	4	5	6	7
EB TH at PEL-WYAN	457.5	2375.9	2035.2	2647.2	2398.7	2408.4	390.1
WB TH at PEL-WYAN	42.8	26.0	34.7	20.4	22.9	25.4	37.9
NB TH at PEL-WYAN	147.4	1209.0	1474.9	983.1	1486.4	1585.8	496.3
EB TH at OUEL-WYAN	373.9	1209.7	917.5	1475.0	1219.7	1126.9	328.4
SB TH at OUEL-WYAN	37.0	20.4	18.8	21.7	49.6	23.7	45.0
WB TH at OUEL-WYAN	25.3	26.6	19.1	18.4	29.1	18.4	16.8
NB TH at OUEL-WYAN	184.4	703.4	409.7	932.6	642.8	841.2	166.4
NB TH at OUEL-WYAN upstrem of LT bay	18.6	1909.9	1117.3	2127.7	1853.7	2301.2	267.6
NB LT at OUEL-WYAN	48.2	16.9	46.5	29.4	30.3	13.9	49.0
WB TH at GOY-WYAN	26.6	16.3	166.6	21.4	1.7	4.0	4.1
WB TH at GOY-WYAN upstream of LT bay	562.9	1213.5	1099.6	1573.6	1296.9	1360.4	990.2
WB LT at GOY-WYAN	325.2	598.5	863.9	693.3	566.5	559.0	868.0
SB TH at GOY-WYAN	5.3	7.2	8.4	5.2	2.0	10.7	9.7
SB TH at GOY-WYAN upstream of LT bay	131.8	145.9	253.0	245.7	193.2	276.5	583.9
SB LT at GOY-WYAN	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Tunnel entrance RT lane	77.0	82.4	217.3	97.9	75.4	92.9	121.4
Tunnle entrance upstream from RT bay	4.3	5.4	54.7	8.0	8.8	9.8	9.6
Just past tunnel entrance on GOY	0.5	0.6	64.1	0.6	0.6	0.6	0.6
WB RT at GOY-WYAN	87.1	137.2	99.5	195.6	148.9	122.4	142.4
WB TH at GOY-WYAN	64.0	116.4	49.8	33.8	48.6	53.7	69.1
WB TH at GOY-WYAN upstream of LT bay	50.8	218.9	154.6	224.3	191.7	147.9	213.4
WB LT at GOY-WYAN	45.4	56.6	39.4	28.2	49.9	34.7	47.7
NB TH at GOY-WYAN	751.2	685.3	842.5	670.6	671.4	692.3	930.9
EB TH at WIND-WYAN	3.9	2.7	1.5	2.8	2.0	3.6	1.4
SB TH at WIND-WYAN	109.5	96.7	95.2	67.2	96.8	64.9	95.4
WB RT at WIND-WYAN	1.0	2.0	4.3	1.1	2.3	2.3	1.1
WB TH at WIND-WYAN	10.8	117.5	56.5	104.2	70.0	60.5	78.6
WB TH at WIND-WYAN upstream of RT bay	34.5	288.2	193.3	304.7	249.6	189.7	266.5
NB TH at WIND-WYAN	87.6	118.5	83.8	177.6	98.6	58.3	95.5
EB TH at MCD-WYAN	4.0	8.5	10.2	0.3	7.5	0.3	7.6
SB TH at MCD-WYAN	49.9	48.0	47.4	32.4	42.5	34.9	47.8
SB TH at MCD-WYAN upstream of RT bay	0.1	0.1	0.1	0.1	0.1	0.1	0.1
WB TH at MCD-WYAN	47.7	1419.7	1048.8	1561.9	1500.7	1288.8	1596.9
NB TH at MCD-WYAN	35.9	27.4	30.5	28.1	30.4	30.2	31.6
NB TH at MCD-WYAN upstream of LT bay	45.8	194.0	145.5	103.8	141.8	70.4	84.1
NB LT at MCD-WYAN	92.8	165.3	79.9	143.9	107.2	61.3	112.7
EB TH at GOY-TUC	736.4	81.1	855.2	63.7	76.8	61.7	169.5
SB TH at GOY-TUC	26.2	24.0	17.9	18.7	16.9	27.2	17.5
WB TH at GOY-TUC	808.0	271.5	743.2	390.5	263.3	359.7	560.9
NB TH at GOY-TUC	865.9	630.2	791.4	569.3	529.7	571.0	846.2
EB TH at GOY-PARK	54.8	54.4	61.0	54.3	54.3	54.3	50.4
EB TH at GOY-PARK upstream of LT bay	2.9	2.9	2.8	2.9	2.9	2.9	7.8
EB RT at GOY-PARK	67.8	69.5	69.6	71.7	69.5	69.5	78.8
SB TH at GOY-PARK	15.1	14.2	14.0	16.4	18.3	27.2	97.5
WB TH at GOY-PARK	11.6	11.6	11.6	11.6	11.6	11.6	11.6
NB TH at GOY-PARK	8.3	8.4	10.4	9.1	9.2	10.4	10.8
NB TH at OUEL-PARK	162.3	167.7	285.7	159.4	157.8	217.4	361.8
NB RT at OUEL-PARK	91.1	82.0	122.4	106.7	83.3	98.2	99.8
EB RT at OUEL-PARK	91.5	91.2	89.7	92.0	95.9	89.6	92.8
SB TH at OUEL-PARK	120.9	122.7	120.4	120.1	118.2	124.3	121.2
WB RT at OUEL-PARK	31.0	29.5	31.2	29.4	29.7	31.4	31.7
WB TH at OUEL-PARK	40.5	38.9	41.2	39.7	39.4	39.6	40.3
WB LT at OUEL-PARK	34.4	33.1	35.3	33.6	34.1	34.3	34.7
NB TH at OUEL-UNIV	32.9	39.9	23.0	39.5	35.6	33.6	31.1
NB TH at OUEL-UNIV upstream of LT bay	12.8	12.4	12.5	13.1	15.1	13.9	43.9
NB LT at OUEL-UNIV	33.5	37.6	23.7	36.2	32.8	39.5	35.7
EB TH at OUEL-UNIV	18.8	18.9	21.9	17.8	18.7	19.7	28.8
SB TH at OUEL-UNIV	58.3	59.5	54.3	56.4	57.5	55.5	55.2
SB TH at OUEL-UNIV upstream of LT bay	5.0	9.1	54.3	6.6	5.7	5.5	6.4
SB LT at OUEL-UNIV	68.4	80.3	76.6	57.4	64.5	80.8	60.7
EB TH at GOY-UNIV	6.2	6.9	6.3	6.5	6.9	7.7	83.2
SB TH at GOY-UNIV	70.2	69.8	73.8	70.8	72.9	84.7	220.8
WB TH at GOY-UNIV	9.3	9.5	15.8	9.7	10.0	11.8	21.7
NB TH at GOY-UNIV	22.9	26.2	20.4	22.7	24.6	22.8	18.0
To tunnel From WEST	2106.1	2495.1	2554.4	2570.4	3476.6	3593.5	2533.4
Through tunnel	2762.0	2744.5	2746.1	2787.8	2706.4	2829.7	2777.7
To tunnel from East	926.4	2280.9	1839.2	1882.3	2472.8	2174.5	2165.1
To tunnel from South	1740.1	1741.0	2121.9	1709.2	1845.2	1877.7	2325.9

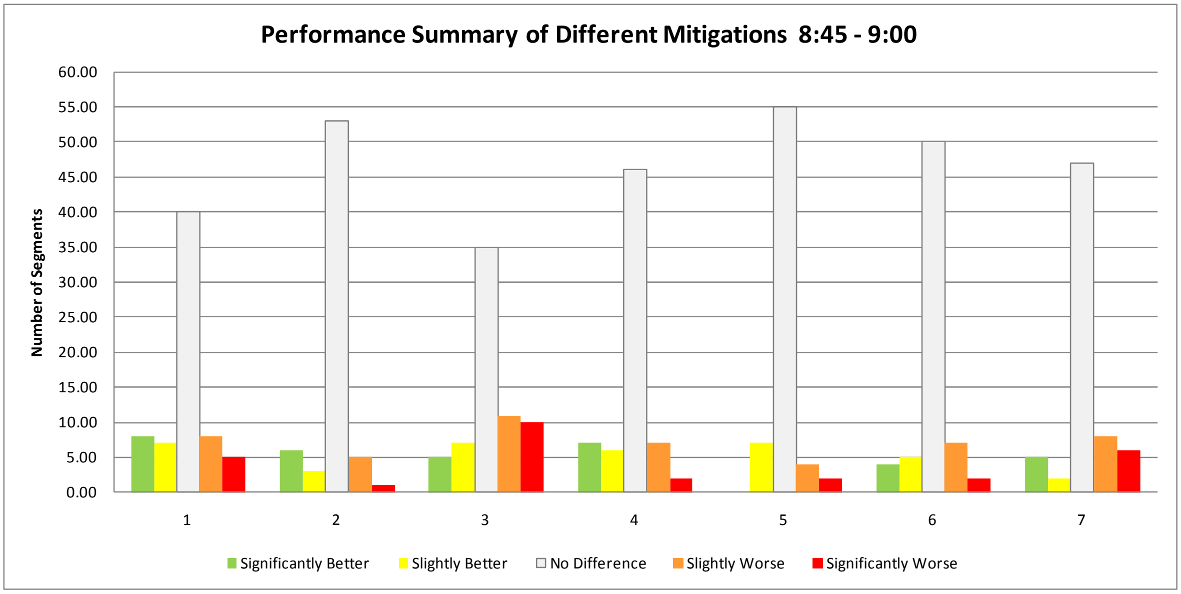


Figure 151. Performance summary 8:45 – 9:00

Table 51. Average delay per link 9:00 – 9:15

Segment	1	2	3	4	5	6	7
EB TH at PEL-WYAN	91.0	551.9	784.8	639.1	704.9	723.5	402.5
WB TH at PEL-WYAN	14.5	32.9	33.1	27.5	27.0	24.4	15.6
NB TH at PEL-WYAN	87.7	375.8	481.0	297.4	695.5	590.5	114.3
EB TH at OUEL-WYAN	78.4	339.2	512.6	377.1	460.8	413.4	297.7
SB TH at OUEL-WYAN	30.9	24.9	18.1	18.7	13.8	13.8	25.6
WB TH at OUEL-WYAN	13.8	22.1	16.6	9.0	19.4	13.7	11.7
NB TH at OUEL-WYAN	63.0	236.4	419.5	319.6	248.2	249.7	184.0
NB TH at OUEL-WYAN upstrem of LT bay	0.4	337.4	670.7	576.0	452.4	559.5	138.7
NB LT at OUEL-WYAN	51.2	63.9	50.9	33.7	30.5	38.0	35.4
WB TH at GOY-WYAN	15.8	6.9	46.9	4.4	4.0	1.4	3.2
WB TH at GOY-WYAN upstream of LT bay	145.2	327.6	651.6	407.8	522.2	463.5	573.8
WB LT at GOY-WYAN	153.8	190.2	372.0	186.3	261.5	241.9	372.4
SB TH at GOY-WYAN	0.0	0.2	0.5	0.3	1.3	1.8	1.4
SB TH at GOY-WYAN upstream of LT bay	0.0	0.0	0.0	0.0	2.1	3.5	0.0
SB LT at GOY-WYAN	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Tunnel entrance RT lane	31.2	37.1	70.9	34.2	40.3	37.8	58.0
Tunnle entrance upstream from RT bay	3.8	2.8	11.1	3.8	5.8	3.2	4.0
Just past tunnel entrance on GOY	0.5	0.5	13.4	0.4	0.5	0.5	0.5
WB RT at GOY-WYAN	21.0	54.5	32.7	45.6	58.9	50.2	77.5
WB TH at GOY-WYAN	32.4	42.5	27.7	18.7	33.5	19.7	42.8
WB TH at GOY-WYAN upstream of LT bay	5.0	65.8	47.4	40.2	42.1	40.7	112.9
WB LT at GOY-WYAN	26.2	44.0	40.9	17.2	27.1	28.9	30.4
NB TH at GOY-WYAN	159.8	209.5	324.3	208.7	301.5	238.3	356.9
EB TH at WIND-WYAN	1.1	0.9	1.3	1.7	0.8	2.3	0.8
SB TH at WIND-WYAN	89.8	96.8	89.6	62.2	96.9	62.1	89.6
WB RT at WIND-WYAN	0.5	0.6	0.7	5.1	3.1	5.5	0.4
WB TH at WIND-WYAN	3.0	20.4	14.3	36.1	19.9	29.6	52.1
WB TH at WIND-WYAN upstream of RT bay	0.6	58.6	35.4	83.5	37.8	57.1	126.1
NB TH at WIND-WYAN	91.3	77.9	99.4	70.8	77.8	58.9	100.9
EB TH at MCD-WYAN	1.0	1.8	1.6	0.2	2.0	0.2	1.0
SB TH at MCD-WYAN	54.6	52.4	56.6	34.7	58.5	34.9	51.5
SB TH at MCD-WYAN upstream of RT bay	0.0	0.1	0.0	0.0	0.0	0.0	0.0
WB TH at MCD-WYAN	7.8	275.8	165.8	362.6	197.5	357.1	420.2
NB TH at MCD-WYAN	54.7	44.7	58.2	34.0	42.3	34.7	44.0
NB TH at MCD-WYAN upstream of LT bay	9.7	14.3	11.6	7.3	14.9	5.2	12.3
NB LT at MCD-WYAN	63.1	44.7	64.2	51.9	40.7	64.9	53.1
EB TH at GOY-TUC	93.5	73.1	120.6	51.6	72.7	52.0	71.7
SB TH at GOY-TUC	20.6	12.7	11.9	17.3	12.7	19.1	14.1
WB TH at GOY-TUC	91.2	82.5	110.7	55.4	73.9	41.1	152.2
NB TH at GOY-TUC	66.1	45.7	106.6	63.5	58.0	65.1	179.9
EB TH at GOY-PARK	49.5	51.4	74.4	51.4	51.4	51.4	51.4
EB TH at GOY-PARK upstream of LT bay	2.2	2.1	2.1	2.1	2.1	2.1	2.1
EB RT at GOY-PARK	63.4	62.9	62.9	62.9	62.9	62.9	62.9
SB TH at GOY-PARK	8.9	7.4	8.9	7.6	7.7	6.9	8.4
WB TH at GOY-PARK	3.8	3.8	3.8	3.8	3.8	3.8	3.8
NB TH at GOY-PARK	4.2	4.3	4.7	4.9	5.4	6.3	4.6
NB TH at OUEL-PARK	100.2	330.4	225.0	212.2	216.0	258.5	187.3
NB RT at OUEL-PARK	51.8	82.9	72.3	84.2	76.3	82.8	68.7
EB RT at OUEL-PARK	114.0	112.1	113.7	110.0	112.3	118.6	113.7
SB TH at OUEL-PARK	120.6	112.6	113.5	116.4	115.7	120.2	119.2
WB RT at OUEL-PARK	20.6	24.5	24.8	25.4	24.8	27.4	23.0
WB TH at OUEL-PARK	4.2	5.4	6.7	6.1	6.8	5.4	6.4
WB LT at OUEL-PARK	24.6	31.2	28.9	29.7	30.5	29.1	28.2
NB TH at OUEL-UNIV	36.7	11.6	18.8	21.1	15.3	18.3	23.2
NB TH at OUEL-UNIV upstream of LT bay	6.2	10.1	9.6	9.1	10.5	10.7	8.3
NB LT at OUEL-UNIV	24.0	20.2	14.0	15.0	14.6	16.1	15.9
EB TH at OUEL-UNIV	13.2	16.9	15.4	16.8	16.1	19.5	14.4
SB TH at OUEL-UNIV	71.1	52.3	72.1	62.7	54.6	54.4	66.4
SB TH at OUEL-UNIV upstream of LT bay	2.4	2.1	72.1	2.4	2.0	2.1	2.4
SB LT at OUEL-UNIV	40.2	36.2	42.3	41.7	26.9	36.8	41.0
EB TH at GOY-UNIV	3.4	2.2	1.2	2.0	2.5	3.4	3.1
SB TH at GOY-UNIV	67.6	67.5	67.1	67.6	67.6	67.8	66.3
WB TH at GOY-UNIV	6.9	7.0	13.2	7.0	7.2	7.1	7.1
NB TH at GOY-UNIV	24.3	32.3	32.4	32.3	32.3	32.4	32.3
To tunnel From WEST	578.5	862.7	999.1	998.7	1474.0	1254.6	1253.4
Through tunnel	2321.5	2428.4	2201.5	2367.1	2488.4	2406.5	2395.1
To tunnel from East	292.9	611.1	549.7	322.8	439.8	676.0	801.6
To tunnel from South	463.2	593.2	802.0	556.6	682.8	604.2	806.8

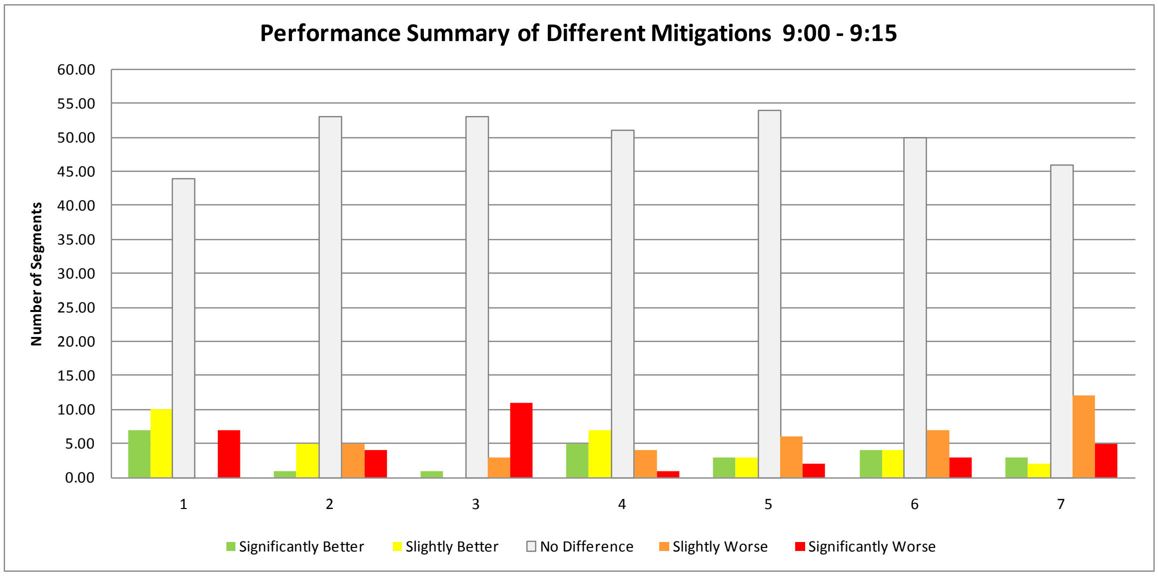


Figure 152. Performance summary 9:00 – 9:15

Table 52. Average delay per link 9:15 – 9:30

Segment	1	2	3	4	5	6	7
EB TH at PEL-WYAN	14.6	13.6	10.8	13.8	37.0	28.8	13.6
WB TH at PEL-WYAN	9.4	20.1	13.3	13.9	14.6	7.8	17.3
NB TH at PEL-WYAN	57.2	63.2	157.9	65.7	71.4	55.6	60.1
EB TH at OUEL-WYAN	10.5	9.8	13.8	12.3	49.9	18.7	12.3
SB TH at OUEL-WYAN	29.6	31.5	22.0	30.1	10.8	28.8	27.9
WB TH at OUEL-WYAN	5.1	1.5	6.3	3.3	22.1	0.8	5.9
NB TH at OUEL-WYAN	62.9	55.0	47.9	67.2	31.6	68.3	69.6
NB TH at OUEL-WYAN upstrem of LT bay	0.2	0.2	0.1	0.2	0.1	0.1	0.2
NB LT at OUEL-WYAN	51.3	61.3	37.0	35.4	25.1	42.7	44.8
WB TH at GOY-WYAN	5.4	2.2	58.6	5.1	5.5	5.8	3.4
WB TH at GOY-WYAN upstream of LT bay	0.4	0.3	13.3	12.0	23.8	13.6	10.7
WB LT at GOY-WYAN	90.1	96.7	66.4	90.5	55.2	70.2	94.0
SB TH at GOY-WYAN	0.0	0.7	0.0	0.0	0.0	0.0	0.0
SB TH at GOY-WYAN upstream of LT bay	0.0	1.1	0.0	0.0	0.0	0.0	0.0
SB LT at GOY-WYAN	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Tunnel entrance RT lane	3.3	0.6	8.3	1.6	4.4	1.2	1.2
Tunnle entrance upstream from RT bay	1.2	0.2	1.1	0.3	0.3	0.4	0.5
Just past tunnel entrance on GOY	0.1	0.0	0.8	0.1	0.1	0.1	0.1
WB RT at GOY-WYAN	2.6	2.4	2.2	2.4	9.7	4.0	2.8
WB TH at GOY-WYAN	7.1	8.2	10.4	1.8	5.8	2.6	4.5
WB TH at GOY-WYAN upstream of LT bay	0.5	0.4	0.7	0.5	0.5	0.7	0.7
WB LT at GOY-WYAN	7.6	8.4	5.2	5.7	6.6	9.2	3.8
NB TH at GOY-WYAN	95.3	104.1	90.2	86.5	91.8	88.0	99.4
EB TH at WIND-WYAN	1.4	1.1	1.4	3.6	2.5	9.7	1.2
SB TH at WIND-WYAN	78.6	78.6	78.6	56.6	78.7	56.6	78.6
WB RT at WIND-WYAN	1.4	0.7	2.6	0.6	0.7	2.3	1.3
WB TH at WIND-WYAN	3.5	3.3	2.9	28.0	4.4	27.3	3.5
WB TH at WIND-WYAN upstream of RT bay	0.5	0.5	0.5	3.8	0.7	4.0	0.6
NB TH at WIND-WYAN	68.6	68.6	68.6	45.5	68.6	45.5	68.6
EB TH at MCD-WYAN	0.2	0.2	0.2	0.2	0.2	0.2	0.3
SB TH at MCD-WYAN	48.4	48.7	44.0	30.9	48.3	28.8	46.5
SB TH at MCD-WYAN upstream of RT bay	0.0	0.0	0.0	0.0	0.0	0.0	0.0
WB TH at MCD-WYAN	8.1	8.2	7.8	16.4	8.1	16.5	8.3
NB TH at MCD-WYAN	49.9	47.2	56.7	21.5	52.7	23.1	46.3
NB TH at MCD-WYAN upstream of LT bay	13.6	10.8	11.1	6.9	11.8	6.9	10.1
NB LT at MCD-WYAN	45.6	47.2	65.5	18.5	55.3	12.4	51.8
EB TH at GOY-TUC	62.6	85.2	77.1	30.9	84.8	30.2	87.1
SB TH at GOY-TUC	6.0	5.7	3.2	8.9	7.7	7.4	4.7
WB TH at GOY-TUC	45.7	55.6	57.0	46.1	55.6	45.0	47.8
NB TH at GOY-TUC	3.9	4.0	3.5	19.3	3.9	19.3	3.8
EB TH at GOY-PARK	69.3	57.7	60.5	57.7	57.7	57.7	57.7
EB TH at GOY-PARK upstream of LT bay	1.7	1.6	1.6	1.6	1.6	1.6	1.6
EB RT at GOY-PARK	73.7	84.2	84.3	84.2	84.2	84.2	84.2
SB TH at GOY-PARK	8.3	9.4	9.0	8.6	8.1	7.8	8.4
WB TH at GOY-PARK	3.6	3.6	3.6	3.6	3.6	3.6	3.6
NB TH at GOY-PARK	2.6	2.7	2.8	6.3	3.9	4.9	2.7
NB TH at OUEL-PARK	99.6	101.7	88.6	88.3	67.2	90.9	88.4
NB RT at OUEL-PARK	64.4	46.5	54.5	32.6	54.3	60.9	73.8
EB RT at OUEL-PARK	109.0	112.4	115.2	117.3	111.8	115.2	115.4
SB TH at OUEL-PARK	132.6	132.5	135.5	135.2	134.9	131.9	136.5
WB RT at OUEL-PARK	11.7	13.0	10.6	13.1	14.1	12.7	13.5
WB TH at OUEL-PARK	11.7	11.9	11.2	11.1	11.6	10.7	11.5
WB LT at OUEL-PARK	15.3	12.6	12.9	12.1	16.0	12.7	12.9
NB TH at OUEL-UNIV	33.9	42.3	32.3	29.6	27.2	29.7	33.8
NB TH at OUEL-UNIV upstream of LT bay	2.5	3.1	4.6	3.7	3.3	4.5	3.4
NB LT at OUEL-UNIV	18.4	30.5	25.8	18.1	19.9	40.3	15.2
EB TH at OUEL-UNIV	10.8	11.0	11.3	9.9	12.4	11.5	9.7
SB TH at OUEL-UNIV	64.8	64.2	64.4	69.8	70.1	67.6	73.4
SB TH at OUEL-UNIV upstream of LT bay	2.4	2.5	64.4	2.5	2.5	2.5	2.6
SB LT at OUEL-UNIV	21.3	20.9	22.3	21.3	21.4	21.4	21.4
EB TH at GOY-UNIV	6.5	3.8	4.6	4.9	3.6	3.1	5.5
SB TH at GOY-UNIV	69.9	69.7	71.8	70.7	69.7	69.9	69.1
WB TH at GOY-UNIV	3.8	3.9	12.3	3.8	3.8	3.8	3.8
NB TH at GOY-UNIV	11.2	11.5	6.1	11.6	11.6	11.5	11.6
To tunnel From WEST	216.5	178.8	257.9	195.3	215.1	194.4	149.5
Through tunnel	347.8	109.8	84.9	348.4	220.6	125.9	84.2
To tunnel from East	76.3	88.9	66.7	118.8	87.9	104.3	60.2
To tunnel from South	173.3	198.8	193.8	196.3	194.2	194.0	183.0

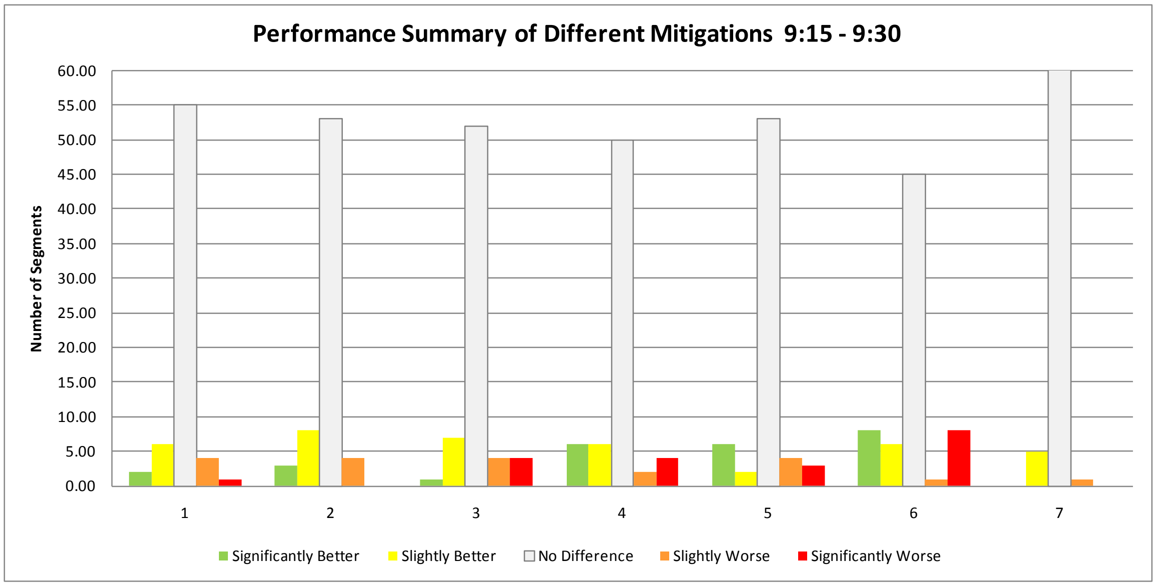


Figure 153. Performance summary 9:15 – 9:30

Table 53. Average delay per link total (3 hours)

Segment	1	2	3	4	5	6	7
EB TH at PEL-WYAN	74.8	169.2	177.2	163.1	161.5	174.0	62.3
WB TH at PEL-WYAN	13.0	12.8	12.4	12.2	11.2	11.1	12.8
NB TH at PEL-WYAN	37.3	136.7	181.7	91.6	142.9	159.4	80.3
EB TH at OUEL-WYAN	37.2	61.4	72.5	60.7	63.0	68.4	27.6
SB TH at OUEL-WYAN	18.4	15.3	19.3	21.7	15.4	8.0	16.7
WB TH at OUEL-WYAN	3.5	3.4	3.4	3.1	3.7	11.9	2.5
NB TH at OUEL-WYAN	33.7	58.5	64.2	63.9	59.3	57.6	32.2
NB TH at OUEL-WYAN upstream of LT bay	7.5	123.3	149.1	149.7	124.8	146.5	35.6
NB LT at OUEL-WYAN	20.3	20.8	19.6	19.7	19.3	12.5	20.7
WB TH at GOY-WYAN	2.9	3.0	11.5	2.6	2.7	2.2	3.6
WB TH at GOY-WYAN upstream of LT bay	66.7	90.9	131.3	87.2	86.5	95.0	64.0
WB LT at GOY-WYAN	53.0	67.2	84.0	67.3	67.9	68.0	69.8
SB TH at GOY-WYAN	49.1	4.7	4.4	4.6	4.2	4.6	4.8
SB TH at GOY-WYAN upstream of LT bay	158.9	206.9	219.5	179.5	174.4	217.2	248.2
SB LT at GOY-WYAN	276.6	4.6	3.2	6.5	7.5	9.0	0.0
Tunnel entrance RT lane	11.4	11.8	17.3	11.9	12.3	12.1	13.4
Tunnle entrance upstream from RT bay	1.2	1.0	5.8	0.9	1.4	1.3	1.1
Just past tunnel entrance on GOY	0.1	0.1	9.5	0.1	0.1	0.1	0.1
WB RT at GOY-WYAN	9.3	14.3	14.0	14.3	15.2	16.8	14.3
WB TH at GOY-WYAN	8.5	11.0	11.9	8.8	8.9	13.7	11.8
WB TH at GOY-WYAN upstream of LT bay	10.4	21.4	20.7	18.1	20.3	21.1	21.8
WB LT at GOY-WYAN	7.9	11.2	9.5	8.6	9.0	12.5	10.1
NB TH at GOY-WYAN	155.4	129.9	152.1	132.2	136.1	129.8	142.0
EB TH at WIND-WYAN	0.6	0.5	0.4	0.6	0.5	1.3	0.5
SB TH at WIND-WYAN	32.4	56.5	50.3	52.5	54.6	57.1	46.5
WB RT at WIND-WYAN	0.5	0.4	0.6	0.6	0.6	0.8	0.4
WB TH at WIND-WYAN	3.8	7.2	7.6	7.7	7.3	7.9	7.5
WB TH at WIND-WYAN upstream of RT bay	9.1	22.2	23.2	22.0	22.3	22.9	22.9
NB TH at WIND-WYAN	27.2	74.6	47.3	65.3	89.8	52.0	48.8
EB TH at MCD-WYAN	2.6	2.9	3.1	1.8	2.8	1.8	2.8
SB TH at MCD-WYAN	16.5	21.0	19.8	16.8	17.9	17.5	19.2
SB TH at MCD-WYAN upstream of RT bay	0.0	0.0	0.0	0.0	0.0	0.0	0.0
WB TH at MCD-WYAN	47.8	151.6	159.2	156.1	151.8	158.6	153.3
NB TH at MCD-WYAN	10.3	9.7	11.0	9.2	9.9	9.5	9.8
NB TH at MCD-WYAN upstream of LT bay	15.8	52.0	61.2	40.2	45.3	33.6	40.5
NB LT at MCD-WYAN	26.8	44.4	46.1	40.9	40.0	39.3	41.5
EB TH at GOY-TUC	169.0	47.3	153.1	46.1	48.6	45.0	61.3
SB TH at GOY-TUC	5.3	4.1	4.1	4.6	4.1	5.2	4.0
WB TH at GOY-TUC	192.9	91.7	126.2	94.9	83.7	83.9	113.9
NB TH at GOY-TUC	109.1	82.5	96.2	83.4	82.8	82.5	91.8
EB TH at GOY-PARK	19.6	19.3	21.2	19.1	19.0	19.2	19.1
EB TH at GOY-PARK upstream of LT bay	3.5	1.2	1.2	2.0	2.1	1.1	1.3
EB RT at GOY-PARK	28.9	27.6	27.9	28.5	28.7	27.6	28.3
SB TH at GOY-PARK	8.0	9.0	11.5	8.8	9.5	15.2	21.3
WB TH at GOY-PARK	13.7	11.7	12.1	11.7	11.7	12.1	12.1
NB TH at GOY-PARK	2.8	2.6	2.8	2.8	2.8	3.1	2.7
NB TH at OUEL-PARK	55.5	53.4	59.7	45.9	44.5	71.2	71.1
NB RT at OUEL-PARK	31.3	30.6	33.7	31.5	29.9	32.8	29.2
EB RT at OUEL-PARK	31.0	30.9	30.5	31.0	31.0	30.9	30.9
SB TH at OUEL-PARK	41.9	42.0	42.1	42.1	42.0	42.2	42.5
WB RT at OUEL-PARK	12.1	12.1	12.3	12.0	12.0	12.2	12.2
WB TH at OUEL-PARK	11.1	11.0	11.0	11.1	11.0	11.2	11.1
WB LT at OUEL-PARK	11.1	11.1	11.2	11.0	11.0	11.1	11.1
NB TH at OUEL-UNIV	9.6	9.9	9.6	10.1	9.8	8.8	9.5
NB TH at OUEL-UNIV upstream of LT bay	5.1	4.8	4.8	4.9	5.0	4.9	6.0
NB LT at OUEL-UNIV	11.2	11.6	11.6	11.7	11.8	10.6	11.8
EB TH at OUEL-UNIV	8.4	8.2	8.1	8.1	8.0	9.1	8.7
SB TH at OUEL-UNIV	17.4	17.8	19.0	18.0	17.8	17.2	17.6
SB TH at OUEL-UNIV upstream of LT bay	2.9	3.1	19.0	3.1	3.0	2.9	2.9
SB LT at OUEL-UNIV	20.0	21.2	20.8	20.5	21.0	21.2	20.9
EB TH at GOY-UNIV	5.7	4.1	7.3	5.2	3.4	6.9	11.9
SB TH at GOY-UNIV	26.2	25.4	28.8	26.3	26.9	31.9	35.9
WB TH at GOY-UNIV	7.1	7.4	11.2	6.4	5.2	9.5	6.2
NB TH at GOY-UNIV	5.7	5.8	5.4	5.7	5.8	7.1	5.7
To tunnel From WEST	480.5	566.8	689.6	584.2	608.3	601.4	531.2
Through tunnel	850.3	850.4	883.5	852.5	848.5	865.0	854.5
To tunnel from East	260.4	318.2	412.6	343.8	344.5	319.1	348.8
To tunnel from South	412.8	390.9	445.7	398.7	411.0	403.0	421.3

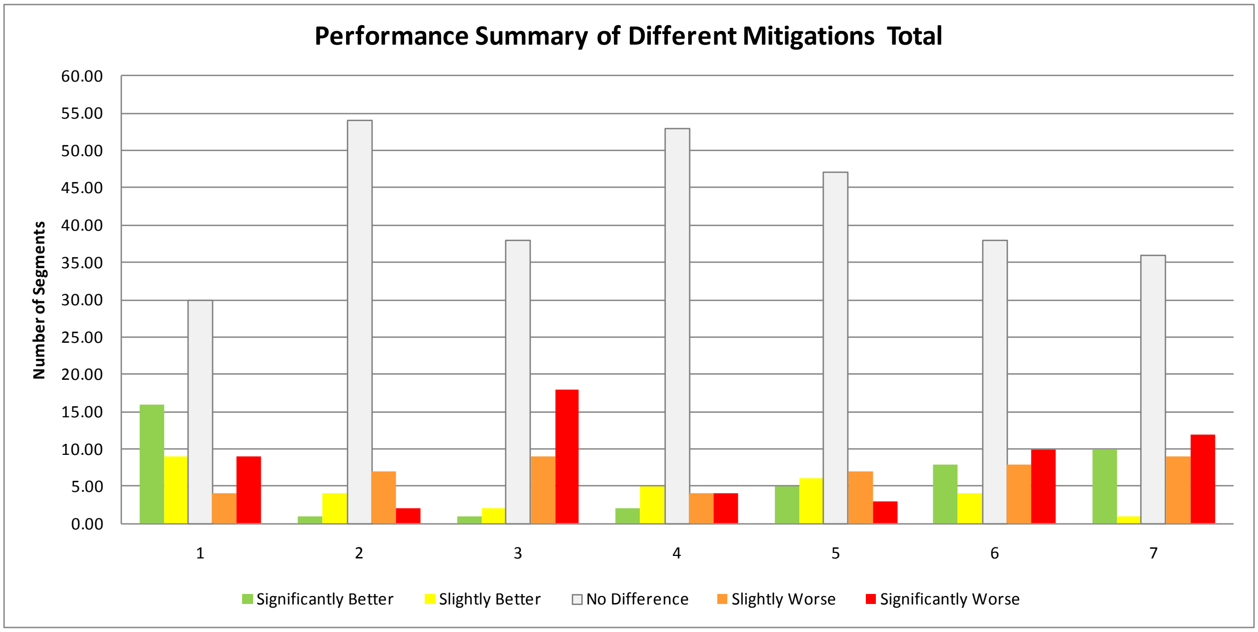


Figure 154. Performance summary total (3 hours)

Once the situation is oversaturated, phase failures are unavoidable and TOSI quickly grows to 100% and higher on every link in the system. A significant number of links experience SOSI > 0, sometimes up to 50% or more of upstream green time is wasted when the downstream queue cannot move. It appears that more extensive mitigations such as the combined metering approaches make the situation worse than doing nothing at all, but most mitigation strategies had a combination of movements that were better and worse and mostly cancelled each other out. Doubling the capacity of the left-turn movement (Strategy 3) is a particularly poor concept that tends to make situations much worse at more than 25 movements, without any corresponding improvements on other movements.

Only the re-routing strategy produced significant improvement during the first 45 minutes following the triggering of the incident. After the 8:00-8:15 A.M. period, the initial queue caused by the prolonged processing time clears and improvement is not apparent after this time because all tunnel traffic is being re-routed to enter through the north. Interestingly, the simplest approach envisioned by the Province and the City initially (without any simulation modeling or extensive analysis), Strategy 1, appears to improve conditions on the largest number of movements. This is an important lesson learned for the research.

Throughput Analysis

The number of vehicles in the system was calculated by comparing the vehicle input and output data recorded by the simulation. The average input rates for each mitigation strategy as well as the Vissim demand input are shown in Figure 155. This figure illustrates that overflow queues on the arterials are inhibiting vehicles to even enter the system at the rate that the model is demanding.

As shown in the figure, each mitigation strategy resulted in very similar input rates. Most, but not all, mitigations are able to slightly outperform doing no mitigation at all. As reflected by the average delay performance results presented above, the Windsor Logic (Strategy 1) had the highest input rate of the mitigations during the time of the incident. Recall that the incident begins at time 3600s and extends until 7200s. Due to the length of the tunnel (approximately one mile), it is not for approximately 45 minutes after the incident starts that vehicles are restricted from entering the system.

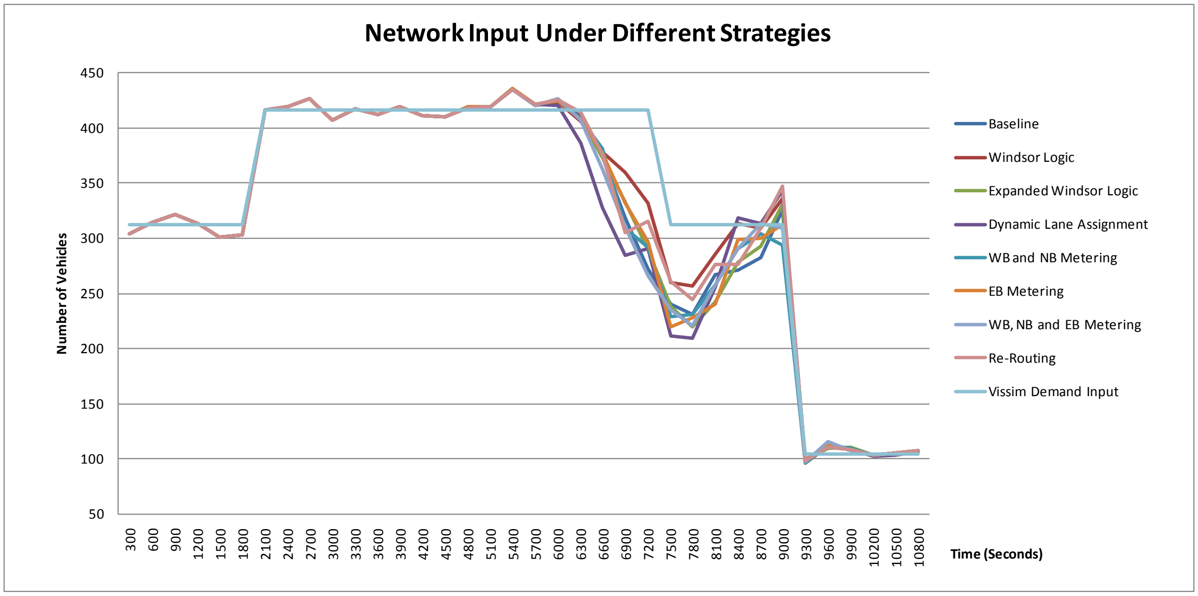


Figure 155. Average input rates under different mitigations

The network output graph shown in Figure 156 illustrates that each mitigation strategy impacts the output differently as compared to the baseline no mitigation condition. The Windsor Logic strategy and the Re-Routing strategy have the greatest impact on output. This matches the performance results presented for the average delay by link.

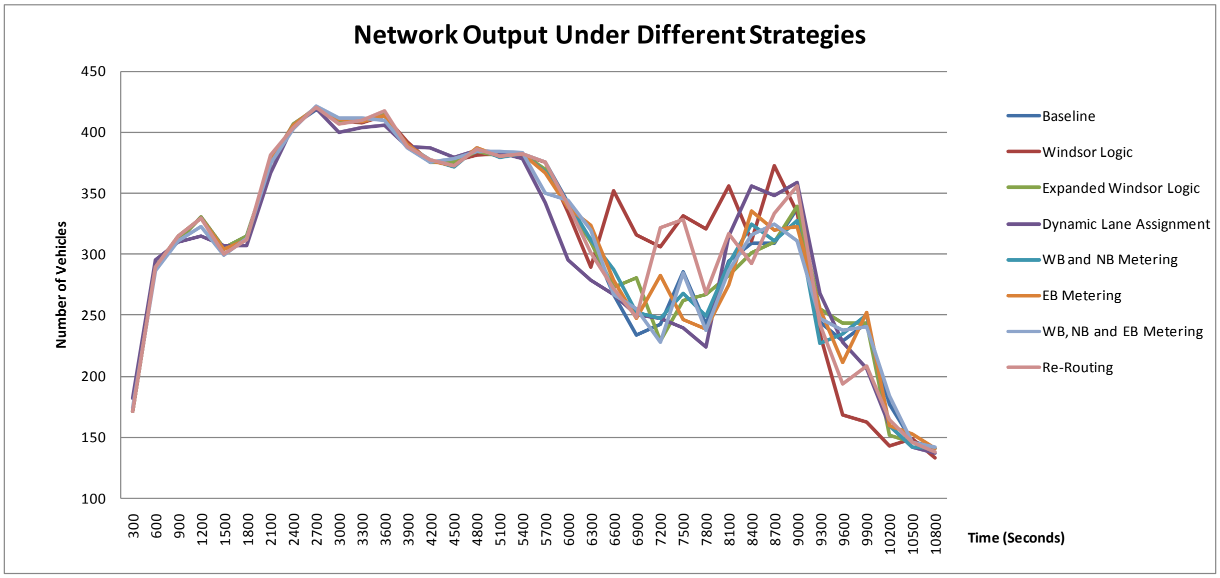


Figure 156. Average output rates under different mitigations

The average number of vehicles in the system was calculated using the input and output data shown above. Figure 157 shows the resulting number of vehicles in the system for each strategy.

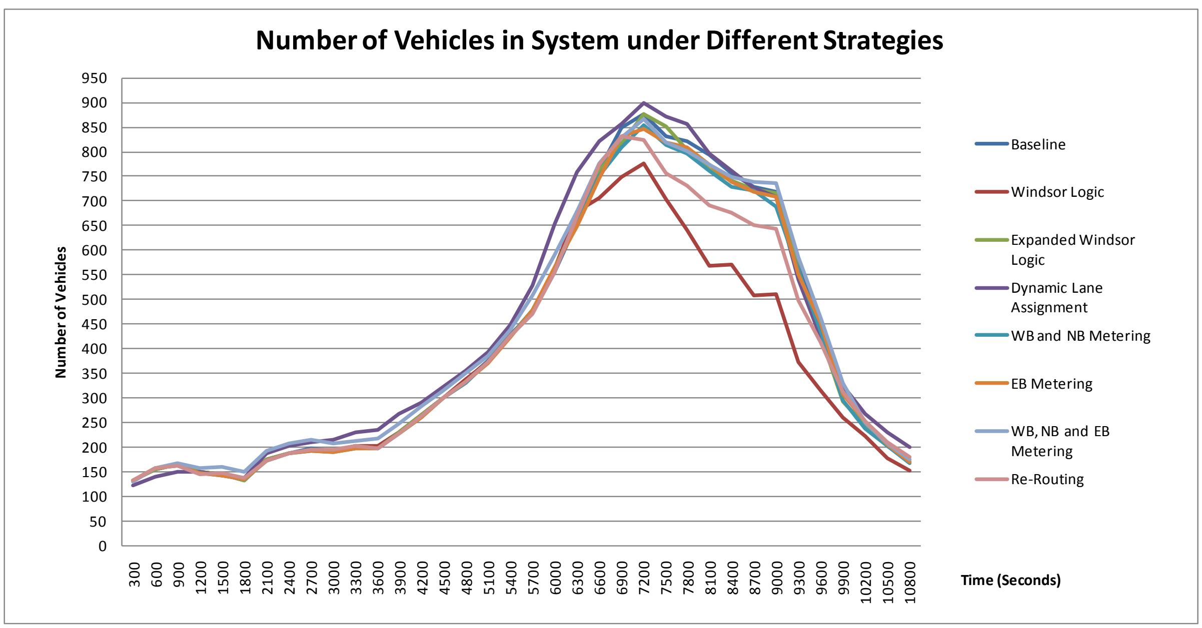


Figure 157. Average vehicles in the system

Determining the best strategy from this presentation of data depends on the desired outcome of the strategy. For example, if the objective is to reduce the number of vehicles stuck in the system, one would choose a strategy which results in a data line which falls below the baseline. However, if the objective is to utilize the storage space within the system, choosing a mitigation which falls above the baseline date would be appropriate. The Dynamic Lane Assignment mitigation strategy appears to result in the highest number of vehicles in the system while still recovering in a time consistent with the no mitigation scenario. Performance improvements for total average delay and throughput do not support that the dynamic left-turn strategy is particularly effective. The Windsor Logic mitigation strategy keeps the number of vehicles in the system lower, and, as a result, appears to recover sooner than any other strategy. This appears to indicate a strong lesson learned from this test case that it is typically better to use a simple solution in an extremely challenging situation.

Summary

This test case demonstrated the application of the online congestion management tool to oversaturated conditions. A number of logic statements were constructed to select different types of mitigation plans based on the TOSI and SOSI values measured from detectors near the incident location. The first strategy used logic conditions envisioned by the City and the Province that enact actions just at the critical intersection. An extension of this logic was theorized that brought in data from additional detectors and enacted additional actions. A number of other offline mitigation strategies were applied that considered the start time of the incident would be easily able to be identified shortly after the queuing in the tunnel began using existing CCTV resources.

Because of the single point of failure condition at the tunnel entrance, the mitigations that were tested had very similar performance and did not show significant improvements to the "do nothing" strategy. Some were able to improve total system input and output and reduce the total number of vehicles in the system. Notably, as stated earlier, the original Windsor logic which takes the fewest number of actions possible was the most effective. This validates the principle that in a challenging situation with a single point of failure, taking a minimum number of mitigation actions may be optimal.

Since more than 50% of the vehicles were destined for the same location, it was not possible to clear enough storage blocking queues to get other (<50%) vehicles destined for other locations moving enough to offset the delays experienced by the vehicles destined for the tunnel. We had originally envisioned that we might be able to construct a strategy where the vehicles not destined for the tunnel could have been provided greater mobility, but the situation was too challenging. In such a situation, it appears that choosing the smallest number of actions was the most effective. However, it is important to note that we did find that the Windsor mitigation logic was statistically more effective than the baseline "do nothing" operation. This shows again, as has been shown in all of the other test cases, that mitigation strategies can be effective in improving system-wide conditions during oversaturation. There is no reason to resort to the aphorism that "there is nothing that can be done. There is simply too much traffic".

In the next section, we will present another test case that illustrates that significant improvements to performance are possible with mitigation strategies. In this test, we followed a process based on engineering judgment, with the intended outcome to gather additional evidence and experience with a wide range of potential mitigations.

Test Case: Arterial with Special Event Traffic

This real-world test case concerns a heavily traveled arterial that becomes significantly oversaturated on several critical routes. This oversaturation happens intermittently during A.M. and P.M. peak periods due to surges in traffic due to day-to-day variability but also when there are crashes or incidents. In addition, the arterial experiences oversaturation during P.M. peak periods when event traffic is overlaid on the already heavy through flows. In this test case, we studied the application of various mitigation strategies at the five intersections near the event location.

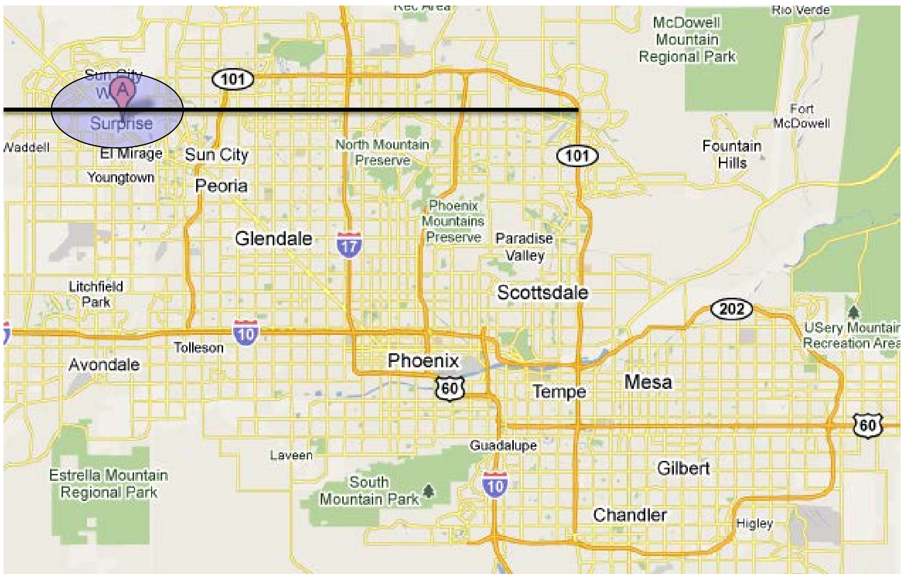


Figure 158. Location of test case in the Phoenix, AZ metropolitan area

Bell Road is located in the northwest Phoenix, AZ metropolitan area as illustrated in Figure 158. Portions of this roadway carry 70,000 vehicles per day. Because of its location relative to major freeway connections to central Phoenix, there is a pattern of high commuter traffic eastbound in the A.M. peak and westbound in the P.M. peak. The relative P.M. peak travel directional flows are illustrated in Figure 159.

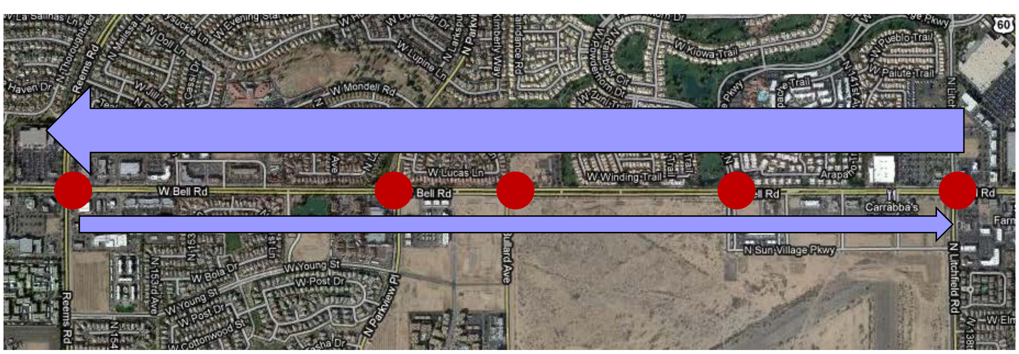


Figure 159. Illustration of relative flows along the arterial during P.M. peak

The current signal coordination timing during P.M. peak period consists of a 130s cycle length and offsets which favor forward progression in the westbound direction. The splits, offsets, and corresponding progression pattern are illustrated in Figure 160. The event traffic is destined for Bell Road and Bullard Avenue which is the fourth intersection down from the top of the figure.

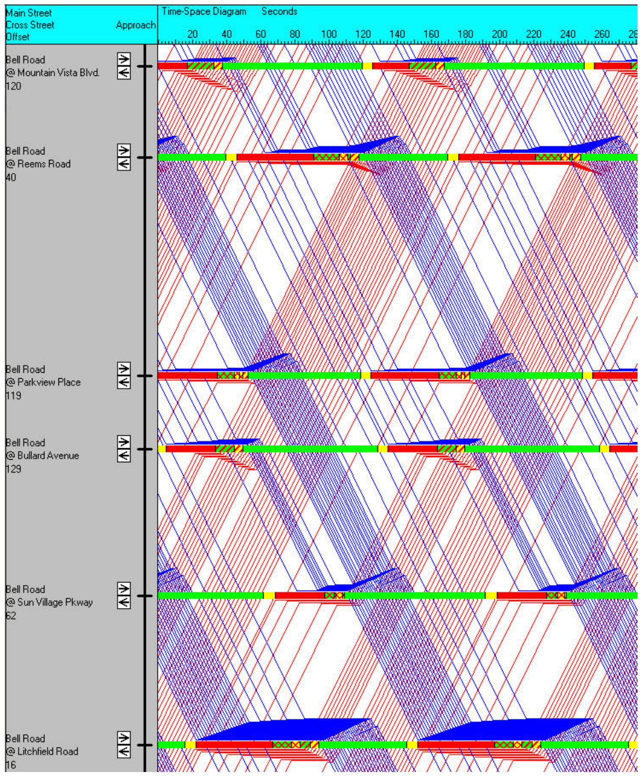


Figure 160. Progression patterns during the P.M. peak

During late February, March, and early April, the commuter traffic along Bell Road is elevated due to additional vehicles traveling to the baseball Spring Training facility for the Texas Rangers and Kansas City Royals located just south of Bell Road. These additional critical routes are illustrated in Figure 161. The result is a significant increase in westbound left turning vehicles at Bullard which causes blocking and starvation for the through phases at the intersections upstream of Bullard to the east since the primary access to this part of the Valley is via the northbound or westbound 101 Loop freeway. During night games, the game traffic is coupled with the normal commuter traffic and the arterial operation breaks down quickly. Some oversaturation is also caused by the increase in traffic heading eastbound towards the facility, but these backups are not nearly as significant as the westbound problem.

The forward progression offsets fail to operate efficiently just a few minutes after the game traffic begins to arrive for the start of the game. Queue lengths quickly approach the entire length of the links between Bullard and Sun Village (2600ft), and between Sun Village and Litchfield Road (2600ft). With the baseline 130s plan, forward progression offsets, and split times, it typically takes two to three cycles for vehicles to traverse each of the three links between Litchfield and

Bullard. What is typically a four and a half minute trip along the three mile segment becomes more than 15 minutes; with more than 12 of those minutes spent in queues between Litchfield and Bullard.

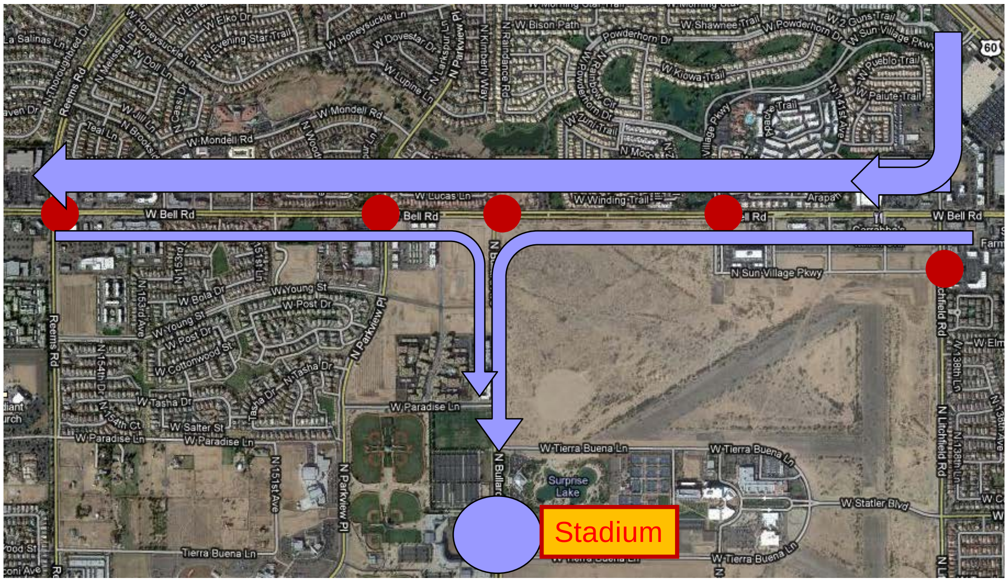


Figure 161. Critical routes during game overlaid with P.M. peak flows

This case study is allocated to the oversaturated scenario taxonomy as shown in Table 54.

Table 54. Allocation of Bell Road game traffic case study on the oversaturated scenario 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		Unplanned Events		
Two-way arterial		**Planned Events**		
Interchange				
Grid				
Network				

This scenario is definitely a two-way arterial problem with several significant critical routes. The queuing lasts for at least an hour and a half as game attendees typically arrive approximately an hour ahead of the first pitch and continue arriving approximately 15 minutes after that. Another 15-20 minutes is required before the queues dissipate and normal traffic operations resume. A dynamic map of the typical queue construction and dissipation process is illustrated in Figure 162 and Figure 163.

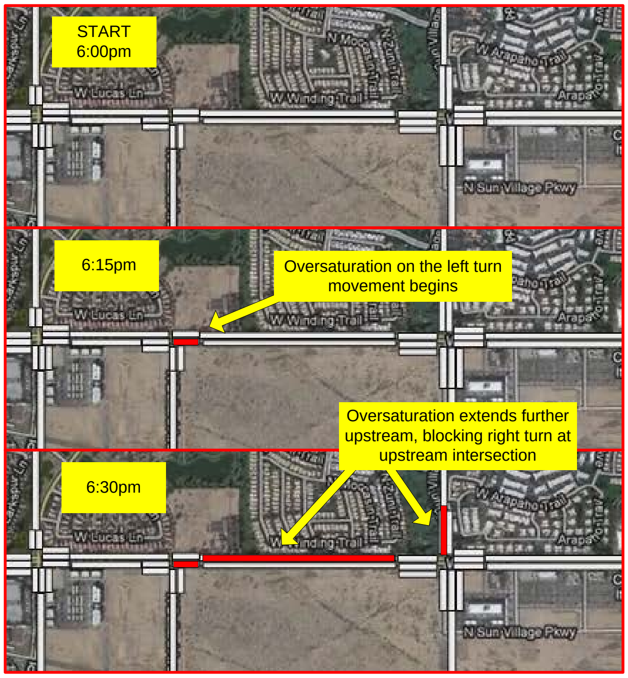


Figure 162. Queue growth at the beginning of the arriving event traffic

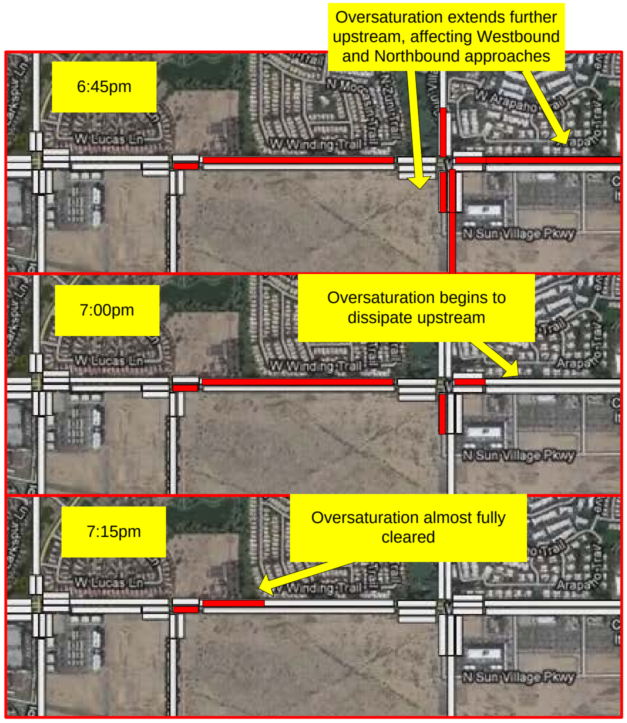


Figure 163. Queue dissipation as the event traffic flows subside

This condition is caused by three major factors (a) the event traffic, (b) the existence of a single lane left-turn bay with only 225 ft of storage, and (c) the lack of adequate split time for the left turn. In addition, the common cycle time along the arterial cannot (or was not being) be adjusted because further upstream to the east of the Litchfield intersection is an intersection that is managed by another agency. The situation is recurrent and predictable, since the start time of every Spring Training game is published and known in advance. Attendance at each game is relatively stable as the attendance nears or meets the capacity of the stadium for most night games.

Simulation Test Configuration

To simplify the test conditions for simulation analysis without onerous data collection, some basic assumptions were made about the game traffic and their origins based on agency and engineer experience with the location and the travel conditions. The baseline condition consisted of increased P.M. peak hour volumes due to event traffic with no change to the P.M. peak coordination timing plan parameters. An additional 7,000 vehicles (approximating that two persons per vehicle attend the event; the stadium holds 14,000) were overlaid on the background vehicles in the P.M. peak period. Of that traffic, 85% was assumed to approach from the west and 15% from the east.

The over-all traffic profile was adjusted according to the time of day profile shown in Figure 164.

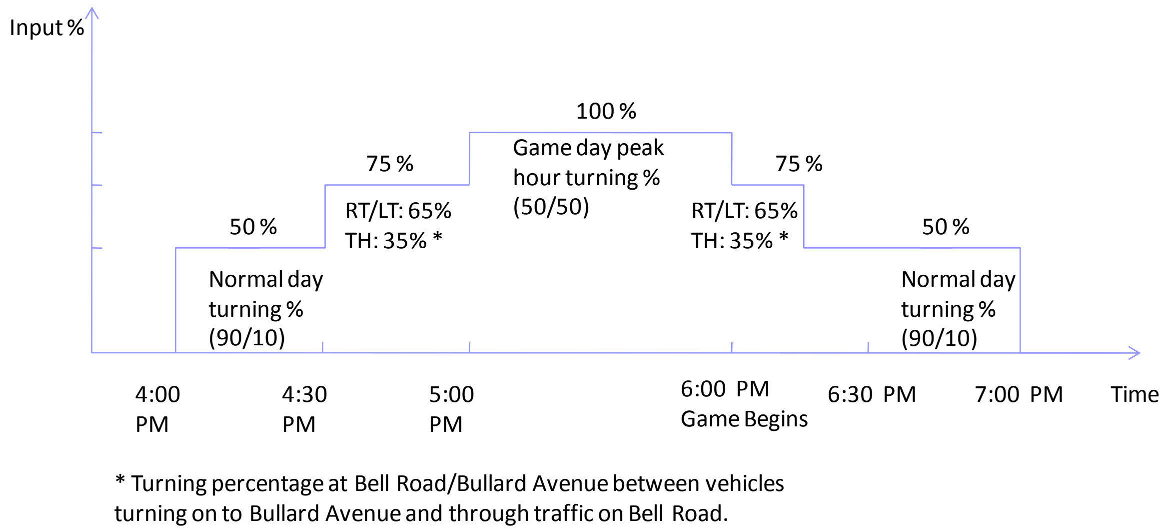


Figure 164. Traffic arrival volumes and turning percentage profile during game traffic

As shown, in addition to the ramp-up and ramp-down of the arrival volumes, the route proportions at Bell and Bullard in the westbound direction were adjusted in the following manner to represent the change in the mix of game and through traffic destinations. The eastbound approach volumes were adjusted to represent the additional game traffic, but the turning percentages were not modified. Five simulation runs were executed for the baseline case and for all mitigation strategies the performance data was averaged over the iterations in the presentation figures. Common random number seeds were used for all five runs for all strategies to reduce the variance effects.

The resulting input, output, and number of vehicles in the system are shown in Figure 165. The x-axis units are the number of seconds since the beginning of the simulation time.

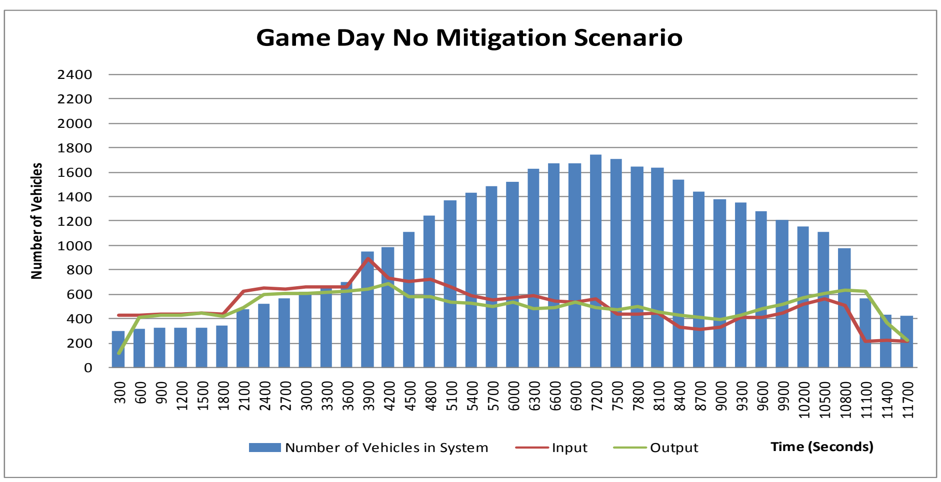


Figure 165. Number of vehicles in system and I/O rates of baseline scenario

Notice that until the game traffic begins to arrive (approximately one hour or 3600s into the simulation) the traffic condition is relatively stable. At 75% of the peak hour total volume, the number of the vehicles in the system remains relatively constant, but the queues begin to form at the westbound left-turn movement and quickly spill back into the next two intersections causing SOSI > 0 and subsequent queuing on northbound left turn at Sun Village and right turn blockages at Sun Village and Litchfield as illustrated in the dynamic maps in Figure 162 and Figure 163. The left-turn queue at Bullard also blocks the through movement at Bullard westbound causing starvation.

Mitigation Strategy Development

The process of choosing mitigation strategies began with addressing the most obvious problem(s) first and working outward to identify other symptoms and potential mitigations. In this test case, we considered only one change to timing plan parameters to implement the mitigation (i.e. we did not expressly consider timing plans for loading, processing, and recovery). Each of the mitigation strategies was implemented by time of day schedule one hour and 30 minutes before the scheduled game time. This start time was selected based on the previous experience which indicated that the most severe increase in volumes begins one hour before the first pitch. An additional 30 minutes was provided to allow the controllers along the arterial to transition and settle into the selected mitigation operation during the time that the volumes were beginning to ramp up at 75% of the peak hour flow rate. In mitigation strategies which include changes to offsets, the offset at Bullard remained constant to avoid transition time at the critical intersection. Five iterations of each mitigation strategy were run and the results were compared to the baseline to determine better and worse conditions in terms of average delay, throughput, and travel time. All simulations used Vissim and the Virtual D4 traffic controller at all intersections in the model.

This oversaturated scenario is generated from the inadequacy of the left-turn phase split westbound off of Bell Road at Bullard Avenue toward the Spring Training facility, so the major objective of the mitigations was to solve this problem first. All of the mitigations included some level of green time re-allocation to increase the left-turn split. Additional objectives were to attempt to alleviate the westbound through movement blocking at Bullard, managing westbound queues interactions at Sun Village and Litchfield and managing the eastbound queues developed because of the adjustments to the other parameters to help the westbound movements.

Other mitigation strategies were theorized and developed as the results of the previous mitigations were analyzed. The mitigation approaches were developed in the order shown in Table 55, but the purpose of some of the strategy development was done simply to evaluate the differences between the various mitigation strategies. In some cases, one strategy would create queuing or delay in another segment of the arterial so additional strategies were tested which attempted to mitigate the 'new' congestion as well as the original critical movement.

Table 55. Mitigation strategies evaluated in this test case

Mitigation Strategy	Description
Extreme Left-Turn Split at Bullard	WB Left-Turn was significantly increased to address the queue causing blocking.
Negative Offsets	In addition to the increased left-turn split at Bullard, negative offsets were implemented in the westbound direction to potentially clear downstream residual queue.
Simultaneous Offsets	In addition to the increased left-turn split at Bullard, simultaneous offsets were implemented in the westbound direction.
Double Cycle at Bullard	Cycle length at Bullard was reduced to 65 seconds to serve the critical movement more frequently.
Resonant Cycle	135 second cycle length along the entire corridor
Moderate Left-Turn Split at Bullard	Game traffic is also generated from the west. The greatly increased left-turn split westbound at Bullard caused significant backup for the eastbound direction at Bullard. To address this, a more moderately increased left-turn split was tested.
Dynamic Lane Assignment	In this mitigation strategy, the WB left-turn movement changed from a single turn lane to duals (converting one westbound through lane to an additional left-turn lane) by a time of day schedule.
Reduced Cycle Length	A cycle length of 90 seconds along the entire corridor was implemented to test if the residual queue lengths could be reduced by reducing the queue at downstream intersections by reducing the available green at upstream intersections.

Design of Negative Offsets

- 130s cycle is used to coordinate with the Arizona Department of Transportation (ADOT) signal to East of evaluated system
- Natural spacing is ½ and ¼ mile grid network (90% of all of Phoenix)
- Adjust offsets appropriately; provide all additional split (5s) to coordinated movements

Dynamic lane assignment

- Convert left-most through lane at Bullard to left-turn only lane during game time
- Five-section head, blank out sign, upstream DMS sign warning traffic to merge RIGHT if not going to game
- Easy in Vissim by modifying signal head model and vehicle route logic by TOD (no mods necessary to Virtual D4 controller)
- Illustrated in Figure 166, below

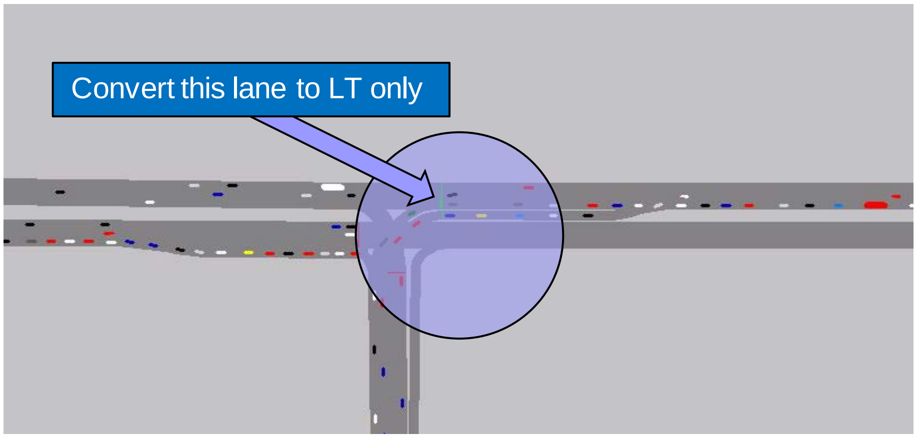


Figure 166. Illustration of dynamic lane allocation for two-lane left-turn movement

Average Delay Analysis

The results for the average delay analysis are presented in table format per system link for each half hour of the simulation and are color coded to illustrate the degree of the impact on average delay. In Figure 167 through Figure 171and Table 56 through Table 60, green indicates the average delay experienced during the mitigation is significantly better as compared to the no mitigation scenario. Yellow indicates slightly better, white indicates little or no change, orange indicates slightly worse, and red indicates significantly worse than average delay results. The number of links experiencing each level of performance is also included. In addition, Figure 172 and Table 61 illustrate the results for total average delay per system link for the three hour simulation. For each of the figures below, the following number identifiers were used:

2a – Extreme Left-Turn Split at Bullard

3 – Negative Offsets

4 – Simultaneous Offsets

5 – Double Cycle

6 – Resonant Cycle

7 – Moderate Left-Turn Split at Bullard

8 – Dynamic Lane Assignment

9 – Reduced Cycle Length

The results of the average delay comparison indicated that each mitigation strategy reduces delay on links on the east end of the system while increasing delay on links located on the west end of the system.

Table 56. Average delay per link 4:30 – 5:00

Segment	2a	3	4	5	6	7	8	9
EB TH at Reems-Bell	21.8	21.6	22.1	22.2	20.4	22.0	22.1	20.0
EB TH at Reems-Bell Upstream of LT Bay	1.3	1.3	1.4	1.4	1.2	1.4	1.4	1.3
NB RT at Reems-Bell	8.9	8.2	8.3	8.2	8.6	8.5	8.4	8.6
NB TH at Reems-Bell Upstream of RT Bay	0.9	0.9	0.9	0.9	1.0	0.9	0.9	0.9
SB LT at Reems-Bell	38.0	42.1	41.7	41.7	43.2	41.7	41.7	32.7
SB TH at Reems-Bell Upstream of LT Bay	1.8	2.6	2.6	2.6	2.6	2.6	2.6	1.3
EB TH at Parkview-Bell	4.2	5.3	13.3	16.2	12.1	12.9	13.0	14.4
EB TH at Parkview-Bell Upstream of LT Bay	2.2	3.4	4.6	9.1	3.7	4.4	4.2	5.6
NB RT at Parkview-Bell	3.8	4.0	3.6	4.4	3.6	4.1	3.7	3.6
NB TH at Parkview-Bell Upstream of LT Bay	0.2	0.2	0.2	0.3	0.2	0.2	0.2	0.2
SB LT at Parkview-Bell	44.0	39.2	41.5	39.4	43.3	40.1	41.3	21.9
SB TH at Parkview-Bell Upstream of LT Bay	5.9	3.1	3.3	3.3	3.1	3.4	3.4	1.1
EB TH at Bullard-Bell	15.8	23.1	24.5	21.1	24.9	15.2	16.0	11.6
EB RT at Bullard-Bell	12.3	18.5	17.3	22.7	17.9	11.9	12.0	13.1
EB TH at Bullard-Bell Upstream of RT Bay	3.9	5.3	6.6	17.3	7.6	2.8	3.0	3.6
WB LT at Bullard-Bell	16.8	13.5	14.1	15.4	12.3	19.4	19.1	15.8
WB TH at Bullard-Bell	0.9	1.2	3.8	2.7	4.0	3.4	3.7	2.9
WB TH at Bullard-Bell Upstream of LT Bay	13.1	8.7	12.1	12.6	8.6	35.1	32.4	7.8
NB LT at Sun Village-Bell	60.2	55.4	53.6	60.9	64.8	56.6	59.3	39.5
NB TH at Sun Village-Bell Upstream of LT Bay	1.3	1.4	1.4	1.4	1.4	1.4	1.4	1.4
WB TH at Sun Village-Bell	0.8	3.0	2.1	2.0	1.9	3.7	3.6	5.9
WB TH at Sun Village-Bell Upstream of LT Bay	1.6	7.7	1.8	1.7	1.7	3.5	3.3	5.2
SB TH/RT at Sun Village-Bell	0.1	1.9	1.1	0.9	2.4	1.1	1.1	2.0
SB TH at Sun Village-Bell Upstream of LT Bay	1.8	1.8	1.8	1.8	1.8	1.8	1.8	1.8
NB LT at Litchfield-Bell	59.3	58.0	58.8	60.1	62.2	58.8	56.8	48.2
NB TH at Litchfield-Bell Upstream of LT Bay	0.7	0.7	0.7	0.7	0.8	0.7	0.7	0.6
WB TH at Litchfield-Bell	19.1	19.0	18.0	18.3	18.2	17.8	17.9	26.1
WB TH at Litchfield-Bell Upstream of LT Bay	15.1	17.8	13.7	14.4	14.4	13.1	13.6	68.4
SB TH at Litchfield-Bell	45.2	43.5	48.2	47.6	51.8	48.2	48.2	31.4
SB TH/RT at Litchfield-Bell Upstream of LT Bay	1.1	1.1	1.1	1.2	1.2	1.2	1.2	0.9

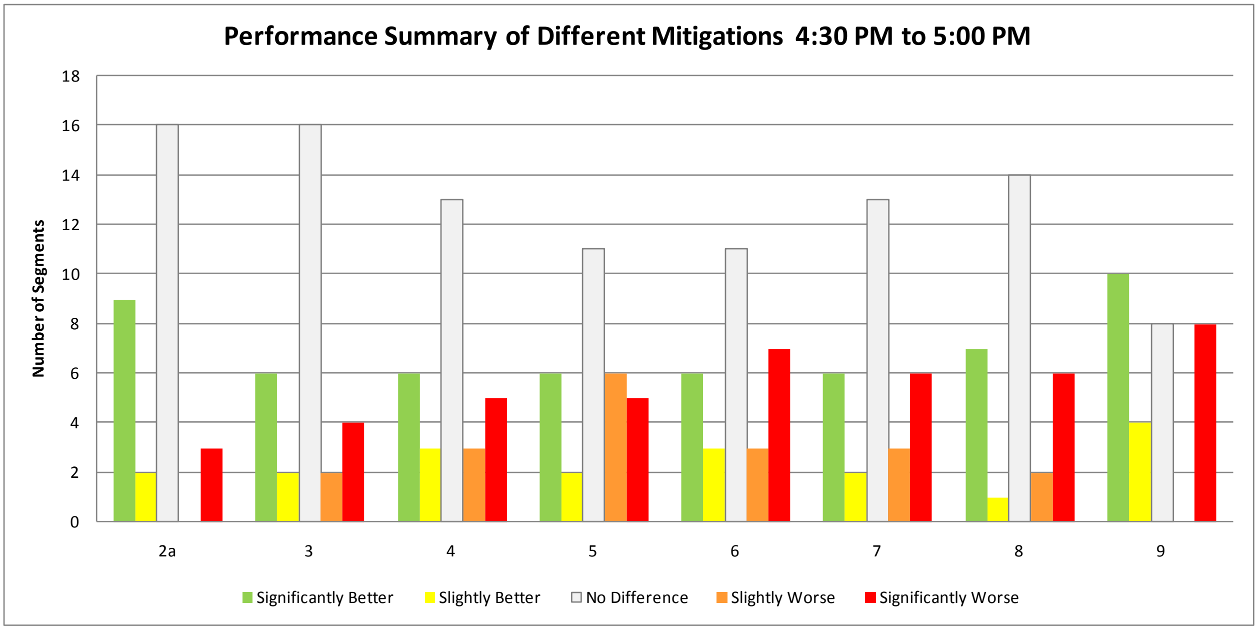


Figure 167. Performance summary 4:30 – 5:00

Table 57. Average delay per link 5:00 – 5:30

Segment	2a	3	4	5	6	7	8	9
EB TH at Reems-Bell	23.1	22.6	23.3	27.6	23.0	22.0	22.3	23.9
EB TH at Reems-Bell Upstream of LT Bay	3.2	2.9	3.0	4.9	3.4	2.9	3.0	2.9
NB RT at Reems-Bell	15.9	16.1	16.5	41.6	22.2	13.3	13.8	15.7
NB TH at Reems-Bell Upstream of RT Bay	12.2	8.5	9.0	19.2	16.1	9.0	11.5	5.0
SB LT at Reems-Bell	54.2	54.0	54.4	54.1	59.7	54.6	54.6	34.0
SB TH at Reems-Bell Upstream of LT Bay	27.4	33.3	33.6	31.1	49.2	36.0	33.3	4.1
EB TH at Parkview-Bell	17.3	17.2	24.1	29.8	25.4	21.2	22.6	22.0
EB TH at Parkview-Bell Upstream of LT Bay	67.4	71.7	74.6	115.9	82.2	59.6	66.4	75.1
NB RT at Parkview-Bell	5.4	5.3	6.2	5.3	5.5	7.4	6.1	5.5
NB TH at Parkview-Bell Upstream of LT Bay	0.4	0.4	0.5	0.3	0.3	0.4	0.3	0.3
SB LT at Parkview-Bell	59.6	52.3	54.3	62.5	60.9	50.6	52.5	24.8
SB TH at Parkview-Bell Upstream of LT Bay	41.4	19.3	26.1	41.6	55.4	21.5	18.8	1.9
EB TH at Bullard-Bell	18.4	42.3	23.8	25.2	30.3	13.0	12.9	13.8
EB RT at Bullard-Bell	22.7	23.7	23.3	29.6	26.7	18.5	21.2	22.4
EB TH at Bullard-Bell Upstream of RT Bay	40.1	40.3	40.9	61.1	47.9	30.5	35.5	41.2
WB LT at Bullard-Bell	19.9	16.5	18.2	16.1	14.6	25.7	25.4	21.9
WB TH at Bullard-Bell	2.3	2.1	2.9	3.0	3.1	2.9	3.0	3.0
WB TH at Bullard-Bell Upstream of LT Bay	104.0	86.0	93.9	94.8	72.4	183.4	99.2	80.8
NB LT at Sun Village-Bell	164.8	70.5	145.6	130.2	153.2	134.1	118.5	49.4
NB TH at Sun Village-Bell Upstream of LT Bay	5.0	1.5	7.6	2.0	5.2	13.7	2.1	1.5
WB TH at Sun Village-Bell	13.0	13.1	13.5	13.2	11.0	25.5	16.4	14.5
WB TH at Sun Village-Bell Upstream of LT Bay	71.5	67.9	71.6	67.4	59.5	127.7	99.0	62.3
SB TH/RT at Sun Village-Bell	3.8	4.3	2.7	2.3	3.9	2.7	2.8	3.3
SB TH at Sun Village-Bell Upstream of LT Bay	1.8	1.8	1.9	1.8	1.8	1.9	1.9	1.8
NB LT at Litchfield-Bell	112.1	90.8	140.1	102.3	144.3	187.6	178.5	80.5
NB TH at Litchfield-Bell Upstream of LT Bay	18.2	7.3	41.2	10.8	39.2	111.8	94.6	6.6
WB TH at Litchfield-Bell	30.0	28.0	29.6	28.8	25.9	38.2	35.4	35.9
WB TH at Litchfield-Bell Upstream of LT Bay	181.9	171.8	177.6	169.9	165.6	212.3	229.6	229.6
SB TH at Litchfield-Bell	68.0	70.3	74.8	80.4	80.2	78.4	79.7	38.7
SB TH at Litchfield-Bell Upstream of LT Bay	13.3	16.1	22.5	18.7	18.8	27.1	22.5	1.8

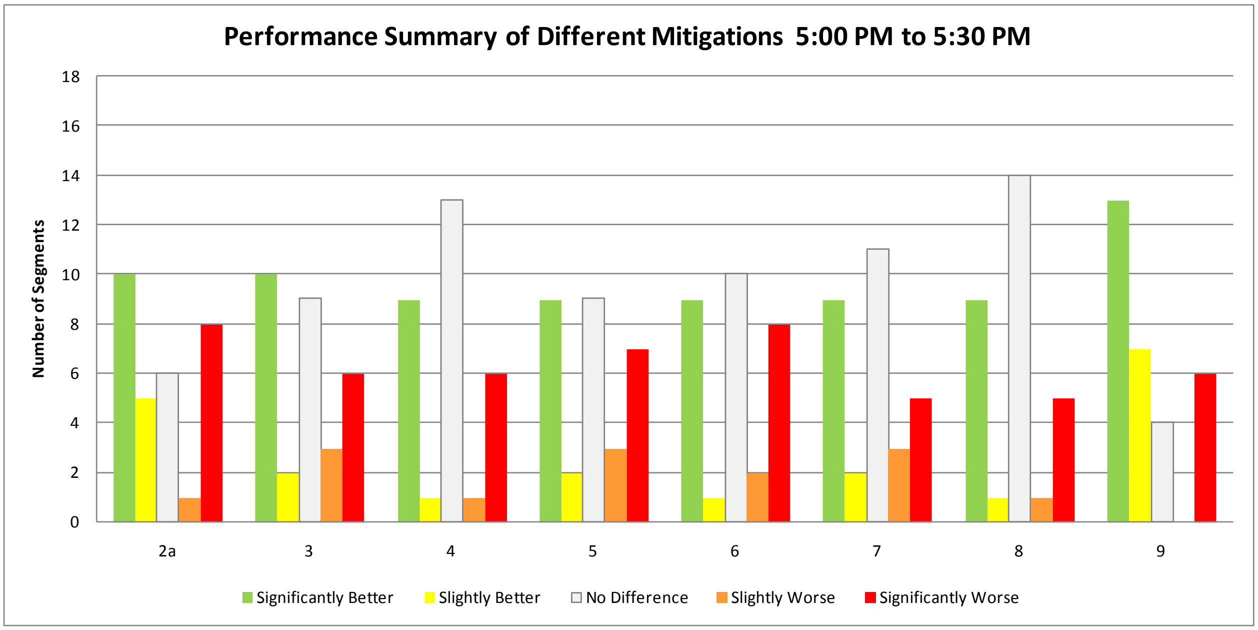


Figure 168. Performance summary 5:00 – 5:30

Table 58. Average delay per link 5:30 – 6:00

Segment	2a	3	4	5	6	7	8	9
EB TH at Reems-Bell	50.3	50.4	49.9	83.1	63.5	32.3	43.5	64.6
EB TH at Reems-Bell Upstream of LT Bay	42.7	38.2	37.1	167.3	78.8	6.4	23.4	72.6
NB RT at Reems-Bell	129.3	196.9	277.5	307.4	288.0	98.3	216.3	223.4
NB TH at Reems-Bell Upstream of RT Bay	137.3	203.4	224.1	417.9	326.8	74.0	117.7	136.6
SB LT at Reems-Bell	59.5	54.0	52.6	71.1	57.0	55.1	55.9	42.3
SB TH at Reems-Bell Upstream of LT Bay	109.6	87.1	116.6	120.1	146.5	135.2	129.0	11.2
EB TH at Parkview-Bell	41.3	40.3	40.0	63.9	49.8	29.3	35.7	45.2
EB TH at Parkview-Bell Upstream of LT Bay	366.0	359.0	352.5	586.6	445.9	229.6	301.4	412.8
NB RT at Parkview-Bell	5.6	5.1	6.1	5.8	6.1	5.4	5.7	6.0
NB TH at Parkview-Bell Upstream of LT Bay	0.5	1.6	0.6	0.8	0.5	0.5	0.8	0.4
SB LT at Parkview-Bell	67.3	61.6	69.1	87.5	83.9	51.7	64.8	29.0
SB TH at Parkview-Bell Upstream of LT Bay	41.4	30.5	44.5	118.1	108.7	13.6	34.4	2.1
EB TH at Bullard-Bell	29.4	33.6	28.2	23.6	38.8	12.6	13.6	16.5
EB RT at Bullard-Bell	23.6	23.9	24.2	29.7	26.6	19.4	23.4	23.9
EB TH at Bullard-Bell Upstream of RT Bay	118.6	116.1	116.4	186.6	148.9	73.1	103.8	122.0
WB LT at Bullard-Bell	16.7	16.3	17.2	14.7	13.6	28.1	24.6	18.9
WB TH at Bullard-Bell	3.0	3.3	2.6	3.5	2.7	2.1	2.7	3.7
WB TH at Bullard-Bell Upstream of LT Bay	276.5	259.0	260.5	239.4	208.0	442.5	164.3	279.3
NB LT at Sun Village-Bell	355.2	194.4	295.3	255.4	361.0	177.2	179.7	135.2
NB TH at Sun Village-Bell Upstream of LT Bay	215.7	25.8	114.9	66.3	199.4	69.5	13.5	8.5
WB TH at Sun Village-Bell	34.6	33.1	32.3	28.2	23.8	60.3	21.3	33.9
WB TH at Sun Village-Bell Upstream of LT Bay	230.9	213.6	208.4	190.8	166.5	425.0	167.7	210.4
SB TH/RT at Sun Village-Bell	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
SB TH at Sun Village-Bell Upstream of LT Bay	1.9	1.9	1.9	1.9	2.1	2.0	1.9	1.8
NB LT at Litchfield-Bell	298.0	205.0	263.9	221.0	291.4	403.7	255.1	189.6
NB TH at Litchfield-Bell Upstream of LT Bay	426.3	220.2	508.0	302.3	505.6	763.0	628.0	224.7
WB TH at Litchfield-Bell	53.1	51.0	52.2	48.2	40.1	76.0	48.1	57.9
WB TH at Litchfield-Bell Upstream of LT Bay	364.2	339.1	352.4	333.7	281.8	497.2	401.5	394.0
SB TH at Litchfield-Bell	82.6	93.7	103.1	92.7	96.0	115.9	94.8	37.3
SB TH at Litchfield-Bell Upstream of LT Bay	35.6	58.6	111.8	30.1	28.9	109.1	38.1	1.8

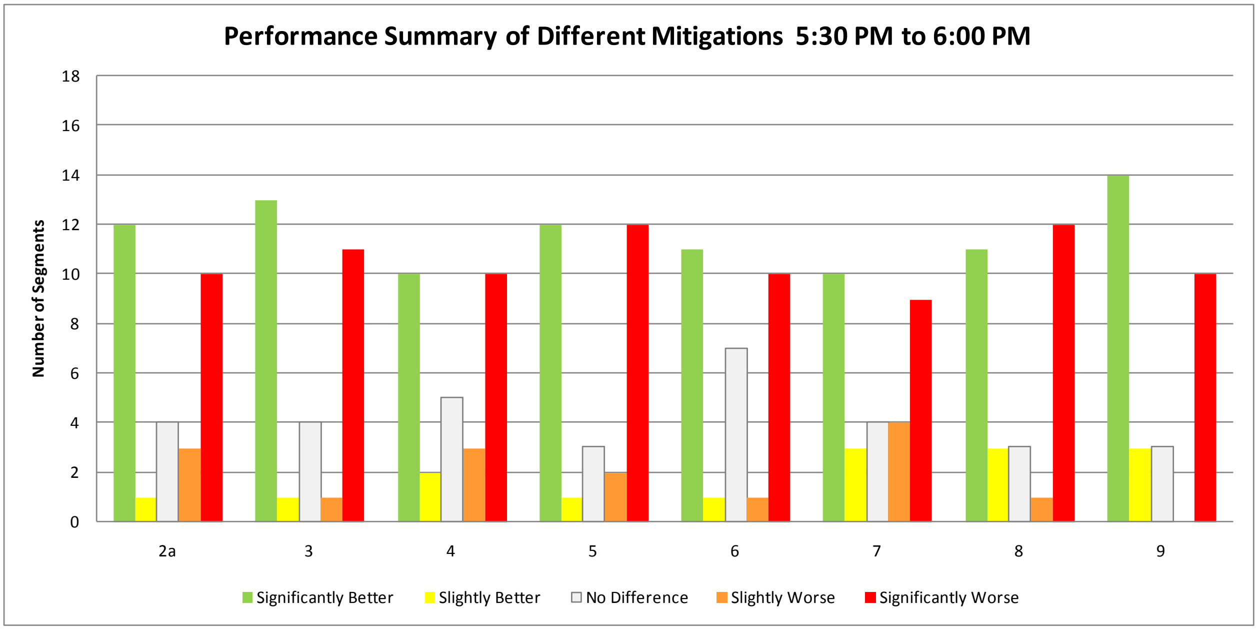


Figure 169. Performance Summary 5:30 – 6:00

Table 59. Average delay per link 6:00 – 6:30

Segment	2a	3	4	5	6	7	8	9
EB TH at Reems-Bell	89.3	80.9	93.3	116.6	99.8	29.9	73.7	85.6
EB TH at Reems-Bell Upstream of LT Bay	297.4	239.0	247.4	739.4	517.8	2.6	113.0	408.5
NB RT at Reems-Bell	601.2	569.3	651.0	792.2	672.6	218.5	426.9	597.7
NB TH at Reems-Bell Upstream of RT Bay	999.2	1194.8	1213.2	1510.3	1346.3	390.9	953.3	1043.4
SB LT at Reems-Bell	66.0	65.9	60.4	81.5	72.3	40.5	51.9	51.3
SB TH at Reems-Bell Upstream of LT Bay	111.4	83.1	81.9	353.1	210.9	44.2	66.8	33.2
EB TH at Parkview-Bell	48.0	46.3	51.4	71.4	56.8	29.5	43.7	49.0
EB TH at Parkview-Bell Upstream of LT Bay	745.9	714.9	761.6	1038.0	839.9	424.0	652.3	738.8
NB RT at Parkview-Bell	4.2	3.8	4.3	4.2	4.5	4.9	4.4	3.8
NB TH at Parkview-Bell Upstream of LT Bay	0.2	3.4	0.3	0.2	0.2	0.2	0.8	0.3
SB LT at Parkview-Bell	72.9	76.1	80.7	99.5	76.7	51.8	72.2	29.9
SB TH at Parkview-Bell Upstream of LT Bay	38.4	43.8	112.5	228.3	140.0	7.6	76.9	1.4
EB TH at Bullard-Bell	27.5	31.1	30.6	24.5	43.5	13.6	16.0	16.6
EB RT at Bullard-Bell	24.0	23.9	23.8	29.6	27.5	19.3	24.0	24.0
EB TH at Bullard-Bell Upstream of RT Bay	133.3	139.7	151.6	199.1	165.2	87.0	126.3	133.6
WB LT at Bullard-Bell	16.3	15.3	15.9	14.6	13.1	26.2	19.5	17.8
WB TH at Bullard-Bell	2.1	2.3	2.3	3.6	2.8	2.2	2.5	2.8
WB TH at Bullard-Bell Upstream of LT Bay	166.7	168.7	174.3	168.6	150.8	312.4	119.4	189.6
NB LT at Sun Village-Bell	316.6	131.5	272.9	222.3	315.8	125.6	102.0	63.1
NB TH at Sun Village-Bell Upstream of LT Bay	219.8	2.3	105.6	53.7	350.2	22.2	1.6	1.4
WB TH at Sun Village-Bell	19.4	20.5	21.1	20.7	18.2	40.1	15.7	21.6
WB TH at Sun Village-Bell Upstream of LT Bay	164.6	169.7	165.5	163.9	140.7	372.8	133.9	175.9
SB TH/RT at Sun Village-Bell	6.2	4.7	6.7	6.1	9.6	6.6	6.6	3.4
SB TH at Sun Village-Bell Upstream of LT Bay	1.8	1.8	1.9	1.8	2.4	1.9	1.8	1.8
NB LT at Litchfield-Bell	210.1	162.1	233.5	185.3	235.6	355.4	221.9	169.4
NB TH at Litchfield-Bell Upstream of LT Bay	729.9	473.7	810.6	600.3	839.9	1058.8	788.1	520.7
WB TH at Litchfield-Bell	37.4	37.6	38.0	38.9	33.1	68.2	37.9	46.7
WB TH at Litchfield-Bell Upstream of LT Bay	271.1	273.4	273.1	281.7	220.5	569.4	306.0	340.5
SB TH at Litchfield-Bell	62.8	80.7	77.2	52.5	58.8	93.3	69.8	33.0
SB TH at Litchfield-Bell Upstream of LT Bay	31.7	65.8	165.0	1.3	2.5	182.1	42.1	0.8

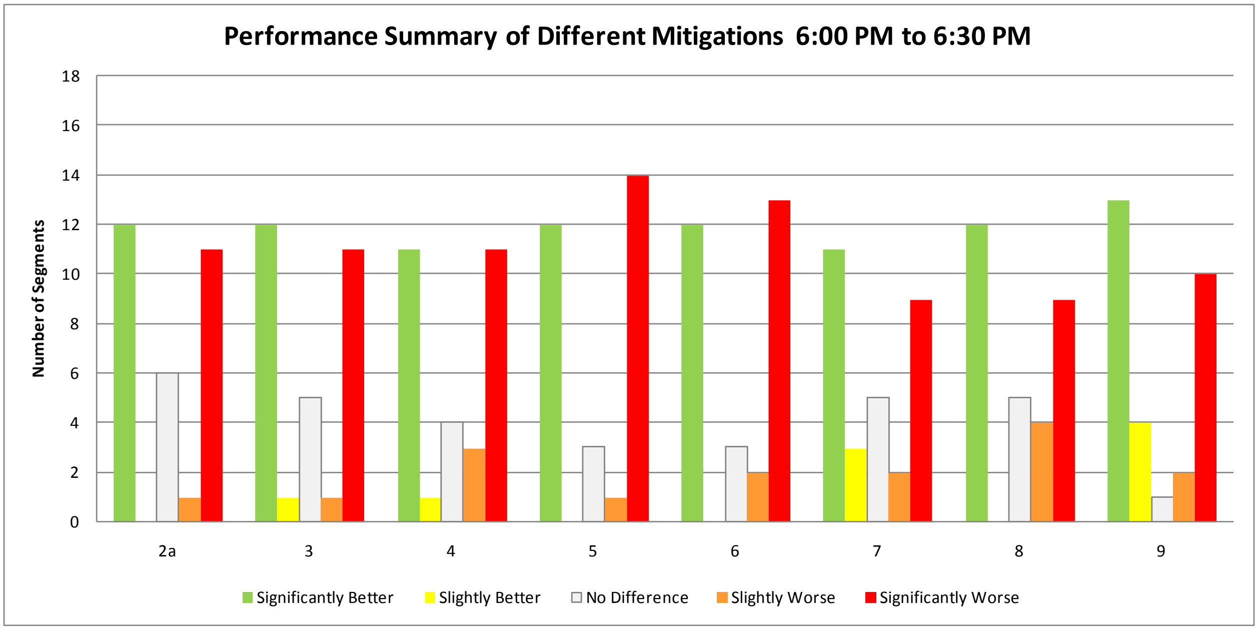


Figure 170. Performance summary 6:00 – 6:30

Table 60. Average delay per link 6:30 – 7:00

Segment	2a	3	4	5	6	7	8	9
EB TH at Reems-Bell	26.8	23.4	26.1	52.7	39.3	19.5	20.5	27.8
EB TH at Reems-Bell Upstream of LT Bay	21.8	2.6	19.5	581.7	192.0	0.5	0.9	42.3
NB RT at Reems-Bell	73.2	34.1	56.1	386.1	153.1	7.1	17.8	82.6
NB TH at Reems-Bell Upstream of RT Bay	652.1	591.4	597.7	1822.7	1122.1	2.2	279.9	733.4
SB LT at Reems-Bell	39.7	40.8	39.3	44.2	45.4	39.4	40.5	25.5
SB TH at Reems-Bell Upstream of LT Bay	3.9	1.6	1.8	43.5	2.8	0.6	0.7	0.5
EB TH at Parkview-Bell	13.6	13.1	15.5	23.3	17.2	12.4	13.2	20.7
EB TH at Parkview-Bell Upstream of LT Bay	273.7	269.4	280.9	545.0	366.2	57.1	192.9	316.0
NB RT at Parkview-Bell	3.3	4.2	3.8	4.6	3.7	4.3	3.6	3.1
NB TH at Parkview-Bell Upstream of LT Bay	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
SB LT at Parkview-Bell	39.3	35.5	35.4	35.1	40.9	32.6	34.4	23.2
SB TH at Parkview-Bell Upstream of LT Bay	3.6	0.6	0.8	1.9	0.9	0.6	0.6	0.5
EB TH at Bullard-Bell	9.9	11.9	9.9	8.6	5.4	10.8	6.3	7.7
EB RT at Bullard-Bell	12.7	11.8	12.8	17.8	12.6	9.2	11.2	15.3
EB TH at Bullard-Bell Upstream of RT Bay	26.0	29.0	27.7	52.4	31.4	8.7	17.8	28.6
WB LT at Bullard-Bell	11.3	7.5	11.1	12.5	12.4	15.4	12.4	13.3
WB TH at Bullard-Bell	1.0	1.0	2.4	1.8	2.2	2.4	2.3	4.0
WB TH at Bullard-Bell Upstream of LT Bay	13.9	12.2	14.4	8.3	3.8	79.4	8.4	23.2
NB LT at Sun Village-Bell	62.2	43.4	73.9	51.4	54.5	62.5	53.0	36.8
NB TH at Sun Village-Bell Upstream of LT Bay	1.2	1.2	1.3	1.2	8.7	1.3	1.2	1.2
WB TH at Sun Village-Bell	1.7	5.2	2.0	1.6	0.9	8.0	1.3	6.4
WB TH at Sun Village-Bell Upstream of LT Bay	8.8	11.0	6.4	2.3	1.0	73.5	1.2	11.1
SB TH/RT at Sun Village-Bell	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
SB TH at Sun Village-Bell Upstream of LT Bay	1.8	1.8	1.8	1.8	1.8	1.7	1.7	1.8
NB LT at Litchfield-Bell	63.2	62.7	64.0	66.1	66.1	105.9	68.9	44.2
NB TH at Litchfield-Bell Upstream of LT Bay	23.5	2.0	15.6	4.2	11.7	281.3	11.5	1.3
WB TH at Litchfield-Bell	18.7	18.6	17.8	18.7	17.4	22.6	18.1	20.5
WB TH at Litchfield-Bell Upstream of LT Bay	5.7	3.3	6.1	7.0	3.2	115.5	4.0	17.0
SB TH at Litchfield-Bell	42.0	43.7	61.1	40.9	42.3	66.0	44.8	24.3
SB TH at Litchfield-Bell Upstream of LT Bay	0.6	3.0	72.7	0.6	0.6	127.5	0.6	0.5

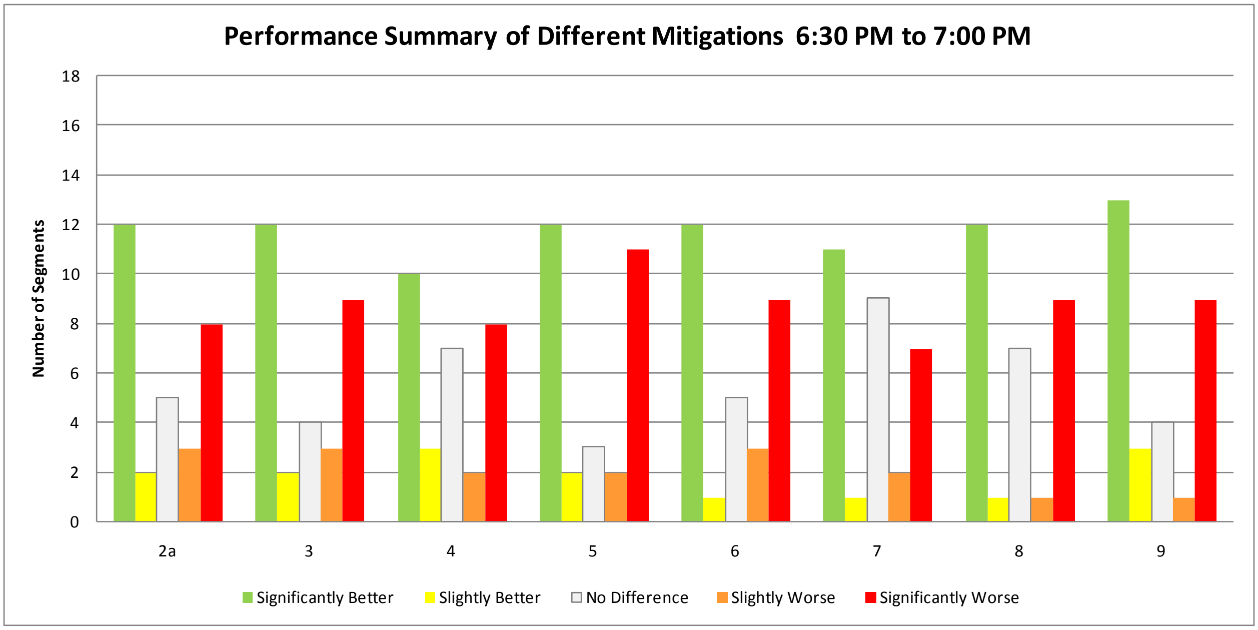


Figure 171. Performance summary 6:30 – 7:00

Table 61. Average delay per link (3 hour total)

Segment	2a	3	4	5	6	7	8	9
EB TH at Reems-Bell	37.7	36.3	38.1	48.6	42.1	24.7	33.6	39.9
EB TH at Reems-Bell Upstream of LT Bay	59.3	46.9	49.0	231.2	125.1	2.9	24.0	87.3
NB RT at Reems-Bell	64.5	65.2	66.3	97.9	79.7	40.6	55.5	67.1
NB TH at Reems-Bell Upstream of RT Bay	249.3	267.5	273.6	435.7	352.7	82.1	200.7	255.3
SB LT at Reems-Bell	50.3	49.7	48.9	54.1	52.9	46.2	48.2	37.5
SB TH at Reems-Bell Upstream of LT Bay	48.1	40.4	46.3	92.8	76.2	45.9	47.2	9.0
EB TH at Parkview-Bell	20.5	20.3	24.1	29.8	25.3	19.2	22.0	24.9
EB TH at Parkview-Bell Upstream of LT Bay	250.5	244.9	251.6	347.9	287.8	146.0	216.4	264.0
NB RT at Parkview-Bell	4.7	4.7	5.1	5.0	5.0	5.6	5.0	4.8
NB TH at Parkview-Bell Upstream of LT Bay	0.3	1.3	0.4	0.4	0.3	0.3	0.5	0.3
SB LT at Parkview-Bell	54.5	53.2	54.9	61.4	59.2	46.2	52.7	27.7
SB TH at Parkview-Bell Upstream of LT Bay	26.9	19.9	34.6	75.2	60.3	10.2	25.6	1.5
EB TH at Bullard-Bell	15.9	23.2	18.3	15.6	20.9	11.1	10.6	10.8
EB RT at Bullard-Bell	19.5	20.8	20.8	24.5	21.9	17.1	19.3	20.1
EB TH at Bullard-Bell Upstream of RT Bay	52.3	54.6	56.0	74.2	61.4	37.6	48.3	53.3
WB LT at Bullard-Bell	16.3	14.6	15.5	14.7	12.9	22.6	20.4	17.8
WB TH at Bullard-Bell	1.4	1.6	2.5	2.5	2.6	2.3	2.4	2.8
WB TH at Bullard-Bell Upstream of LT Bay	86.4	82.3	83.7	81.1	70.3	138.7	65.3	86.8
NB LT at Sun Village-Bell	186.4	107.0	166.2	146.6	181.8	116.7	111.2	73.7
NB TH at Sun Village-Bell Upstream of LT Bay	89.0	7.4	44.7	25.3	103.0	22.3	4.2	2.9
WB TH at Sun Village-Bell	10.4	11.6	10.8	10.2	9.0	17.5	9.0	12.7
WB TH at Sun Village-Bell Upstream of LT Bay	71.6	72.1	68.1	65.3	58.8	127.2	59.2	68.7
SB TH/RT at Sun Village-Bell	57.0	52.3	50.6	47.4	73.8	50.4	50.7	49.3
SB TH at Sun Village-Bell Upstream of LT Bay	1.8	1.8	1.8	1.8	2.0	1.9	1.8	1.8
NB LT at Litchfield-Bell	115.7	103.0	119.7	109.7	122.0	148.4	124.8	94.6
NB TH at Litchfield-Bell Upstream of LT Bay	174.9	115.3	201.2	143.9	199.5	298.5	232.6	123.3
WB TH at Litchfield-Bell	27.1	26.7	26.8	26.6	24.4	32.7	26.6	31.6
WB TH at Litchfield-Bell Upstream of LT Bay	124.1	121.4	122.9	122.4	109.4	177.8	134.5	157.1
SB TH at Litchfield-Bell	61.8	67.4	71.2	64.5	67.2	78.4	68.1	34.7
SB TH at Litchfield-Bell Upstream of LT Bay	16.4	27.8	63.6	11.6	11.7	73.4	20.4	1.2

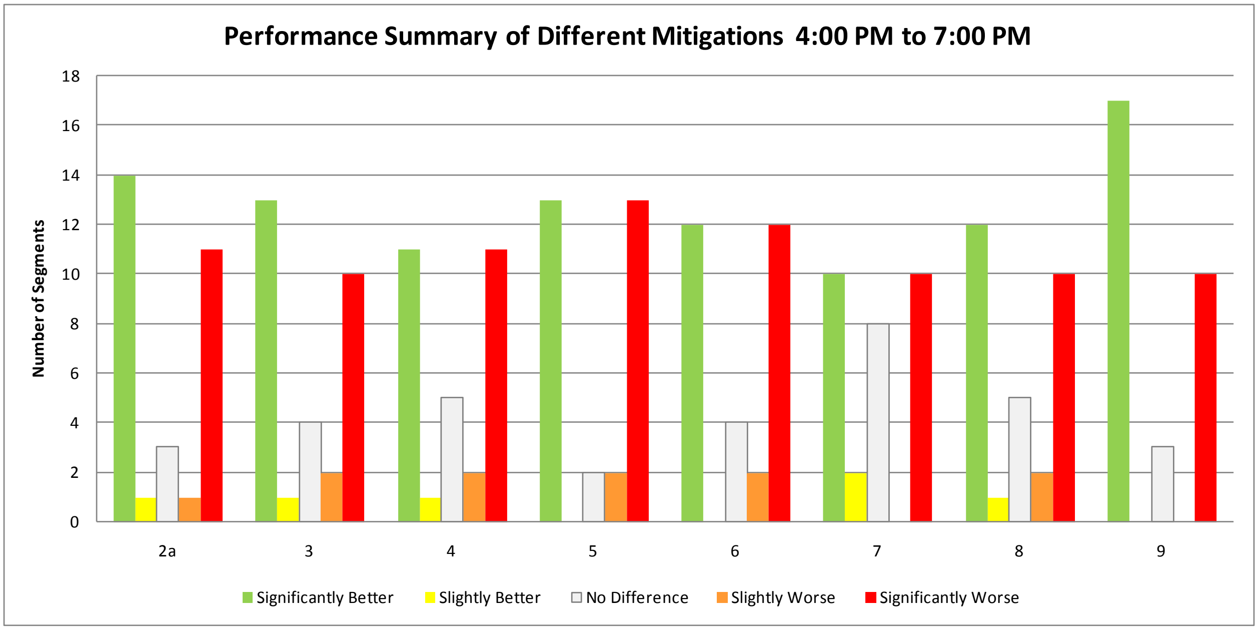


Figure 172. Performance summary (3 hour total)

As shown in Figure 167 through Figure 171, each of the mitigation scenarios performed similarly in terms of average delay per link. Because the logical first mitigation strategy that would be implemented by a traffic engineer facing this problem would be to extend the westbound left-turn phase at Bullard, the results of each mitigation strategy was also compared to the results of the extended left-turn split strategy. The results of this comparison are shown in Figure 173 and Table 62.

Table 62. Average delay comparison with extended left turn at Bullard (3 hour total)

Segment	3	4	5	6	7	8	9
EB TH at Reems-Bell	36.3	38.1	48.6	42.1	24.7	33.6	39.9
EB TH at Reems-Bell Upstream of LT Bay	46.9	49.0	231.2	125.1	2.9	24.0	87.3
NB RT at Reems-Bell	65.2	66.3	97.9	79.7	40.6	55.5	67.1
NB TH at Reems-Bell Upstream of RT Bay	267.5	273.6	435.7	352.7	82.1	200.7	255.3
SB LT at Reems-Bell	49.7	48.9	54.1	52.9	46.2	48.2	37.5
SB TH at Reems-Bell Upstream of LT Bay	40.4	46.3	92.8	76.2	45.9	47.2	9.0
EB TH at Parkview-Bell	20.3	24.1	29.8	25.3	19.2	22.0	24.9
EB TH at Parkview-Bell Upstream of LT Bay	244.9	251.6	347.9	287.8	146.0	216.4	264.0
NB RT at Parkview-Bell	4.7	5.1	5.0	5.0	5.6	5.0	4.8
NB TH at Parkview-Bell Upstream of LT Bay	1.3	0.4	0.4	0.3	0.3	0.5	0.3
SB LT at Parkview-Bell	53.2	54.9	61.4	59.2	46.2	52.7	27.7
SB TH at Parkview-Bell Upstream of LT Bay	19.9	34.6	75.2	60.3	10.2	25.6	1.5
EB TH at Bullard-Bell	23.2	18.3	15.6	20.9	11.1	10.6	10.8
EB RT at Bullard-Bell	20.8	20.8	24.5	21.9	17.1	19.3	20.1
EB TH at Bullard-Bell Upstream of RT Bay	54.6	56.0	74.2	61.4	37.6	48.3	53.3
WB LT at Bullard-Bell	14.6	15.5	14.7	12.9	22.6	20.4	17.8
WB TH at Bullard-Bell	1.6	2.5	2.5	2.6	2.3	2.4	2.8
WB TH at Bullard-Bell Upstream of LT Bay	82.3	83.7	81.1	70.3	138.7	65.3	86.8
NB LT at Sun Village-Bell	107.0	166.2	146.6	181.8	116.7	111.2	73.7
NB TH at Sun Village-Bell Upstream of LT Bay	7.4	44.7	25.3	103.0	22.3	4.2	2.9
WB TH at Sun Village-Bell	11.6	10.8	10.2	9.0	17.5	9.0	12.7
WB TH at Sun Village-Bell Upstream of LT Bay	72.1	68.1	65.3	58.8	127.2	59.2	68.7
SB TH/RT at Sun Village-Bell	52.3	50.6	47.4	73.8	50.4	50.7	49.3
SB TH at Sun Village-Bell Upstream of LT Bay	1.8	1.8	1.8	2.0	1.9	1.8	1.8
NB LT at Litchfield-Bell	103.0	119.7	109.7	122.0	148.4	124.8	94.6
NB TH at Litchfield-Bell Upstream of LT Bay	115.3	201.2	143.9	199.5	298.5	232.6	123.3
WB TH at Litchfield-Bell	26.7	26.8	26.6	24.4	32.7	26.6	31.6
WB TH at Litchfield-Bell Upstream of LT Bay	121.4	122.9	122.4	109.4	177.8	134.5	157.1
SB TH at Litchfield-Bell	67.4	71.2	64.5	67.2	78.4	68.1	34.7
SB TH at Litchfield-Bell Upstream of LT Bay	27.8	63.6	11.6	11.7	73.4	20.4	1.2

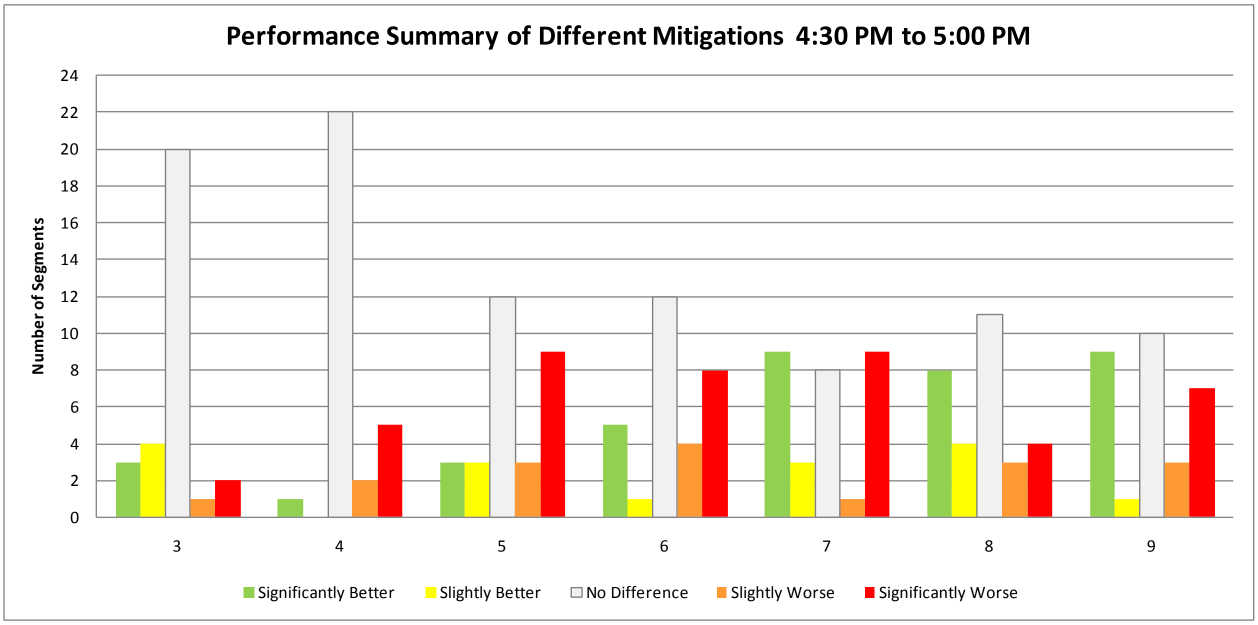


Figure 173. Performance summary comparison to extended left-turn split at Bullard (3 hour total)

The results of this comparison illustrate the various mitigation strategies do not produce significantly different results. The mitigations which do reduce average delay on some links do so at the expense of other links which experience increased delay.

Throughput Analysis

The number of vehicles in the system was calculated by comparing the vehicle input and output data recorded by the simulation. The average input rates for each mitigation strategy as well as the Vissim Demand input are shown in Figure 174. This figure illustrates that the arterial is so congested that vehicles cannot enter the system at the rate that the model is demanding. Each mitigation strategy tested results in input rates which are higher than the no mitigation scenario indicating that each mitigation strategy allows for more vehicles to enter the system.

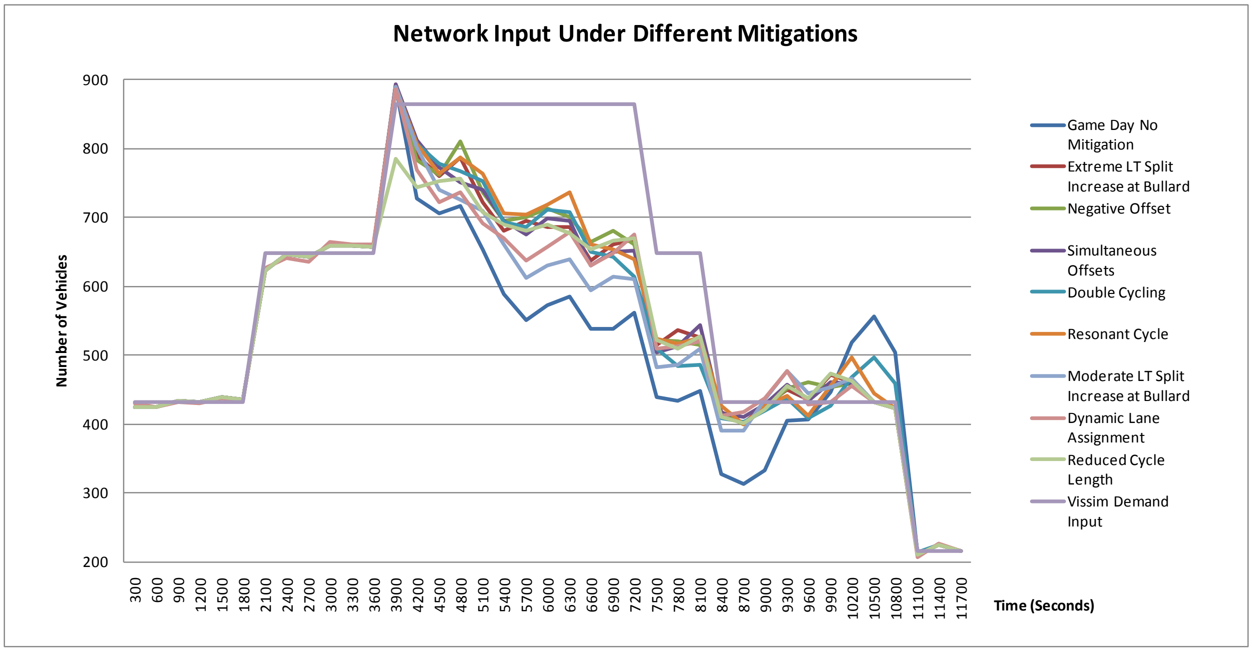


Figure 174. Average input rates under different mitigations

Similarly, the network output graph shown in Figure 175 illustrates that each mitigation strategy improves the baseline condition. The greatest impact of the mitigation strategies can be seen during the "recover" portion of the curve which occurs after the peak hour at approximately 7800s into the simulation. Without any mitigation, the output continues to decrease until the input volumes decrease to 75% of the peak hour volume while each mitigation strategy causes the output to increase and return to steady state sooner than the baseline condition. The Dynamic Lane Assignment strategy resulted in the highest output rate and also yields the earliest recovery.

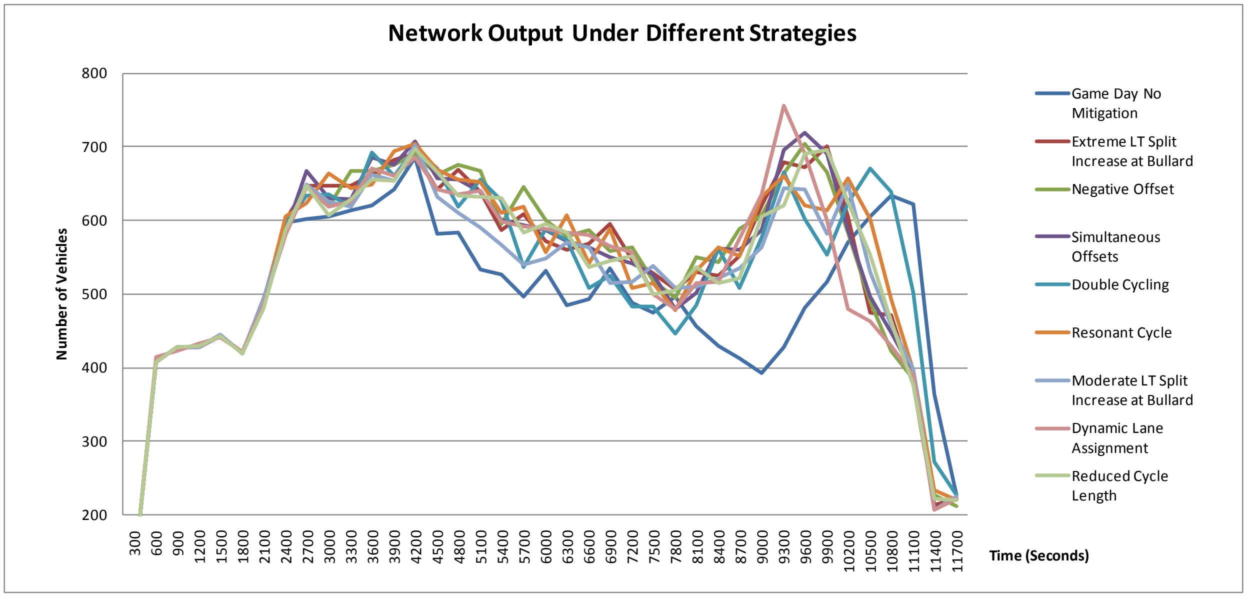


Figure 175. Average output rates under different mitigations

The average number of vehicles in the system was calculated using the input and output data shown above. Figure 176 shows the resulting number of vehicles in the system for each strategy.

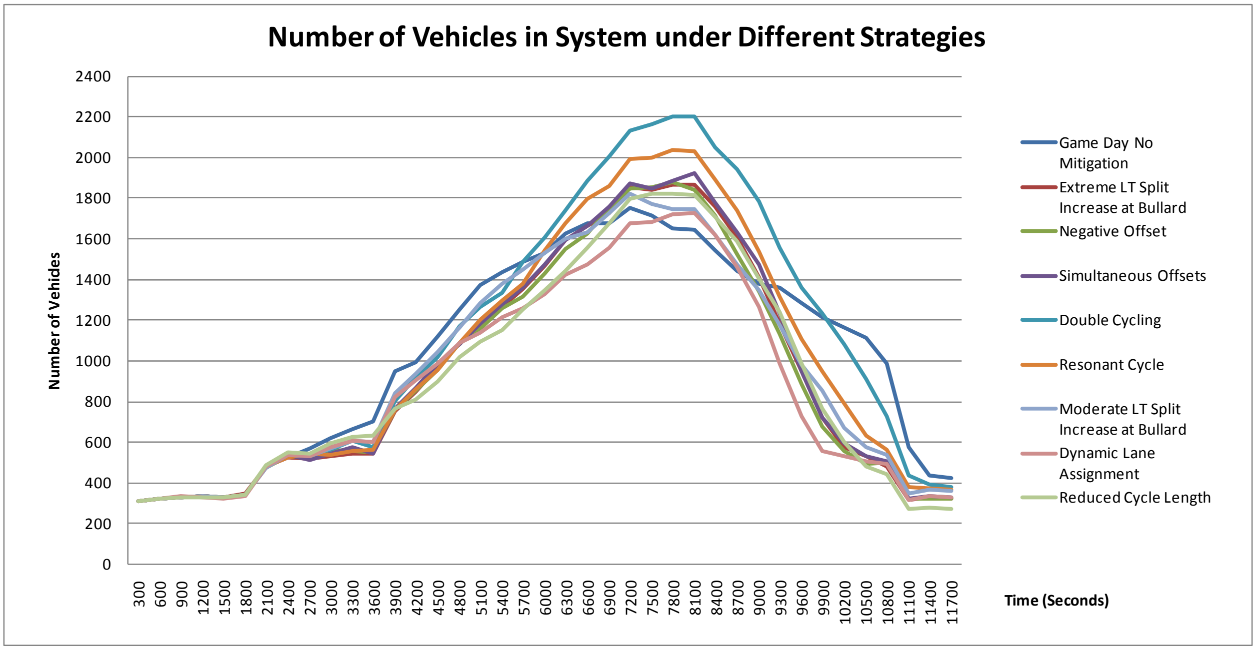


Figure 176. Average vehicles in the system

Determining the 'best' strategy from this presentation of data depends on the desired outcome of the strategy. For example, if the objective is to reduce the number of vehicles 'stuck' in the system, one would choose a strategy which results in a data line which falls below the baseline. However, if the objective is to utilize the storage space within the system, choosing a mitigation which falls above the baseline date would be appropriate. The Double Cycling mitigation strategy appears to result in the highest number of vehicles in the system but Figure 177 also indicates that this strategy take the longest to recovery. The Dynamic Lane mitigation strategy keeps the number of vehicles in the system lower during the loading portion of the simulation and appears to recover sooner than any other strategy.

Travel Time Analysis

Strategies for mitigating the Bell Road corridor under game conditions was also analyzed in terms of average travel time through the network. The results of this analysis, shown in Figure 178, are similar to the delay analysis which implied that strategies that improve the performance for the westbound direction of travel, will negatively impact other directions of travel.

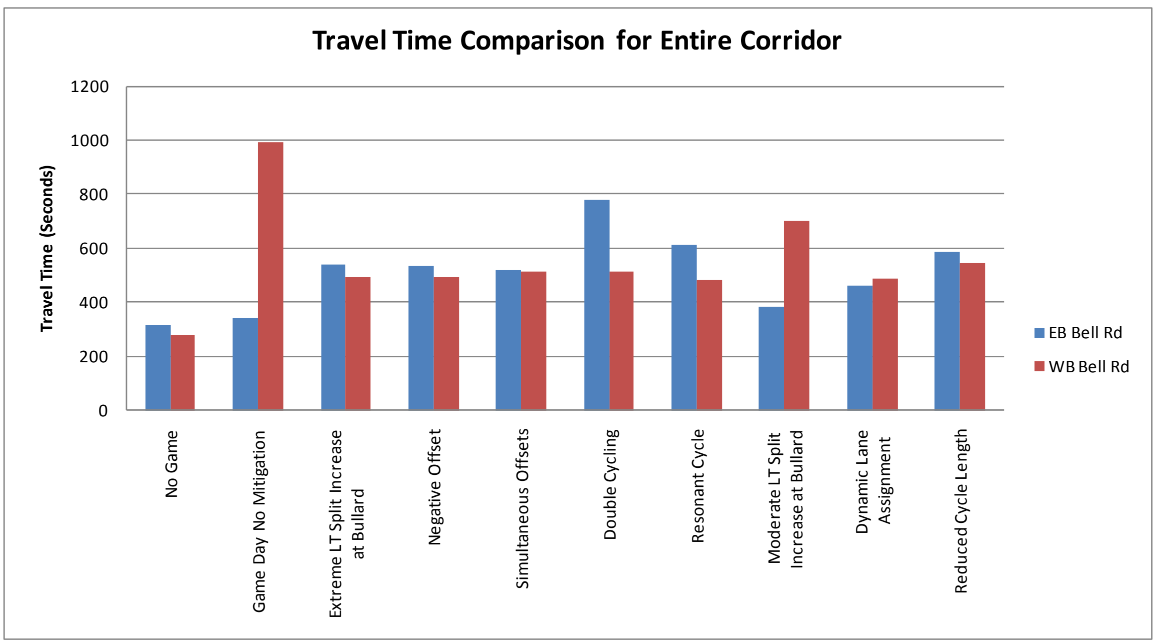


Figure 177. Average travel time under different mitigation strategies

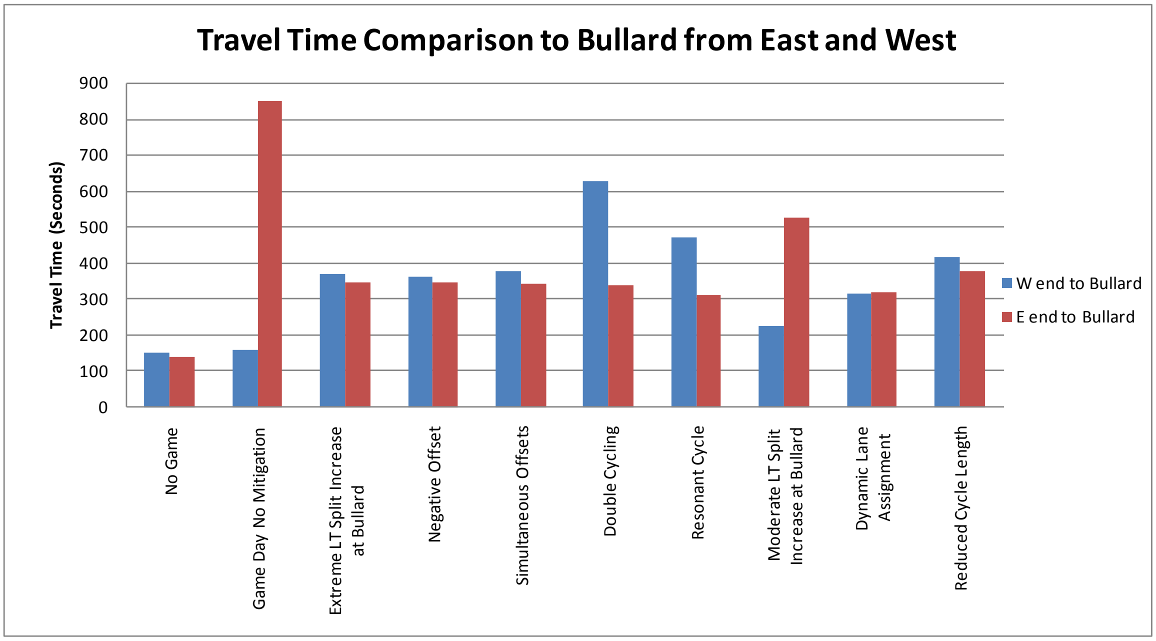


Figure 178. Travel time comparison to Bullard from eastbound and westbound directions

Summary

In this test case, we applied the guidance process to a real-world situation with event traffic overlaid on normal heavy P.M. peak flows. This situation, unmitigated, produces extensive queuing which increases the travel time on the arterial by 400%. A number of different mitigation strategies were applied, working up from the basic mitigation to increase the left-turn split time at the critical intersection. All of the mitigations were found to be effective in reducing the westbound travel time. Some detriment to eastbound travel time resulted from the improvements in the westbound direction. The largest effects of the mitigations were observed during the recovery period.

Chapter 4: Conclusions and Future Directions

In this project, Dr. Douglas Gettman at Kimley-Horn and Associates (KHA), Dr. Henry Liu at the University of Minnesota, and Dr. Montasir Abbas at Virginia Tech performed experimental research on the mitigation of oversaturated traffic conditions on arterials and networks with the support of a number of graduate students and staff. Dr. Alex Skabardonis provided advisory support. This research was divided into five experimental areas:

- Diagnosis and quantification of the type and cause of oversaturated conditions
- Development of a methodology for testing and evaluating the relative performance of mitigation strategies on specific scenarios
- Development of tools to relate diagnosis of conditions with mitigation strategies
- Testing of mitigation strategies on real-world scenarios
- Distillation of the findings into a rational guide for practitioners to diagnosis and select the appropriate mitigation strategies

The research focused on identifying traffic control strategies that can be implemented by traffic signal systems using traditional actuated-coordinated fixed-parameter signal timing plans, time-of-day scheduling, or traffic-responsive plan selection methods.

Definitions and Diagnosis

A substantial set of definitions was identified during the project to develop a taxonomy for describing oversaturated scenarios. This taxonomy includes spatial extent (approach, route, network), duration (intermittent, persistent, pervasive), causation (demand, incidents, timings, etc.), recurrence, and symptoms (storage blocking, starvation, etc.). The basic definition of an oversaturated condition starts from the presence of an overflow queue on a traffic movement after the termination of the green-time for that movement. From this basic concept, higher level definitions were developed for approaches, phases, routes, arterials, and networks.

To diagnose the severity of an oversaturated condition a methodology was developed to estimate queue length from high resolution second-by-second occupancy data from advance detectors and second-by-second phase timing data. This methodology allows measurement of queues that grow substantially upstream of the detector location. In addition, two quantitative measures of oversaturation intensity were developed. These measures (TOSI and SOSI) quantify the relationship between the length of the residual queue at an intersection approach or movement with the available green-time. SOSI measures how much green-time is wasted when vehicles cannot move due to downstream blockage and TOSI measures how much green-time is spent dissipating overflow queuing from the previous cycle. Oversaturated routes, arterials, intersections, networks, and so on are then defined as having TOSI and/or SOSI > 0 on the constituent approaches and movements at the same time. The characteristics of how TOSI and SOSI change over time were described in Task 2 in field tests in Minneapolis on TH55.

Development of Management Objectives and Characterizing Oversaturated Conditions Scenarios

In Task 3, we defined three broad operational objectives of signal timing strategies:

1. Minimizing (user) delay
2. Maximizing throughput
3. Managing queues

The minimize delay objective drives strategies and operational principles that assume undersaturated operation and encapsulates all objectives that might be considered to be effective during undersaturated operation. This includes (a) minimizing the delay at a single intersection and (b) minimizing delay in a network or series of intersections on a travel route.

Minimizing user delay is not appropriate when the situation is oversaturated, particularly since it is no longer possible to avoid phase failures. However, minimizing total vehicle delay was found to be an acceptable objective during the loading regime of an oversaturated scenario. In some situations, maximizing throughput was also appropriate during the loading regime.

An objective of maximizing throughput de-emphasizes the individual user's experience and focuses more heavily on the performance of the system as a whole. Timing strategies designed with this objective in mind tend to punish lighter movements to the benefit of the greater (system) good. This is done by moving heavier phases for longer amounts of time more frequently than by the typical cycle-failure minimizing methods of actuated control.

Strategies that maximize throughput can be categorized as:

- Strategies that make best use of the physical space (e.g., lag heavy left turns; run closely-spaced intersections on single controller).
- Strategies that make best use of green-time in the cycle (e.g., prevent actuated short greens, separate congested movements from the uncongested ones, phase reservice).
- Strategies that reduce the negative impact of other influences (e.g., bus and pedestrian movements) on the overall ability of the signal system to process vehicle flows.

Throughput maximization strategies have the goal of either increasing input to a system that has spare storage capacity, increase output from a system that is severely limited in storage capacity, or both. At some point, however, no further revision to the signal timing will increase maximum throughput and queues will continue to grow until demand diminishes.

The queue management objective applies when congestion is pervasive to the extent that additional green-time on a critical route will make the situation worse. Strategies that minimize total delay and maximize throughput were found to have similar performance in situations where queue management is a more appropriate objective. The only choice in these situations becomes the arrangement of the operation of signals within a network to prevent the queues from multiplying the problem. Traditional signal timing parameters (cycle, split, offset, etc.) are not particularly suited for operation during situations where queue management is the primary goal. Synchronizing the actions of multiple controllers in a system of intersections for the purpose of queue management is very difficult within the context of actuated-coordinated control by commanding patterns with different parameters. In the scenarios where signal timing plans were optimized for different objectives and were compared against each other, we approximated queue management strategies with timing plans (cycle, split, offset, sequence, etc.) that minimized the total degree of saturation on the critical routes.

In tandem with these three operational objectives, we determined that there are three operating regimes during an oversaturated conditions scenario:

Loading

1. Processing
2. Recovery

During the loading regime, traffic volumes are increasing, route proportions are changing, and in the case of non-recurrent events, the triggering event has started. During loading, overflow queuing and other symptoms such as storage blocking and starvation begin to emerge. Early application of mitigation strategies can delay the onset of debilitating queuing. Early application of mitigation strategies is easier to conceptualize when the causal factors are recurrent.

During the ***processing*** regime, the traffic volumes and route proportions are such that queues and congestion are not going to dissipate 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. This is the operational situation that many practitioners might characterize as there is nothing that can be done. Queue management strategies can be applied during this regime to help the system return to steady-state operation sooner than continuing to apply the normal operational strategies designed for undersaturated conditions. Queue management strategies might also be able to move more traffic that are not on the critical routes, since allocation of more green-time to the critical routes might actually make the problem worse.

During the ***recovery*** regime, traffic volumes and/or route proportions and/or restrictive downstream capacity (e.g. clearance of crash, removal of construction cones, reduction in traffic flow, etc.) have been adjusted so that the overflow queues begin to dissipate. In this regime of operation, mitigation strategies were found to be especially effective in returning the system to steady state sooner than continuing to apply normal operational strategies that assume undersaturated operation. The test cases performed in this research indicate that the recovery regime is where the most substantial performance improvements can be achieved by applying mitigations.

In two of the test cases in Task 7, we explored the performance of timing plans designed for each of the operational objectives during each of the three operational regimes. It was found that there is significant value in applying different timing plans during the different regimes of a scenario.

Mitigation Strategies

We developed and tested mitigation strategies using three different methodologies. First, since little is published about traditional signal timing strategies for mitigating oversaturated conditions a research methodology was developed to compare the performance of traditional signal timing plans (cycle, splits, offsets, etc.) under different assumptions about the critical routes and operating regimes. This methodology compared mitigation strategies designed to maximize throughput, minimize delay, or manage queues against several realizations of critical route flows in two real-world test cases. This timing plan development framework and evaluation methodology extended previous work by Akcelik, Abu-Lebdeh, Lieberman, and Rathi. The principles by which the green-times, offsets, and objective functions were determined were extended from the principles developed by these previous works to address the inclusion of critical routes within the network and to comprehensively compare the mitigations on multiple objectives.

The performance of each mitigation strategy was compared for both delay and throughput measures in a Pareto analysis to identify non-dominated strategies during each time period of a scenario. Typically it was found that no specific mitigation is optimal for both minimizing delay and maximizing throughput, and certainly no strategy was uniformly dominant during each time period of a scenario with loading, processing, and recovery regimes.

The Pareto front analysis also revealed the importance of identifying critical routes through a network of intersections. Mathematical approaches akin to O-D estimation were explored, but are not detailed in the final report. Significant differences in performance were observed for differing definitions of the critical routes in a complex scenario. Since (demand) volumes are not easily measurable during oversaturation, identification of critical routes was described as an ad-hoc inspection process in the resulting practitioner guidance. Additional research is needed in the area of critical route identification, particularly to bring together the concepts of measurement of TOSI and SOSI with traditional O-D estimation techniques.

In addition to this structured and mathematically-based development methodology, we tested a variety of potential mitigations including phase reservice, negative offsets, cycle time adjustment, left-turn treatment, phase sequence, dynamic lane allocation, and phase truncation using engineering judgment. After identifying the critical routes for a specific scenario, combinations of mitigations were tested on four different simulation test cases. All mitigation strategies were implemented as either scheduled timing plan changes or based on the detection of certain oversaturated conditions using the if…then software tool developed as part of this project. Detection of oversaturated conditions was tested with both traditional measures (detector occupancy) and the derived measures we developed in this project, TOSI and SOSI. This if…then tool is available from NCHRP for universal use.

Finally, in addition to the multi-objective strategy development framework and the variety of investigations based on engineering judgment, we also developed a methodology for deriving green-time re-allocation directly from the quantitative TOSI and SOSI measures that were developed in Task 2. After measurement of TOSI and SOSI either with mathematical or observational methods, a two-pass heuristic algorithm was developed to calculate increases and decreases to green-time to drive TOSI and SOSI on an oversaturated route to as close to zero as possible. The potential effectiveness of this two-pass analytical procedure was demonstrated in Task 7. Iterative application (i.e. adaptive control for oversaturated route management) of the procedure seems a reasonable next step in extending the research in this direction.

Practitioner Guidance

In Task 6, we developed guidance for practitioners to identify mitigation strategies that apply to various oversaturated conditions. The guidance follows a systems engineering approach to problem resolution, starting with problem characterization. In the initial steps of this process, the goal is to answer several basic questions:

- How many intersections and directions of travel are affected? (**Spatial extent**)
- How long does the oversaturated condition last? How does it evolve over time? How does it dissipate during recovery? (**Temporal extent**)
- How frequently does the oversaturated condition occur? (**Recurrence**)
- What is the cause or causes of this oversaturated condition? (**Causes and Symptoms**)

In subsequent steps, it is recommended that the practitioner identify the objectives and approximate regimes of operation such as the duration of the loading, processing, and recovery regimes by generating a dynamic map of how the queues grow and dissipate through the scenario. The dynamic map also serves to identify the critical routes through the network. The guidance provides a litany of mitigation strategies and identifies how each strategy applies to various oversaturated scenarios. Rules of thumb and design principles are provided for some of the mitigations, in particular identifying where measurements of TOSI and SOSI can be used to identify appropriate actions. The guide concludes by specifying a generic systems-engineering-based process for deploying methodologies and evaluating their relative effectiveness. Requirements for on-line application of strategies using the logic tool developed in Task 4 are provided in the guide. Requirements for central system features, field controller features, additional detector stations, and other communications or field equipment for each type of mitigation are also detailed in the guidance.

Significant future research and development will be needed to extend the existing guidance into a comprehensive tool for quantitatively determining mitigations in a cookbook fashion.

Test Applications

In Task 7, we applied methodologies developed in this project to six test networks. Two of the test networks (Reston Parkway in Herndon, VA and the Post Oak area in Houston, TX) were used in the development and testing of the methodology for developing mitigating strategies and testing those strategies using a multi-objective Pareto analysis. In the first test case (Reston Parkway), we considered the application of a single signal timing strategies for an entire oversaturated scenario. In the second test case, we explicitly considered the three regimes of the scenario in applying a sequence of three signal timing plans during the three operational regimes. Two other networks (TH55 in Minneapolis, MN and downtown Pasadena, CA) were used in the development and testing of strategies directly related to TOSI and SOSI. TH55 was used to prove and refine the concepts of TOSI and SOSI and to test the forward-backward procedure (FBP) in a relatively simple situation. The Pasadena, CA downtown network was used for developing and testing the FBP in a more stressing and complicated routing scenario. . Finally, two other test cases (an arterial in Surprise, AZ and a small network in Windsor, ON) were used to test a variety of mitigation strategies using engineering judgment and to apply the guidance methodology developed in Task 6. The Windsor, ON network was also used to demonstrate the application of the if…then on-line mitigation strategy selection tool.

All of the test applications were done in simulated using Vissim with either the RBC or the Virtual D4 traffic controller. While route proportions and demand flows were changed over time, no dynamic traffic assignment was used, i.e. vehicles in the simulation did not react to the congestion conditions to change their route, change their destination, or forgo travel.

Test Cases for the Multi-Objective Pareto Analysis

Two scenarios were used to develop and test the strategy development methodology. These test cases focused on recurrent, daily congestion and oversaturation on an arterial and then in a more complicated network. The first test case applied a single mitigation timing plan to the entire oversaturated scenario and the second test case applied a sequence of three timing plans during the scenario to address the loading, processing, and recovery regimes. This methodology for developing mitigation strategies focuses on the identification of critical routes in a network.

From the identification of these critical routes and the approximation of the arrival demands on the routes, an optimization problem is solved to obtain a range of feasible cycle, split, offset values that meet an operational objective (minimize delay, maximize throughput, or manage queues). These solutions were then evaluated in Vissim to determine the differences in the performance of each combination of mitigation strategies for both delay and throughput measures. Non-dominated solutions were then identified using Pareto analysis. Different mitigation strategies are non-dominated at different times during the scenario and tend to result in clustering of similar mitigations during the three regimes of operation.

The first test case analyzed an oversaturated scenario on Reston Parkway in Herndon, VA. This scenario is an arterial that intersects with the heavily traveled Dulles Toll Road. Combinations of cycle, splits, and offsets designed for operation in oversaturated conditions were tested. In this test case, the loading, processing, and recovery regimes were not explicitly considered in the timing plan design process. Only one timing plan was applied during the entire duration of the scenario.

The timing plans were then combined with either upstream metering on the critical route or with phase reservice for the northbound left turn at the critical interchange. In both cases it was found, in general, that short cycle lengths (e.g. 100s) with close to simultaneous offsets would minimize total system delay. Medium-length cycle times (e.g. 140s) were found to maximize throughput. Strategies that were optimized to maximize throughput and combined with upstream metering generally decreased total delay by 20% and increased total throughput by 15%. Strategies that were optimized to minimize total delay and combined with upstream metering could reduce delay by up to 40%, but increased throughput by only 7%. Strategies that were focused on maximizing throughput and combined with phase reservice decreased total delay by 27% and increased total throughput by 22%. Strategies that were focused on minimizing total delay with phase reservice could reduce delay by up to 63%, but increased throughput by only 5%. This test case illustrated the importance of considering different objectives and focusing signal timing plan development on mitigating conditions for a given or assumed set of critical routes.

In the second test case in the Post Oak area of Houston, TX, the three regimes of operation were explicitly considered in the timing plan development process. The same Pareto front evaluation procedure was applied to determine the differences in performance of each combination of mitigation strategies under different assumptions about the critical routes through the network. Two combinations of critical routes were evaluated. The first considered the combination of critical routes that pass through and into the network area. The second considered the combination of critical routes that result from travelers inside of the network area that are departing the area from the numerous parking facilities inside the network. Since this test case was the most complex and large-scale scenario of those that were tested in the project, each combination of critical routes was overlaid with each other to identify the critical movements and approaches throughout the network. Timing plans were then designed with consideration of these critical movements and considered a sequence of timing plan changes at the beginning of the processing regime and then at the beginning of the recovery regime. The timing plan strategies considered green flaring, phase reservice, negative and simultaneous offsets, and harmonic (2:1 and 3:1) cycling at many of the intersections in the network.

The strategy development process directly considered development of timing plans that were focused on one of the three regimes of the scenario, loading (minimize delay or maximize throughput), processing (manage queues), or recovery (maximize throughput). A pre-processing step to identify the timing plan among all of the plans that were developed that performed best during each regime of operation for either minimizing delay or maximizing throughput. This optimization process also identified the recommended points in time for switching between the three timing plans during the scenario. Notably, the switching points for the (maximize throughput→manage queues→maximize throughput) strategy and the (minimize delay→manage queues→maximize throughput) strategy were different times.

These two sequences of timing plans were then tested in the simulation and compared with the baseline operation of the network. The baseline operation with a single timing plan with a common cycle for all intersections throughout the network. In both strategies, the cycle time of the mitigations was reduced during the processing regime (from 150s to 100s or 90s) and then slightly increased during the recovery regime (from 150s to 160s). In addition, approximately half of the intersections in the network were double-cycled during the processing regime (80s or 75s cycle time) in order to manage the growth and interaction of long queues on the short network links in the interior of the network.

Both strategies were found to provide modest 5-10% improvements over the baseline strategy for total delay. For total stops and average stops per vehicle, the (minimize delay→manage queues→maximize throughput) strategy combination showed small detriments and the (maximize throughput→manage queues→maximize throughput) strategy improved both measures over the baseline. For total throughput, the (minimize delay→manage queues→maximize throughput) strategy produced improved throughput on approximately 1/3 of the intersections and decreased throughput on the remaining intersections when compared to the baseline. The (maximize throughput→manage queues→maximize throughput) strategy produced improved throughput on approximately 2/3 of the intersections and decreased throughput on the remaining intersections when compared to the baseline. The detriments to throughput at the intersections with reduced performance were less significant than the detriments produced by the (minimize delay→manage queues→maximize throughput) strategy. Those locations with throughput reductions were typically at the locations where SOSI was non-zero (portions of the green-time were wasted because no vehicles could move). In addition, significant throughput improvements were found at the intersections that were double-cycled.

Test Cases for Direct Application of TOSI and SOSI to Re-allocate Green-Time

In the test applications for direct application of the TOSI and SOSI measures, we found significant improvements were possible. In the TH55 in Minneapolis, MN application, the TOSI and SOSI values were used to identify offset and green-time re-allocation recommendations to improve the throughput performance. This test case could represent both recurrent and/or non-recurrent oversaturation. In this case, which had one oversaturated approach, it was demonstrated that direct measurement of TOSI and SOSI and adjustment of the green-times on the arterial using the FBP could drive the resulting TOSI and SOSI measurements to zero. In the test case using the Pasadena, CA downtown network, two scenarios were tested. The first was a single oversaturated route in one direction on an arterial. The FBP was developed and applied using the average TOSI and SOSI values measured along the route in the do nothing condition. The resulting adjustments to the green splits along the oversaturated route resulted in a 30% improvement to the throughput along the route. Notably because of the downstream blocking conditions along portions of the route, some of the green-time splits were actually reduced, but the throughput performance along the route was still increased.

In the second test, two intersecting oversaturated routes were generated (one southbound and one westbound). The same FBP was applied to calculate the green-time adjustments along both routes. In this test, the average throughput for southbound route showed no appreciable improvement but the westbound route was improved by 10%. In both of the test cases, modifications to cycle time, phase sequence, and other mitigation approaches were not considered.

Both of these test cases indicated that there is promise in directly considering the quantitative measures of oversaturation intensity in the re-computation of green-time allocation. More research and development will be required to formulate more comprehensive analytical procedures that integrate this approach with the consideration of the loading, processing, and recovery phases of a specific oversaturated scenario; essentially solving for the green-time re-allocation as a rolling-horizon adaptive control problem.

Test Cases for the Application of the Practitioner Guidance and On-line Evaluation of Mitigations

Two additional test applications focused on following the process defined in the practitioner guidance. The first test case in Windsor, ON, evaluated mitigating strategies for handling oversaturated conditions on two critical routes (eastbound and northbound) competing for access to a single capacity-limited destination. In this test, a non-recurrent incident is generated at the entrance to a border-crossing tunnel. In normal operation, the westbound approach to the tunnel entrance has priority access to the tunnel since those vehicles are making a right turn into the critical link, however this route is not critical for the management of oversaturation. Six different mitigation strategies including metering, dynamic-lane assignment, phase omits, and green-time re-allocation were tested to see if more equitable allocation of green-time could be provided for the two critical routes with minor effect on the non-critical routes. This test was also envisioned to explore improvement of performance to non-critical routes that were previously blocked by vehicles on the critical routes using the standard operations.

This test case also demonstrated how the on-line calculation of TOSI and SOSI measures with if…then logic could be used to select appropriate mitigation strategies in a closed-loop manner. Typical locations for advance detection were used for the measurement of TOSI and SOSI and triggering the selection of a new plan. Initially it was conceived that the application of if…then rules for selecting timing plans would be less complicated for practitioners than the use of traditional target-based traffic-responsive methods. In straight-forward situations with only one or two critical oversaturated approaches this is true. When applying the approach to three, four, or more detection points, the truth table becomes onerous to generate manually. It was clear in these experiments that adaptive control algorithms are necessary to make appropriate decisions in complex situations. This is an important direction for future research and development.

All of the mitigation strategies applied to this test case showed that more efficient use of available space could be achieved by applying metering and offset strategies to reduce the occurrence of SOSI > 0 along the critical route. The performance results for total travel time, delay, and throughput were largely inconclusive as most link-by-link performance was either not significantly different than the baseline, or performance improvements on some links were offset by performance detriments on other links. Because this scenario involves a single point of failure at the tunnel entrance, truncating phase green-time or omitting phases that have SOSI > 0 does not result in significant performance improvements for other approaches, since those movements are also blocked by the same downstream incident. It was found however that more equitable treatment for all routes could be obtained by disallowing right-turn-on-red and implementing gating on the non-priority routes.

The final test case in Surprise, AZ was focused on a heavily traveled arterial with pre-planned special event traffic overlaid on P.M. peak traffic that is already near oversaturated. The special event occurs roughly in the middle of the arterial network and most event traffic approaches from the east. In normal operation, the westbound approach is heavily queued at the entrance to the special event facility. Eight different mitigation strategies were formulated and tested on this scenario. The mitigations included combinations of cycle time adjustment, green-time re-allocation, negative and simultaneous offsets, dynamic lane allocation, and double-cycling. The results for this scenario indicated that, in general, all of the mitigations outperform the baseline operation for both total travel time, throughput and delay measures. Improvements in the performance of the westbound oversaturated route were offset by detriments to the non-critical eastbound route. However, the mitigations provided more equitable performance in the two route travel times, whereas when applying the baseline strategy the performance of the critical route was three to four times worse than the non-critical route. The mitigation strategies showed considerable improvement over the baseline operation during the recovery regime. Most of the mitigating strategies improved the system recovery time by more than 20 minutes.

Project Summary and Directions for Further Research

This goal of this project was to conduct research and develop guidance for practitioners in the application of mitigation strategies for oversaturated conditions. The body of knowledge in this area was quite limited when the project began. We focused our efforts on four topics:

1. Development of techniques to quantitatively characterize oversaturated conditions
2. Development of a technique for generating signal timing plan parameters and evaluating the effectiveness of those strategies
3. Developing an experimental tool for linking the quantitative measures of oversaturation to selecting of mitigating strategies in an on-line manner
4. Evaluating a wide range of strategies and providing evidence of effectiveness in a series of simulation experiments

These activities were then used to generate a practical guide for selection and application of strategies to particular oversaturated situations. Much more research and development in this area is needed to establish understanding, generalize the guidance, and develop implementable procedures for analysis and application of strategies. The complexity of issues that must be considered during oversaturated conditions is still daunting for the development of closed-form solutions or cookbook type procedures.

Some key take-away findings were identified during the project. The first key finding was that identifying the critical routes through a network of intersections is the first a critical first step in identifying appropriate mitigations. The methodology we developed for designing timing plans that explicitly consider critical routes showed promise that alternative formulation for the optimization process can result in significant gains in total system performance. In addition, this methodology is one of the first we know of that can optimize both the timing plan parameters of individual timing plans and the sequence and duration of application of those plans during a scenario. There is still much effort necessary to bring this complicated and experimental methodology closer to being able to be applied by a typical practitioner much like they might run a tool like Synchro or Transyt.

The maximization of system throughput is only possible by equitable allocation of green-time to the critical routes and movements in the system. 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. Furthermore, traditional thinking such as “more green is always better” can work directly against the throughput objective since if the downstream link is already significantly queued, the upstream traffic will not be able to move anyway. It is not intuitively obvious how green-time can be re-allocated most appropriately in this situation. The second key development in this project was the development of the TOSI and SOSI quantitative measures and the development of the heuristic FBP to directly compute re-allocation of green-times from those measurements. This process was developed and tested in an off-line manner in this project and shown to make improvements to an oversaturated scenario. These experiments point toward on-line application of the FBP to continually re-compute the green-time allocation in an adaptive manner.

The final key finding of the project was that it is important to consider operating the system differently during the three regimes of operation during oversaturated conditions:

- Loading
- Processing
- Recovery

During the loading regime, systems are best operated by continuing to minimize total delay or maximizing total throughput. When TOSI and SOSI become significantly > 0, strategies which manage queues on the critical routes are most effective. These strategies minimize the degree of saturation on critical routes in order to minimize SOSI and TOSI effects. Minimizing SOSI is the most important goal during the processing regime. Finally, in the recovery regime, strategies which maximize throughput are much more effective in clearing the overflow queues that were generating during the processing phase than other strategies.

Figure 179 illustrates the difference between the total output processing rate of a mitigating strategy versus a no-mitigation baseline operation. In particular, note the substantial difference between the performances of the mitigation strategy with the baseline control strategy *after the peak period ends*.

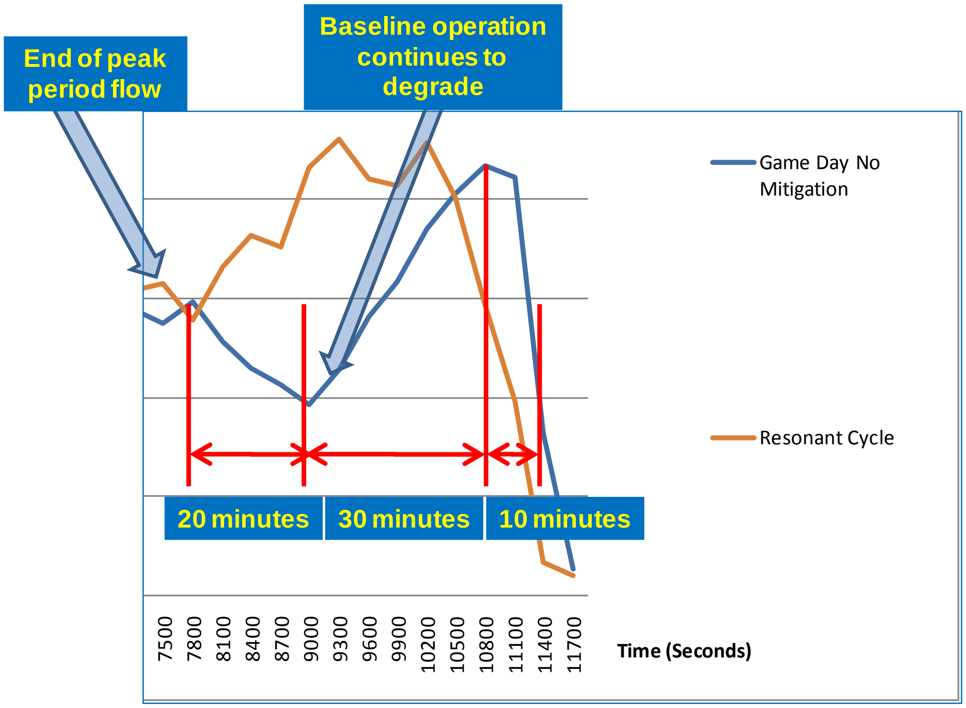


Figure 179. Comparison of output processing rates during recovery period

It is quite clear that the mitigation strategy begins processing more of the overflow queues much more efficiently after the peak period input flows subside. The no-mitigation strategy continues to decrease in total system output for an additional 20 minutes and takes another 30 minutes before its' peak output processing rate is finally reached (approximately 30 minutes behind the mitigation strategy's maximum output processing rate). Finally, the no-mitigation strategy returns the system to steady-state operation at least 10 minutes later than the mitigation strategy. All three of these metrics indicate that mitigation strategies designed to maximize throughput can be effective in improving total system output in the recovery regime. Of the three regimes, the largest performance improvements can be achieved by applying mitigations during the time when the system is recovering from the severe queuing.

Directions for Future Research

It is often said that good research generates more questions than provides answers. This project generated a number of significant directions for further research. While the goal of this project was to initially develop a guide for practitioners, there were simply too many unknowns in this area to distill a limited amount of benefits information into a guide or generate a prescriptive unified theory of operations.

As such, the first future research need is to evaluate additional test cases that will illustrate the performance of certain mitigations in specific situations. Based on the wide variety of potential mitigations and combinations of mitigations, it simply was not possible to conduct an evaluation of every type of treatment. While the real-world cases that were tested in this project are instructive, for the most part it is difficult to extrapolate specific results from a particular test to other scenarios. In order to fully develop a comprehensive guide, at minimum, an example application of each potential technique is needed.

Much more emphasis of researchers that study signal systems issues is needed in this area. As has been discussed several times in the context of TRB Signal Systems Committee meetings, it could be helpful if some standard test bed networks and situations (hypothetical or real world) were developed for researchers to test methodologies and compare the results apples to apples. We propose that the networks tested in this project may be put into the public domain for testing methods and evaluation of strategies developed by other researchers.

The second major need is to test and evaluate mitigating strategies in the real world. All of the tests that were performed during this project were done with simulation tools. It is well known that simulations have challenges in representation of real-world behaviors during oversaturation. Field testing and application of mitigations in real-world sites (among those that were tested in the simulation studies during this project) would certainly be a valuable research activity to follow this effort.

The practitioner guidance could be greatly improved by development of additional "rules of thumb" and more "cookbook" type design principles for mitigations. This could not be achieved during this research due to the immaturity of knowledge in this area. Furthermore, the combination of mitigations into comprehensive strategies is still more art than science and the methodology developed in this project is too cumbersome and complicated for use by a typical practitioner. Development of an off-line analysis tool that can develop mitigation strategies in general network structures will be a valuable future research topic.

Additional research is needed on the role of thresholds, persistence time, and recovery time in the measurement of TOSI and SOSI for selection of mitigations in an on-line manner. In this research, we selected what seemed to be common-sense values for these parameters but did not do any sensitivity analysis on the values of these inputs. Additional scenarios need also to be constructed and tested with the on-line tool for more comprehensive evaluation of the effectiveness of such a tool. In order to truly get such methods into real practice, system operators will have to "spec" and procure such features in upgrades or new installations of ATMS.

Finally, while it was found that real benefits can be achieved through application of fixed-parameter timing plans, it was clear that on-line adaptive feedback control methods would improve the operation of oversaturated systems. In particular, it appears that offsets and green-time on oversaturated critical routes need to be adjusted almost every cycle to mitigate TOSI and SOSI > 0. Negative offsets can be designed for a particular value of TOSI, but if the demand rate remains constant and the green-time is left constant, the queue length will continue to grow until the link is filled. Development of adaptive algorithms and logic that directly consider oversaturated conditions in actuated-coordinated systems would be of benefit to the industry. The FBP could be extended to a rolling-horizon formulation to take another step in that direction. Much additional research would be necessary to extend the basic heuristic of the FBP to consider phase sequencing, cycle time, protected/permitted lefts, and so on.

References

Abbas, M. M., Li, P., Hesham, P. & Rakha, A. (2007). Multi-objective strategies for timing signal system under oversaturated conditions. *Proceedings of the 18th IASTED International Conference: Modeling and Simulation,* 580–585.

Abu-Lebdeh, G. & Benekohal, R. F. (1997). Development of a Traffic Control and Queue Management Procedure for Oversaturated Arterials. *Transportation Research Record 1603*, TRB, National Research Council, Washington, D.C., 119–127.

Abu-Lebdeh, G. (1999). *Development of dynamic traffic signal control procedures for oversaturated arterials and genetic algorithms solutions.* Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign.

Abu-Lebdeh, G. & Benekohal, R. F. (2003). *Design and evaluation of dynamic traffic management strategies for congested conditions*. Transportation Research Part A, 37(2), 109–127.

Abu-Lebdeh, G. & Benekohal, R. F. (2000). Signal Coordination and Arterial Capacity in Oversaturated Conditions. *Transportation Research Record: Journal of the Transportation Research Board*, *No. 1727*, Transportation Research Board of the National Academies, Washington, D.C., 68–76.

Akcelik, R. (1981). *Traffic Signals: Capacity and Timing Analysis. Research Report 123*, Australian Road Research Board, Melbourne, Australia.

Akcelik, R. (1988). The highway capacity manual delay formula for signalized intersections. *ITE Journal,* 58 (3), 23–27.

Beaird, S., Urbanik II, T. & Bullock, D. (2006). Traffic Signal Phase Truncation in Event of Traffic Flow Restriction. *Transportation Research Record: Journal of the Transportation Research Board, No. 1978*, Transportation Research Board of the National Academies, Washington, D.C.

Benekohal R. F. & Kim, S. O. (2005). Arrival-based Uniform Delay Model for Oversaturated Signalized Intersections with Poor Progression. *Transportation Research Record: Journal of the Transportation Research Board, No. 1920*, Transportation Research Board of the National Academies, Washington, D.C., 86–94.

Boettger, R. (1982). On-line optimization of the offset in signalized street networks. IEE International Conference. Road Traffic Signaling, 207.

Boettger, R. (1971). *Optimal coordination of traffic signals in street networks*. V International Symposium on the Theory of Traffic Flow and Transportation. University of California, Berkeley.

Bonneson, J. (2005). "Measuring and Predicting the Performance of Automobile Traffic on Urban Streets". Interim Report, NCHRP Project 03-79(4). Texas A&M Research Foundation.

Bonneson, J., Pratt, M., & Zimmerman, K. (2009). *Traffic Signal Operations Handbook*. http://tti.tamu.edu/documents/0-5629-P1.pdf

Bretherton, R. D. (1989). SCOOT urban traffic control system – philosophy and evaluation. IFAC Symposium on Control. *Communications in Transportation*, 237–239.

Bretherton, R. D. & Bowen, G. T. (1990). Recent enhancements to SCOOT – SCOOT version 2.4. *Proceedings of the 3rd Conference on Road Traffic Control*, IEE, London.

Chang, T.-H., & Lin, J.-T. (2000). Optimal Signal Timing for an Oversaturated Intersection. *Transportation Research Part B*, 34(6), 471–491.

Chaudury, N. & Chu, C.-L. 2000. *Guidelines for Timing and Coordinating Diamond Interchanges with Adjacent Traffic Signals*. http://tti.tamu.edu/documents/4913-2.pdf

Cronje, W. B. (1983a). Analysis of Existing Formulas for Delay, Overflow, and Stops. *Transportation Research Record 905*, TRB, National Research Council, Washington, D.C., 89–93.

Cronje, W. B. (1983b). Derivation of Equations for Queue Length, Stops, and Delay for Fixed-Time Traffic Signals. *Transportation Research Record 905*, TRB, National Research Council, Washington, D.C., 93–95.

Cronje, W. B. (1986). Comparative Analysis of Models for Estimating Delay for Oversaturated Conditions at Fixed-Time Traffic Signals. *Transportation Research Record 1091*, TRB, National Research Council, Washington, D.C., 48–59.

Denney, R. W., Head, L., and Spencer, K. (2008). Signal Timing Under Saturated Conditions, Federal Highway Administration, Report No. FHWA-HOP-09-008.

Diakaki, C., Dinopoulou, V., Aboudolas, K., Papageorgiou, M., Ben-Shabat, E., Seider, E. & Leibov, A. (2003). Extensions and New Applications of the Traffic-Responsive Urban Control Strategy Coordinated Signal Control for Urban Networks. *Transportation Research Record: Journal of the Transportation Research Board, No. 1856*, Transportation Research Board of the National Academies, Washington, D.C., 202–211.

Dineen, M. (2000). *Real-time display of Dublin traffic information on the web.* Unpublished doctoral dissertation, University of Dublin.

Dinopoulou, V., Diakaki, C. & Papageorgiou, M. (2005). Application and evaluation of the signal traffic control strategy TUC in Chania. *Journal of Intelligent Transportation Systems*, 9(3), 133–143.

Dion, F., Rakha, H. & Kang, K. Y. (2004). Comparison of delay estimates at under-saturated and over-saturated pre-timed signalized intersections. *Transportation Research Part B*, 38, 99–122.

Dion F., & Yagar S. (1996). *Real-time control of signalised networks – different approaches for different needs.* Eighth International Conference on Road Traffic Monitoring and Control (Conf.Publ. No.422), IEE, 56–60.

Dunne, M. & Potts, R. (1964). Algorithm for traffic control. *Operations Research*, 12(6), 870–881.

Fambro, D. B. & Rouphail, N. M. (1997). Generalized Delay Model for Signalized Intersections and Arterial Streets. *Transportation Research Record 1572*, TRB, National Research Council, Washington, D.C., 112–121.

Gartner, N. H. (1989). OPAC: Strategy for demand-responsive decentralized traffic signal control. *Control, Computers, Communications in Transportation: Selected Papers from the IFAC - IFIP - IFORS Symposium, Paris, France*, 241–244.

Gartner N. (1983). A Demand Responsive Strategy for Traffic Signal Control. *Transportation Research Record 906*, TRB, National Research Council, Washington, D.C.

Gartner, N, H. (2001), Presentation on the TRB workshop of Adaptive Traffic Signal Control System, Jan., Washington D.C. Accessed June 6, 2009. http://www.signalsystems.org.vt.edu/documents/Jan2001AnnualMeeting/TRB2001_Slides_for_posting.pdf

Gartner, N. H., Pooran, F. J. & Andrews, C. M. (2002). Optimized Policies for Adaptive Control Strategy in Real-Time Traffic Adaptive Control Systems Implementation and Field Testing. *Transportation Research Record: Journal of the Transportation Research Board, No. 1811*, Transportation Research Board of the National Academies, Washington, D.C., 148–156.

Gazis, D. C. (1964). Optimum Control of a System of Oversaturated Intersections. *Operations Research*, 12(6), 815–831.

Gazis, D. C. & Potts, R. B. (1965). The Oversaturated Intersection. In *Proc. 2nd International Symposium on the Theory of Traffic Flow*.

Gettman, D. (2005). Presentation in TRB Signal System Committee Mid-year Meeting Performance Measures Workshop, LA, NV. Accessed June 6, 2009. http://www.signalsystems.org.vt.edu/documents/July2005AnnualMeeting/Presentations/ACSLITE%20Performance%20Measures%20v5.pdf

Gettman, D., Shelby, S. G., Head, L., Bullock, D. M. & Soyke, N. (2007). Data-Driven Algorithms for Real-Time Adaptive Tuning of Offsets in Coordinated Traffic Signal Systems. *Transportation Research Record: Journal of the Transportation Research Board, No. 2035*, Transportation Research Board of the National Academies, Washington, D.C., 1–9.

Girianna, M., & Benekohal, R. F. (2002). Dynamic Signal Coordination for Networks with Oversaturated Intersections. Presented at Annual Meeting of 81st Transportation Research Board, Washington, D.C.

Gordon, R. L. (1969). A Technique for Control of Traffic at Critical Intersections. *Transportation Science*, 4, 279–287.

Gray, B. M. and Ibbetson, L. (1991). Operational Traffic Control Strategy for Congestion. *Traffic Engineering and Control*, 60–66.

Green, D. H. (1968). Control of Oversaturated Intersections. *Operational Research Quarterly*, 18(2), 161–173.

Hallenbeck, M. E., Ishimaru, J. M., Davis, K. D. & Kang, J. M. (2008). Arterial Performance Monitoring Using Stop Bar Sensor Data. Presented at the 87th Annual Meeting of the Transportation Research Board.

Head, K. L. (1995). An Event-Based Short-Term Traffic Flow Prediction Model. *Transportation Research Record 1510*, TRB, National Research Council, Washington, D.C., 45–52.

Henry, J. J. (1989). PRODYN tests and future experiments on ZELT. VNIS `89: Vehicle Navigation and Information Systems, IEEE Conference, Toronto.

Highway Capacity Manual. (2000). Transportation Research Board, National Research Council, Washington, D.C.

Huddart, K. W. & Wright, C. (1989a). Strategies for Urban Traffic Control. The Rees Jeffrey's Road Fund, *Transport and Society*.

Huddart, K. W. & Wright, C. (1989b). Catastrophic Traffic Congestion and Some Possible Ways of Preventing It. *Proc. TRAFFEX International Traffic Engineering Exhibition*. Seminar on Congestion, Control and Parking Enforcement. Brighton, U.K.

Hunt, P. B., Robertson, D. I., Bretherton, R. D. & Winton, R. I. (1981). SCOOT - a traffic responsive method of coordinating signals. TRL Laboratory Report, 1014.

Hurdle, V. F. (1984). Signalized Intersection Delay Model: A Primer for the Uninitiated. Transportation Research Record 971, TRB, National Research Council, Washington, D.C., 101.

Kang, Y. S. (2000). *Delay, Stop and Queue Estimation for Uniform and Random Traffic Arrivals at Fixed-Time Signalized Intersections*. Unpublished doctoral dissertation, Blacksburg, Virginia.

Kim, S. O. & Benekohal, R. F. (2005). Comparison of Control Delays for CORSIM and the Highway Capacity Manual for Oversaturated Signalized Intersections. *Journal of Transportation Engineering*, 131(12), 917–923.

Kim, Y. & Messer, C. J. (1992). Traffic signal timing models for oversaturated signalized interchanges. *TTI Research Report*, 1148-2.

Kimber, R. M. & Hollis, E. M. (1979). Traffic queues and delays at road junctions, Laboratory Report 909. U.K. Transport and Road Research Laboratory, Crowthorne, Berkshire, England.

Khatib, Z., Abdel-Rahim, A., Judd, G., & Jagarapu, K. (2002). Actuated Coordinated Signal Systems. Phase I: Oversaturated

Conditions. http://ntl.bts.gov/lib/11000/11500/11562/KLK202.pdf

Koonce, P., Rodegerdts, L., Lee, K., Quayle, S., Beaird, S., Braud, C., Bonneson, J., Tarnoff, P., & Urbanik, T. (2008). *FHWA Traffic Signal Timing Manual*. http://ops.fhwa.dot.gov/arterial_mgmt/tstmanual.htm

Kovvali, V., Messer, C., Chaudhary, N., & Chu, C.-L. (2002). A Program for Optimizing Diamond Interchanges in Oversaturated Conditions. Transportation Research Record: Journal of the Transportation Research Board, No. 1811, Transportation Research Board of the National Academies, Washington, D.C.

Lan, C.-J., Messer, C. J., Chaudhary, N. A., & Chang, E. (1992). Compromise Approach to Optimize Traffic Signal Coordination Problems During Unsaturated Conditions. *Transportation Research Record 1360*, TRB, National Research Council, Washington, D.C., 112–120.

Li, M. T. & Gan, A. C. (1999). Signal Timing Optimization for Oversaturated Networks Using TRANSYT-7F. *Transportation Research Record: Journal of the Transportation Research Board, No. 1683,* TRB, National Research Council, Washington, D.C.

Li, H., & Prevedouros, P. D. (2004). Traffic Adaptive Control for Oversaturated Isolated Intersections: Model Development and Simulation Testing. *Journal of Transportation Engineering*.

Li, J., Rouphail, N. M., Tarko, A. & Velichansky, L. (1996). Overflow Delay Model for Signalized Arterials. *Transportation Research Record 1555*, TRB, National Research Council, Washington, D.C.

Lieberman, E. (1992). “Traffic Signal Control for Saturated Conditions”. Final report, NCHRP Project 03-38(4). KLD Associates, Inc.

Lieberman, E. B., Chang, J. & Prassas, E. S. (2000). Formulation of Real-Time Control Policy for Oversaturated Arterials. *Transportation Research Record: Journal of the Transportation Research Board, No. 1727*, TRB, National Research Council, Washington, D.C., 77– 88.

Lieberman, E., Chang, J., Bertoli, B., & Xin, W. (2010) New Signal Control Optimization Policy for Oversaturated Arterial Systems. Presented at 89th Annual Meeting of the Transportation Research Board, Washington, D.C.

Lieberman, E. B., Rathi, A. K. & King, G. F. (1986). Congestion Based Control Scheme for Closely Spaced, High Traffic Density Networks. Presented at 65th Annual Meeting of the Transportation Research Board, Washington, D.C.

Lighthill, M. J. & Whitham, G. B. (1955). On Kinematic Waves: II a Theory of Traffic Flow on Crowded Roads. *Proc, Royal Society London, Series A*, 229(1178).

Liu, H. & Ma, W. (2009). A Virtual Probe Approach for Time-Dependent Arterial Travel Time Estimation. *Transportation Research Part C*, 17(1), 11–26.

Liu, H., Wu, X., Ma, W., & Hu, H. (2009a). Real-Time Queue Length Estimation for Congested Signalized Intersections. *Transportation Research Part C*, 17(4), 412–427.

Liu, H., Ma, W., Wu, X., & Hu, H. (2009b). Real-Time Estimation of Arterial Travel Time under Congested Conditions. *Transportmetrica*, in press.

Lo, H. K. & Chow, A. H. F. (2004). Control Strategies for Oversaturated Traffic, *Journal of Transportation Engineering*, 466–478.

Longley, D. (1968). A Control Strategy for Congested Computer Controlled Traffic Network. *Transportation Research*, 2(4), 391–408.

Luyanda, F., Gettman, D., Head, L., Shelby, S., Bullock, D. & Mirchandani, P. (2003). ACS-lite Algorithmic Architecture Applying Adaptive Control System Technology to Closed-Loop Traffic Signal Control Systems. Transportation Research Record: Journal of the Transportation Research Board, No. 1856, Transportation Research Board of the National Academies, Washington, D.C.

Mahalel, D., Yehuda, G. & Shiftan, Y. (1991). Manual Versus Automatic Operation of Traffic Signals. *Transportation Research A*, 25A, 121–127.

Martin, P. (2006). Presentation at TRB 2006 Workshop on Oversaturation, Jan., Washington D.C. Accessed June 6, 2009. http://www.signalsystems.org.vt.edu/documents/Jan2007AnnualMeeting/Presentations/ATCS_Oversaturation_TRB.pdf

May, A. D., Montgomery, F. O. & Quinn, D. J. (1988). Control of Congestion in Highly Congested Networks. *Proc. CODATU IV Conference*, Jakarta.

Messer, C. (1998). Extension and Application of Prosser-Dunne Model to Traffic Operation Analysis of Oversaturated, Closely Spaced Signalized Intersections. *Transportation Research Record 1646*, TRB, National Research Council, Washington, D.C., 106–114.

Mirchandani, P. & Head, L. (2001). A Real-Time Traffic Signal Control System: Architecture, Algorithms, and Analysis. *Transportation Research Part C*, 9(2), 415–43.

Michalopoulos, P. G. & Stephanopoulos, G. (1979). An Algorithm for Real Time Control of Critical Intersection. *Traffic Engineering and Control*, 20, 9–15.

Michalopoulos, P. G., & Stephanopoulos, G. (1977a). Over-Saturated Signal Systems with Queue Length Constraints - I. Single Intersection. *Transportation Research*, 11(6), 413–428.

Michalopoulos, P.G. and Stephanopoulos, G. (1977b) Oversaturated Signal System with Queue Length Constraints – II: Systems of Intersections, *Transportation Research*, Vol. 11(6), 423–428.

Minnesota Department of Transportation. (2011). Traffic Signal Timing and Coordination Manual. http://www.dot.state.mn.us/trafficeng/publ/signaloperations/2011_Timing%20&_Coordination_Manual2011.pdf

Mueck, J. (2002). Using Detectors Near the Stop-Line to Estimate Traffic Flows. *Traffic Engineering and Control*, 43, 429–434.

Mueck, J. (2005). Recent Developments in Adaptive Control Systems in Germany Accession. 12th World Congress on Intelligent Transport Systems, ITS America Publication.

Newell, G. F. (2001). A Simplified Car-Following Theory: a Lower Order Model, *Transportation Research Part B*, 36(3), 195–205.

Papagcorgiro, M., Kosmatopoulos, E., Bielefeldt, C., Dinopoulou, V., Morris, R., Mueck, J., Richards, A., & Weichenmeier, F. (2006). Comparative Field Evaluation of the Signal Control Strategy TUC in Three Traffic Networks. Presented at Annual Meeting of the 85th Transportation Research Board, Washington D.C.

Park, B., Messer, C. J., & Urbanik II, T. (2000). Enhanced Genetic Algorithm for Signal Timing Optimization of Oversaturated Intersections. Transportation Research Record: Journal of the Transportation Research Board, No. 1727, TRB, National Research Council, Washington, D.C., 32–41.

Perrin, J., Martin, P. T. & Coleman, B. (2002). Monitoring Commuter Congestion on Surface Streets in Real Time. *Transportation Research Record: Journal of the Transportation*

Research Board, No. 1881, Transportation Research Board of the National Academies, Washington, D.C., 107–114.

McShane, W. R., Crowley, K. W., Lee, B., & Casey, T. W. (1978). *NCHRP Report 194: Traffic control in oversaturated street networks*. TRB, National Research Council, Washington D.C.

Prosser, N. and Dunne, M. (1994). A Procedure for Estimating Movement Capacities at Signalized Paired Intersections. *Proc., 2nd International Symposium on Highway Capacity*, Vol. 2, Sydney, Australia, 473–492.

Quinn, D. J. (1992). A review of queue management strategies. *Traffic Engineering and Control*, 3311, 600–605.

Rakha, H., Kang, Y. S. & Dion, F. (2001). Estimating Vehicle Stops at Undersaturated and Oversaturated Fixed-Time Signalized Intersections. Transportation Research Record: Journal of the Transportation Research Board, No. 1776, TRB, National Research Council, Washington, D.C., 128–137.

Rathi, A. K. (1988). Control Scheme for High Density Traffic Sectors. *Transportation Research Part B*, 22B(2), 81–101.

Rathi, A. K. (1991). Traffic Metering: An Effectiveness Study. *Transportation Science*, 421–440.

Rathi, A. K. & Lieberman, E. B. (1991). Effectiveness of Traffic Restraint for a Congested Urban Network: A Simulation Study. *Transportation Research Record 1232*, TRB, National Research Council, Washington, D.C., 95–102.

Robertson, D. I. (1969). *TRANSYT: A traffic network study tool*. Ministry of Transport LR253, TRRL, Crowthorne, U.K.

Roess, R. P., Prassas, E. S., & McShane, R. W. (2004). *Traffic Engineering (3rd Edition)*, Pearson Prentice Hall.

Rouphail, N. M & Akcelik, R. (1992). A Preliminary Model of Queue Interaction at Signalized Paired Intersections. *Proceedings of Australian Road Research Board*, 16(5), Vermont South, Victoria, 325–345.

Sayers, T., Anderson, J., & Bell M. (1998). Traffic control system optimization: A multi-objective approach. Cardiff Proceedings of the third conference on Mathematics in Transport Planning and Control, University of Wales, Cardiff.

Sharma, A., Bullock, D. M., & Bonneson, J. (2007). Input-Output and Hybrid Techniques for Real-Time Prediction of Delay and Maximum Queue Length at a Signalized Intersection. Transportation Research Record: Journal of the Transportation Research Board, No. 2035, Transportation Research Board of the National Academies, Washington, D.C., 88–96.

Sharma, A. & Bullock, D. M. (2008). Field Evaluation of Alternative Real-Time Methods for Estimating Delay at Signalized Intersections. *Proceedings of the 10th International Conference on Applications of Advanced Technologies in Transportation*, Athens, Greece, May 27–31.

Shelby, S., Bullock, D., & Gettman, D. (2005). Resonant Cycles in Traffic Signal Control. *Transportation Research Record: Journal of the Transportation Research Board, No. 1925*. Transportation Research Board of the National Academies, Washington, D.C.

Shepherd, S. P. (1992). *A Review of Traffic Signal Control*. Working Paper 349, Institute for Transport Studies, Leeds, U.K.

Sims, A. G. & Dobinson, K. W. (1980). The Sydney coordinated adaptive traffic (SCAT) system philosophy and benefits. *IEEE,* VT-2(2), 130–137.

Sims, A. G. & Finlay, A.B. (1984). *SCATS*. Splits and offsets simplified (SOS). *ARRB Proceedings,* 12(4).

Smaglik E.J., Sharma, A., Bullock, D.M., Sturdevant, J.R., & Duncan, G. (2007). Event-Based Data Collection for Generating Actuated Controller Performance Measures, *Transportation Research Record: Journal of the Transportation Research Board, No. 2035*, Transportation Research Board of the National Academies, Washington, D.C., 97–106.

Smaglik, E J., Bullock, D. M., & Sharma, A. (2007). Pilot study on real-time calculation of arrival type for assessment of arterial performance. *Journal of Transportation Engineering*, July, 415–422

Tian, Z. & Urbanik, T. (2006). Application of Diamond Interchange Control Strategies at Closely-Spaced Intersections. Presented at the Annual Meeting of the 86th Transportation Research Board, Washington, D.C.

Webster, F. V. (1961). *Traffic Signal Settings*. Road Research Laboratory Technical Paper, 39.

Webster, F. V. & Cobbe, B. M. (1996). *Traffic Signals*. Road Research Laboratory Technical Paper, 56.

Appendix A: Literature Review

Summary of the Literature Review

Development of strategies to handle oversaturated conditions is not a new topic of consideration. The research team found a wide variety of work in both diagnosis and estimation of oversaturation and control strategies and scenarios. In diagnosis and estimation we focused on reviewing techniques for measurement of queues and techniques for measuring the degree of saturation and surrogates for the degree of saturation. There has been quite a bit of work in the past on estimation of delay during oversaturated conditions and approaches for modeling oversaturated conditions. These efforts focused primarily on Highway Capacity Manual-type analysis and are thus not directly applicable to this project. Research on queue estimation is dominated by input-output modeling approaches.

Input-Output methods are limited to estimating queues up to the point of the input detector, but not beyond this point. For arterial streets this requires installation of exit-side detection in order to measure a queue that is the full length of the link. Such detector installation can be cost-prohibitive. Methods for measuring the degree of saturation can identify the saturation level up to the point of saturation, but estimates of saturation above 1.0 have not been shown to be reliable, except for those estimates used by SCOOT and SCATS. The methods researched in this project provide some insight into the level of severity of the oversaturation (and queue length), which we believe is critical in identifying appropriate mitigation strategies.

In the review of strategies and scenarios we looked at previous research on adaptive control systems, optimal control formulations, and various other approaches. Features of adaptive control systems and most other strategies are described in the literature in a qualitative manner. Concepts can be leveraged, but specific algorithms are not typically described quantitatively. Notably the features of SCOOT and SCATS that handle oversaturated conditions are mostly if…then type rules with thresholds that change some parameters or impose additional constraints on certain decision variables. Descriptions of these features are not accompanied by research indicating their ability to be effective in the real-world. Simulation studies evaluating the performance of these features could not be found either. Quantitative approaches (i.e. optimal control formulations) that were found in the literature all require information on traffic volumes, queue lengths, or both. Volume information is the most difficult information to obtain during oversaturation using state-of-the-practice detection systems, which makes most of the optimal formulations rendered ineffective.

The literature review is divided into two focus areas. The first focus area is on the diagnosis of oversaturation and the second focus area is on strategies for mitigation of oversaturation. In several cases, the same reference material appears in both sections but the focus of discussion is on either the diagnosis or the strategy component of the particular research literature.

Literature Review on Diagnosis of Oversaturated Conditions

Although there has not been a significant amount of literature devoted to how to manage oversaturated traffic signal systems, there has been even less effort devoted to the identification of oversaturated conditions, with both spatial and temporal extent. Many management strategies as detailed in Part 2 of the literature review assume that arrival volumes can be somehow known and the oversaturated conditions can be accurately predicted based on this information. In the real world, this is simply not the case. Existing detection systems provide observations of flows at a fixed point on a link, which, during saturation, fail to provide the same accuracy of measurements. The techniques identified in this section of the literature review cover methods for estimating oversaturated conditions or measures of saturation at traffic signals. Using traffic data from signal systems to diagnose and identify oversaturation is sporadic and inconsistent in the literature, as we summarize in the following section.

An oversaturated intersection movement (or lane group) can be defined as one in which the traffic demand exceeds the capacity. Based on this definition of oversaturation, the v/c ratio (v is demand and c capacity) has been used in theoretical formulae to identify whether an approach or movement is oversaturated. Analytically, the v/c ratio for each lane group can be estimated directly by dividing the demand flow rate by the capacity, by using the following equation:

$$X_i = \frac{v_i}{c_i} = \frac{\left(v/s\right)_i}{\left(g/C\right)_i} \qquad \text{Eq. A-1}$$

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; $\left(v/s\right)_i$ and $\left(g/C\right)_i$ are the flow ratio and the 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 (1964) expanded this concept to diagnose oversaturation by testing the following inequality:

$$q_1/s_1 + q_2/s_2 > 1 - (L/C) \qquad \text{Eq. A-2}$$

where, q_1 and q_2 are arrival rates for two directions; s_1 and s_2 are saturation flow rates for two directions; L is the total lost time and C is the cycle length. Since Eq. A-2 only fits for intersections with fixed cycle length and lost time, Green (1967) modified it to Eq. A-3, which was called "absolute" oversaturation to deal with the situation when C is not a fixed value.

$$q_1/s_1 + q_2/s_2 > 1 \qquad \text{Eq. A-3}$$

Direct application of the above models, however, is difficult. This concept might be extended from an intersection definition to a network definition by adding all of the demand flows in a network, but it is not clear how such as definition could be used in diagnosis and treatment of specific operational issues. Further difficulties in using such a definition arise because of the uncertainty of the capacity and saturation flow and due to the difficulty in measuring the arrival flow using current data collection systems (especially under congested situations – the very conditions that we are trying to identify). Because of this, other researchers have pursued alternative characterizations of oversaturation.

Definitions of Congestion and Level of Service

Longley (1968) 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 unsignalized 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 unsignalized intersections is not useful in our work here as we are primarily concerned about signalized intersections.

Pignataro et al. (1978) 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 (we presume "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 subcategories: *saturated* and *over-saturated* operations.

    i) **Saturated Operations**: a term that describes that range of congestion where 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 where a queue exists and it has grown to the point where upstream traffic operations are adversely affected.

In Pignataro's taxonomy, the distinction between saturated and oversaturated operations is thus the effect of the queuing on upstream operations. It is an interesting proposal for the distinction of the terms of "saturation" and "oversaturation", but no quantitative measures are provided in Pignataro et al. (1978) on how to identify such situations.

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

Local congestion: occurs when more vehicles face a cycle failure that does not result in damaging or excessive queues.

Extended congestion: cycle failures repeat such that queues extend damagingly through upstream intersections causing the capacity of the upstream intersections to be reduced.

Regional congestion: occurs when the queue from a critical intersection joins or influences the queue at upstream critical intersections.

Intermittent congestion: occurs as a natural result of stochastic traffic arrivals. Even at light volumes timed using the Poisson method, one might expect cycle failures 5% of the time.

Recurrent (cyclical) congestion: occurs at an intersection with insufficient capacity that occurs predictably as a result of foreseeable demand patterns.

Prolonged congestion: congestion at intersections creates such inefficiencies that demand must fall below the reduced capacity for extended periods to permit overflow queues to clear.

These different descriptions for congestion are qualitative. 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 measurement, travel delay/time, travel stops/speed, the flow-density/occupancy relationship, and green time utilization for diagnosis of oversaturation.

Definitions Based on Queue Length

Queuing is one of the most important indicators of oversaturation. Due to heavy demand or insufficient capacity, oversaturation has been characterized as the growth of queues. Gazis (1964) first characterized oversaturation that "a stopped queue cannot be completely dissipated during a green cycle". This is also commonly known as a cycle failure or a phase failure. Later, many other researchers explored the relation between queue length and traffic conditions at intersections. For example, Kimber and Hollis (1979) proposed a model to describe the relationship between queue length and intensity, which is described as the demand-to-capacity ratio. Akcelik (1981,

1988) investigated the queue-based intersection delay under oversaturated conditions. Abu-Lebdeh and Benekohal (2003) defined oversaturation such that “traffic queues persist from cycle to cycle either due to insufficient green splits or because of blockage”. These efforts contribute to the understanding of oversaturation which have been summarized by Roess et al. (2004) in his textbook *Traffic Engineering* 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 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 (Denney et al., 2008) defined different types of traffic conditions based on queue information (associated with an individual intersection).

1) **Light traffic**. Characterized by the expectation of minimized cycle failure. In such conditions, the signal can fully serve arrival queues and cycle failures are expected to be infrequent at less than 25% of all cycles.
2) **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.
3) **Heavy traffic**. Characterized by frequent cycle failures, but with a residual queue that ebbs and flows without growing uncontrollably. The flows exceed capacity nearly half of the time due to stochastic arrivals.
4) **Oversaturated operation**. Characterized either by excessive residual queues that grow without control, therefore causing more widespread damage to the operation of a network.

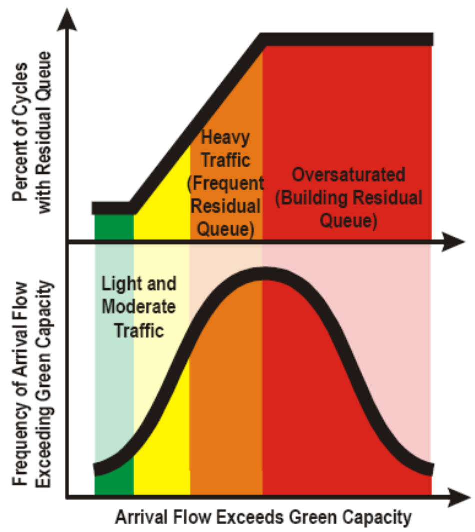


Figure A-1. Saturation and residual queues (Source: Denney et al., 2008)

Figure A-1 graphically illustrates the relationship between such traffic conditions (including light, moderate, heavy and oversaturated) and the residual queue length. 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 demand), which requires installation of upstream detectors. For most agency-standard detection technology deployments at intersections, this requirement cannot be satisfied. Typically, advance detectors are installed within a few hundred feet from the stop line, depending on the speed of approaching traffic, for vehicle actuation purposes on high-speed approaches (dilemma zone protection). Traditional input-output approaches to estimating queue length will not work appropriately if the queue length spills over the advance detector location. Therefore, either additional upstream detectors need to be installed or alternative methods to estimate queue length based on a typical layout of detector placements needs to be developed (or use of alternative measures to identify queue length).

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 A-2 (Dion et al., 2004) compares typical delay accumulation curves for both under-saturated and over-saturated conditions. Figure A-2 also shows that the overflow delay becomes over time, a more and more significant component of over-all intersection delay. How this delay accumulates is extremely difficult to measure.

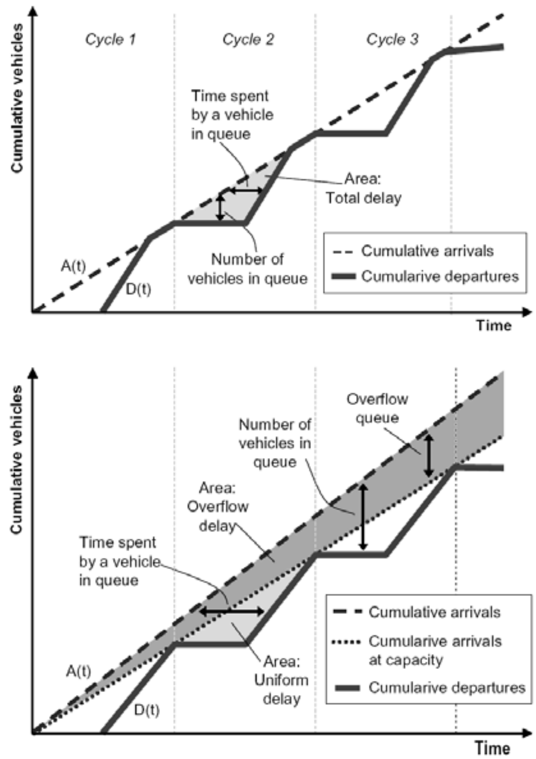


Figure A-2. Idealized cumulative arrivals and departures for under and oversaturated conditions (Source: Dion, F. et al., 2004)

Many different delay models for oversaturated intersections have been proposed and compared with traditional delay models which were designed to measure delay during undersaturated conditions (Fambro & Rouphail, 1997; Kang, 2000, Dion et al., 2004; Benekohal & Kim, 2005; Kim & Benekohal, 2005;). Figure A-3 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 (Roess et al., 2004). As illustrated, the estimated delays at oversaturated intersections significantly increase compared with random delays of undersaturated intersections. Most research has hypothesized that this growth rate in the average delay per vehicle varies linearly as the v/c ratio increases.

These methods 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.

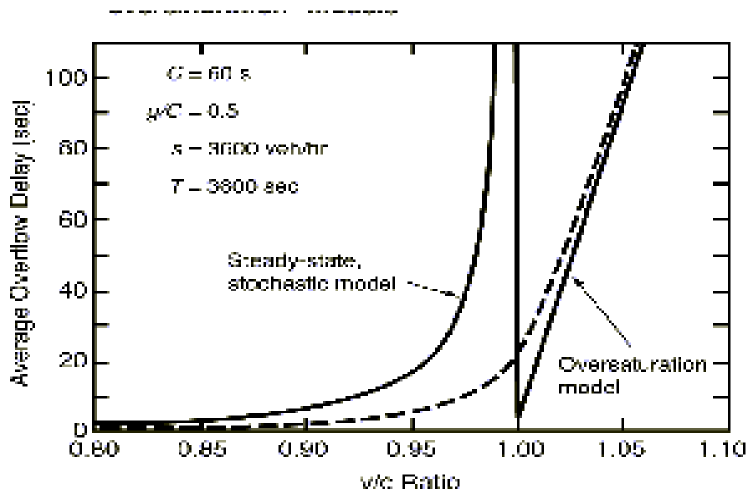


Figure A-3. Random and overflow delay models compared (ource: Hurdle, Roess, et al., 2004)

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. Most models which are designed to estimate speed, stops, and travel time performance during oversaturation are based on measuring queues at intersections. For example, the model developed by Cronje (1983a, 1983b, 1986) 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.

Flow-Occupancy Diagram/Fundamental Diagram

Most of the performance measurements mentioned above (queue, delay, speed, travel time, etc.) can be estimated based on two basic measures, i.e. flow and occupancy of fixed-location detectors. Density of traffic on an arterial can be estimated by occupancy at a fixed location. The relationship between these two measures, i.e. the flow-occupancy diagram or the fundamental diagram, can be used to indicate traffic conditions. Different conditions (undersaturated, saturated and oversaturated) are located at different areas of the fundamental diagram. Recently, some related studies have focused on how to use such information to estimate arterial traffic system performance. For example, Perrin et al. (2002) used occupancy data to estimate the v/c ratio and level of service (LOS); Sharma et al. (2007) combined detector and signal phase information to estimate vehicle delay and queue measurements; and Hallenbeck et al. (2008) used stop-bar detector data combined with signal state data to estimate arterial traffic conditions (congestion) from the perspective of the fundamental flow-occupancy diagram. Figure A-4 (Hallenbeck et al., 2008) indicates that oversaturated conditions can correspond to the right-hand side of the volume-occupancy diagram. Although most of the data in these studies are generated from simulations, these general ideas will have a contribution on our research on the identification of oversaturation.

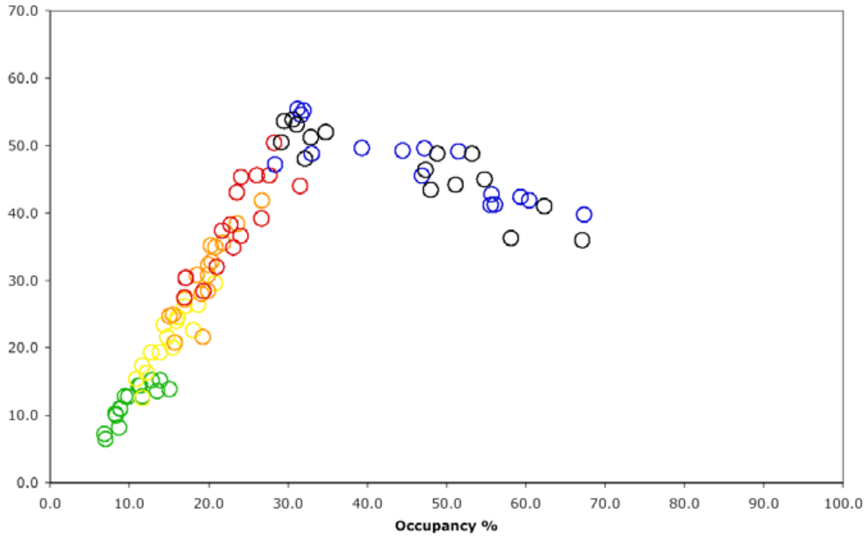


Figure A-4. A comparison of occupancy percentage and corresponding arterial congestion. Different colors indicate different congestion levels. (Source: Hallenbeck et al., 2008)

Utilization of Green Time

Another measure that has been used to identify saturation is green utilization. If a traffic signal phase is oversaturated, vehicles will continue to discharge at the saturation flow rate until the end of the effective green. When the traffic phase is undersaturated, there is some proportion of the green time (or green split) that is not used by traffic and is thus available to other phases when the intersection is operated in actuated mode. Under fixed-time control, this extra time is identified as green time with no associated occupancy on traffic detectors that serve that phase. The proportion between the used green time and the green phase, i.e. green utilization, can be treated as an indicator of the saturation level of a particular traffic phase (Gettman, et al, 2007). This method, however, does not serve to estimate the degree to which a certain traffic phase is oversaturated (i.e. how long the queue actually is). The indicator simply identifies that there is not enough green time to service all of the traffic demand, but it does not indicate how much green time would be necessary to clear the queue.

A similar concept is also proposed by Smaglik et al. (2007). This research uses the proportion of vehicles arriving on green to identify the traffic arrival type. As presented in Figure A-5, by using setback detectors, vehicle arrivals are collected and tabulated into cycle-by-cycle bins; and Eq. A-4 is used to calculate the proportion of vehicles arriving on green, which is the proportion between the number of vehicles arriving on green and the total number of vehicles arriving within a cycle. Applying the formula provided by HCM 2000 (Eq. A-5), the platoon ratio can be estimated as well as the arrival type and the quality of progression (Figure A-6).

$$P = \left(\frac{N_g}{N_r + N_g} \right) \qquad \text{Eq. A-4}$$

where, P is the proportion of vehicle arriving on green; N_g is the number of vehicles arriving on green; and N_r is the number of vehicles arriving on red.

$$P_p = P\left(\frac{C}{g} \right) \qquad \text{Eq. A-5}$$

where, P_p is platoon ratio; C is cycle length; and g is green interval.

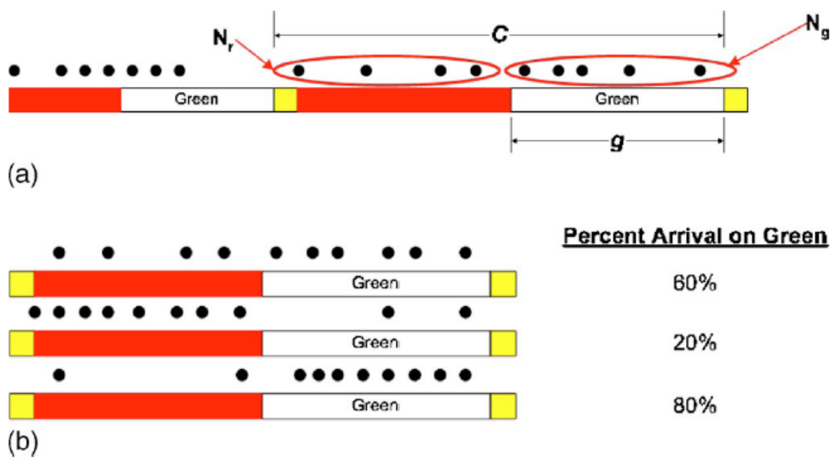


Figure A-5. Calculation of proportion of vehicles arriving on green (P): (a) binning of vehicle arrivals during different indications; (b) cycle by cycle binning (Source: Smaglik et al., 2007)

Arrival type	Range of platoon ratio (R_p)	Default value (R_p)	Progression quality
1	≤0.50	0.333	Very poor
2	>0.50–0.85	0.667	Unfavorable
3	>0.85–1.15	1.000	Random arrivals
4	>1.15–1.50	1.333	Favorable
5	>1.50–2.00	1.667	Highly favorable
6	>2.00	2.000	Exceptional

Figure A-6. Relationship between arrival type and platoon ratio (Source: HCM 2000)

It is necessary to point out that the high green utilization of a given phase is not always necessarily due to oversaturation. In some cases, good coordination design can generate high utilization on a phase by synchronizing the arrival of the incoming platoon with the start of green. Green utilization is not a complete picture of latent demand in much the same way that input-output methods for queue estimation fail to measure the stored demand. Green utilization can only indicate that there is more demand for a given phase, but not by how much.

The State of the Practice in Diagnosis of Oversaturated Conditions

All the measures described above can be used as indicators to diagnose oversaturation with varying levels of effectiveness. Practical applications, however, are restricted by the existence of data collection systems, appropriate field hardware and communications systems, and operational support. At best, the use of data from signal systems to diagnose and identify oversaturation is sporadic and inconsistent. In this section we summarize some capabilities of existing traffic signal systems for identification of oversaturated conditions.

Flow Profile Estimation in the SCOOT Adaptive Traffic Control System

The SCOOT (The **S**plit, **C**ycle and **O**ffset **O**ptimization **T**echnique) (Hunt et al., 1981) adaptive traffic control system has capabilities and features designed to deal with oversaturation (Martin, 2006). As an adaptive system, SCOOT depends on good detection data so that it can respond to changes in flow. Detectors are normally required on every link. Their locations are important and they are usually positioned at the upstream end of the approach link (Other types of detectors such as stop-bar detectors can also be used to improve the traffic state estimation). These detectors (referred to as exit detectors) provide arrival flow information (*i.e.* input) to the SCOOT traffic state estimation module so that measures such as queue length can be determined (Figure A-7). To diagnose oversaturation, SCOOT estimates the degree of saturation by measuring the flow upstream of the stop-line and the "online saturation occupancy" measurement in lieu of a fixed saturation flow rate. The online saturation occupancy of a link is "the rate at which the queue modeled by SCOOT discharges from stop-line" (Bretherton and Bowen, 1990).

The saturation percentage (*%SAT*) in SCOOT is then estimated by following equation (Martin, 2006):

$$\%SAT = q \Big/ (STOC * g) \qquad \text{Eq. A-6}$$

where: g is the green time, q is the combination of flow and occupancy; $STOC$ is maximum outflow rate of a queue over the stop line. The unit of q and $STOC$ are LPU (link profile unit) and LPUs/sec, respectively. LPU is an internal unit defined in SCOOT, where one vehicle is equivalent to approximately 17 LPUs, and a saturation flow of 2000 vehicles/hour is equivalent to a SCOOT saturation occupancy (STOC) of 10 LPUs/Second. Oversaturated conditions are indicated by high *%SAT* (larger than 100%).

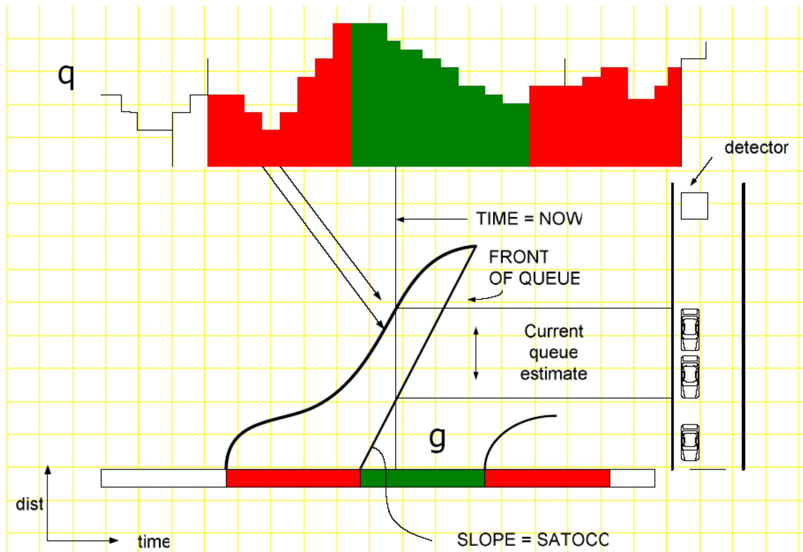


Figure A-7. Queue estimation in SCOOT (Source: Martin, 2006)

Degree of Saturation Measurements from the SCATS Adaptive Traffic Control System

Unlike SCOOT, the SCATS (**S**ydney **C**oordinated **A**daptive **T**raffic **S**ystem) (Sims & Dobinson, 1980) adaptive traffic control system only uses stop-bar detectors as "its primary source of data and so has no information about near-future arrivals" (Dineen, M., 2000). Compared with SCOOT, Dion and Yagar (1996) point out that "whereas SCATS is reactive to short-term traffic fluctuations it does not have SCOOT's predictive capability and is therefore less proactive". SCATS uses the Degree of Saturation (*DS*), *i.e.* the "ratio of the effectively used green time to the total available green time", to diagnose the level of congestion for a given traffic phase. *DS* is estimated by:

$$DS = \frac{NF[g-(T-t*n)]}{g+r} = \frac{NF[g']}{g+r} \qquad \text{Eq. A-7}$$

where: *NF* is a bias factor (weighting factor); g is green time; *T* is Total non-occupancy (space) time; *t* is space time which is unavoidably associated with each vehicle; *r* = remaining (or unused) phase time; and *g'* is effectively used green time (Dineen, M., 2000). This approach requires a short detector and very accurate measurement of occupancy since during oversaturation, the gap times between vehicles can be very small. The ratio can be > 1 only when T < t*n, which indicates that there is less actual "space time" remaining than what would be available if the phase served the saturation flow rate for the entire green time. This is not possible to occur unless r=0 (there is no unused phase time since the phase is forced off (maxed-out) and did not gap-out).

Figure A-8 shows examples of degree of saturation profile from the SCATS (Martin, 2006). In the figure, the measurements of degree of saturation are indicated as the vertical lines (green) for each traffic cycle. A traffic phase is estimated to be oversaturated when the green vertical line exceeds the horizontal (red) line for each phase. Notice in the bottom right corner of the figure that the cycle time of this intersection is gradually increased (and decreased at times) in conjunction with the measurement(s) of oversaturation on phases at this intersection. Notice also how noisy the cycle-by-cycle estimates of DS are over time, particularly for left turn phases (1, 3, 5, and 7).

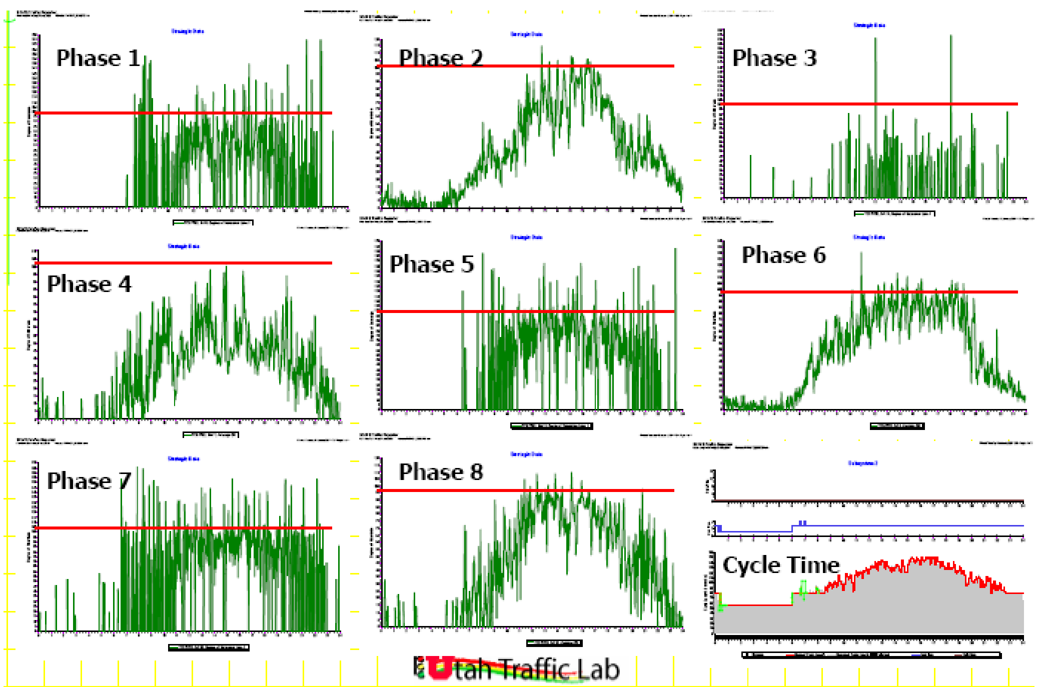


Figure A-8. Degree of saturation estimates over time from the SCATS adaptive traffic control system (Source: Martin, 2006)

Flow Profile Estimation in the OPAC Adaptive Traffic Control System

The OPAC (**O**ptimized **P**olicies for **A**daptive **C**ontrol) (Gartner, 1983) adaptive control system implements a rolling horizon strategy to optimize signal timing by predicting flow profiles as well as predicting queues that will result due to various potential changes to the signal timings (Gartner et al., 2002). Based on loop detectors placed upstream of each approach, a flow profile is developed for each phase (Figure A-9). The head of the profile is actual counts from upstream link detectors; and the tail of the profile is projected for the near future using a simple model consisting of a moving average of all past arrivals on the approach.

A simple input-output method is applied by OPAC to predict queue information, i.e. the estimated queue length is the sum of the initial queue length plus the difference of arrival and departure of each interval in the stage. Clearly, the accuracy of prediction highly depends on the location of upstream detectors which provide the arrival information. Placing the detectors well upstream of the intersection (10 to 15s travel time) will allow actual arrival information to be used for the head period; thereby increasing the accuracy of following predictions. However, such method cannot deal with oversaturated conditions as arrival information is missing once queue spillovers upstream detector. There is little discussion in available literature of any particular features of

OPAC or algorithm enhancements that allow OPAC to estimate oversaturation or adjust its flow profile estimates during congested conditions.

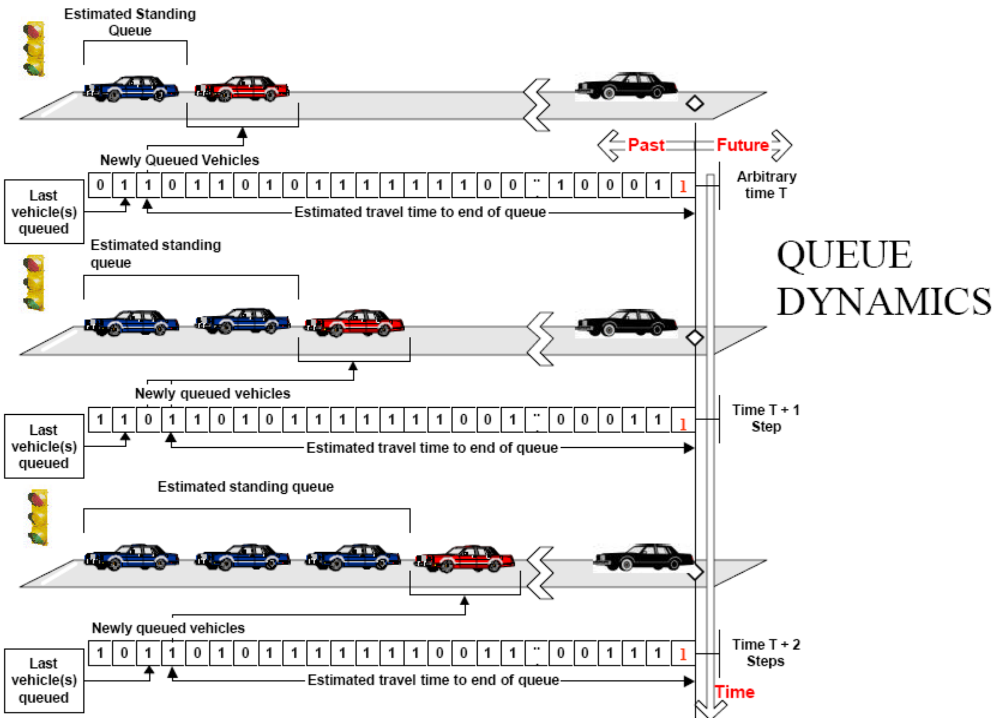


Figure A-9. Flow profile of OPAC (Source: Presented by Nathan H. Gartner on the workshop of Adaptive Traffic Signal Control Systems, TRB, Jan. 7, 2001, Washington D.C.)

Oversaturated Conditions Measurement in the RHODES Adaptive Traffic Control System

The RHODES (**R**eal-time **H**ierarchical **O**ptimized **D**istributed **E**ffective **S**ystem) (Mirchandani & Head, 2001) adaptive traffic control system uses the output of the detectors on the approach of each upstream intersection, together with the traffic state and planned phase timings from the upstream signals to predict future arrivals at the intersection. These inputs are used in the "QUEUE" algorithm (Head, 1995), a simple input-output estimation procedure to determine the queue length of traffic for each control phase. In detail, the queue at cycle t_1, $q(t_1)$, is equal to the residual queue at t_0, $q(t_0)$, plus the predicted arrivals, $a(t_1,t_0)$, and minus the estimated departures, $d(t_1,t_0)$, using a given queue discharge rate (Eq. A-8).

$$q(t_1) = q(t_0) + a(t_1,t_0) - d(t_1,t_0) \qquad \text{Eq. A-8}$$

To keep biases from entering into the estimates, RHODES identifies some certain epochs that the queue length is zero on the basis of the information from stop-bar presence detector (Head, 1995). RHODES handles queue estimates that extend beyond the location of upstream detectors by allowing a traffic agency operator to tune the saturation flow rates of the queue discharge process by time of day. As such, during rush hours the estimated queues can be made to dissipate more slowly on oversaturated approaches allowing the baseline adaptive algorithms to operate as if the intersection were still under-saturated on all approaches. RHODES also includes the ability to weight the delay estimates on one traffic phase higher than the delay on other traffic phases at the same intersection (and adjust the weights by time of day). This feature also helps to overcome the limitations of the queue estimation algorithm during oversaturated conditions by allowing certain phases that are known to have more traffic to be serviced with longer green times. In Pinellas County, FL, the RHODES installation on SR-60 has shown vast improvements during heavy arterial flow scenarios utilizing these key enhancements.

Although these adjustments can improve the effectiveness of the adaptive algorithms during oversaturation, RHODES, along with most other model-based adaptive systems, recommends placement of additional detectors at the upstream ends of links (i.e. exit detection) similar to the detection requirements for SCOOT. Extending detection placement in this way allows estimation of queues that extend to the entire link length. Of course, such additional detection comes with an associated increase in total system costs.

Green Utilization Measures in the ACS-Lite Adaptive Traffic Control System

Similar to the SCATS adaptive signal control system, ACS-Lite (**A**daptive **C**ontrol **S**oftware **Lite**) (Luyanda et al., 2003) uses phase utilization as it's indicator of degree of saturation for each traffic phase. Phase utilization is calculated based on second-by-second occupancy data from stop-bar detectors. Combining phase timing information with volume/occupancy data, the average used green is calculated (Figure A-10). Phase utilization is then used to compare the need for additional green time and availability of excess green time of each phase on the controller. In Figure A-11, Phase 1 is shown to be saturated (red color) since the entire phase split is used over the last seven cycles and occupied with traffic the entire time.

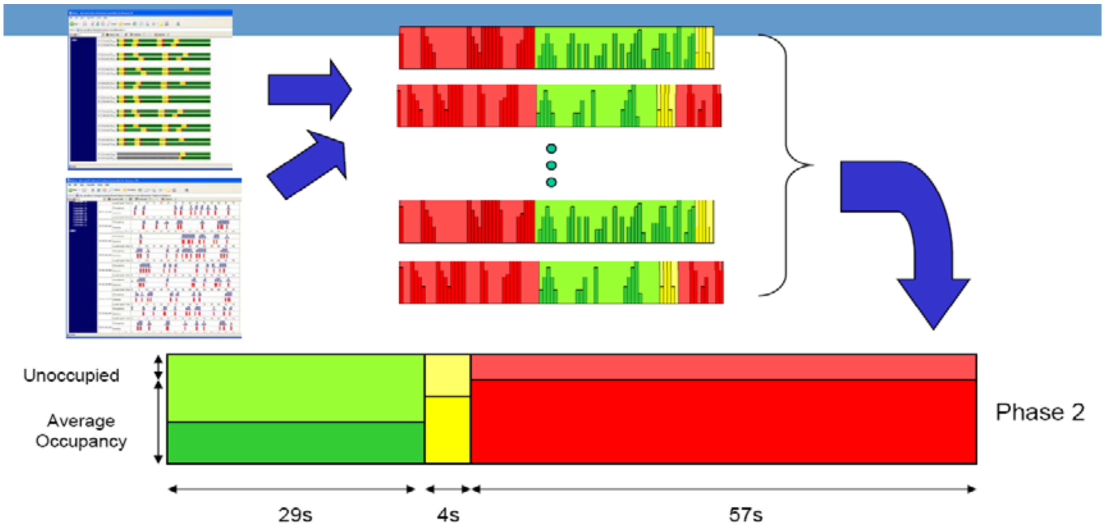
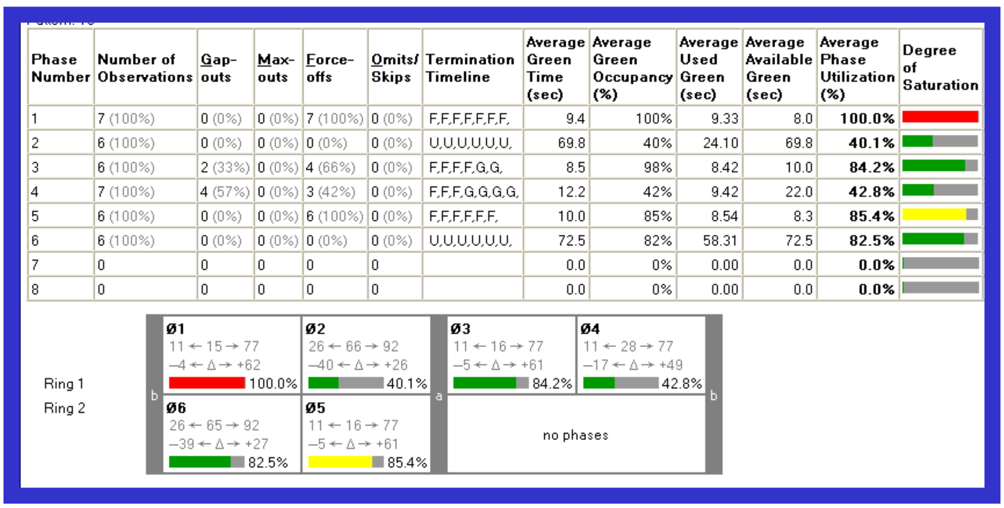


Figure A-10. Phase utilization measures from the ACS-Lite adaptive control system (Source: Gettman, 2005)



Phase Number	Number of Observations	Gap-outs	Max-outs	Force-offs	Omits/ Skips	Termination Timeline	Average Green Time (sec)	Average Green Occupancy (%)	Average Used Green (sec)	Average Available Green (sec)	Average Phase Utilization (%)	Degree of Saturation
1	7 (100%)	0 (0%)	0 (0%)	7 (100%)	0 (0%)	F,F,F,F,F,F,F,	9.4	100%	9.33	8.0	**100.0%**	
2	6 (100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	U,U,U,U,U,U,	69.8	40%	24.10	69.8	**40.1%**	
3	6 (100%)	2 (33%)	0 (0%)	4 (66%)	0 (0%)	F,F,F,F,G,G,	8.5	98%	8.42	10.0	**84.2%**	
4	7 (100%)	4 (57%)	0 (0%)	3 (42%)	0 (0%)	F,F,F,G,G,G,G,	12.2	42%	9.42	22.0	**42.8%**	
5	6 (100%)	0 (0%)	0 (0%)	6 (100%)	0 (0%)	F,F,F,F,F,F,	10.0	85%	8.54	8.3	**85.4%**	
6	6 (100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	U,U,U,U,U,U,	72.5	82%	58.31	72.5	**82.5%**	
7	0	0	0	0	0		0.0	0%	0.00	0.0	**0.0%**	
8	0	0	0	0	0		0.0	0%	0.00	0.0	**0.0%**	

Figure A-11. Phase utilization measures from the ACS-Lite adaptive control system (Source: Gettman, 2005)

ACS-Lite currently contains no explicit handling of oversaturation or adjustment factors to approximate oversaturation on a particular traffic phase. Green utilization is scaled from 0-100%. Like SCOOT, and to some extent OPAC, ACS-Lite includes the measurement of a cyclic flow profile but it does not use this information to estimate oversaturation. Occupancy data from upstream advance detectors (~300ft from the stop bar) are used to measure when the traffic on an approach arrives to the signal during red or green. ACS-Lite includes no explicit handling of oversaturated conditions, but visualization screens can clearly indicate when an approach is saturated as the occupancy percentage typically is not reduced to less than 100% during the cycle.

Queue Estimation in the German ACS/BALANCE/MOTION Adaptive Control Systems

ACS, BALANCE, and MOTION are all adaptive control systems developed in Germany that are based on the estimation of queues to recognize the traffic state (Mueck, 2005). Some progress has been made in recent years in these systems for estimating queues that grow past the detector location. A typical European traffic management agency places detectors 50-100ft upstream of the stop-line (Mueck, 2002). Mueck's algorithm for estimating queue length is based on the measurement of the so-called fill-up time. The fill-up time is "from the beginning of the red time of a signal until continuous occupancy of a detector" (Mueck, 2002). The fill-up time is found to be correlated to the level of congestion on an approach. Essentially, this algorithm identifies that when there is a residual queue that cannot be discharged fully during the green time, the fill-up time to 100% occupancy on the detector is much faster than the fill-up time when the queue is fully discharged and another platoon (or turning traffic from side streets) arrives at the stop bar. When the fill-up time "falls short depending on the distance between detector and stop-line" (Mueck, 2002), it is determined that a residual queue is forming and that the approach is now "congested", such that:

$$\delta = \begin{cases} 1 & dt \le dt_0 \\ 0 & dt > dt_0 \end{cases} \qquad \text{Eq. A-9}$$

where dt is the measured fill-up time; and dt_0 is the reference value. Then, this method estimates the residual queue length by estimating the growth of the surrogate "congestion characteristic" $\overline{\delta}_n$ (for cycle n) using an exponential smoothing method:

$$\overline{\delta}_n = a_\delta \delta_n + (1 - a_\delta)\overline{\delta}_{n-1} \qquad \text{Eq. A-10}$$

Empirical data collected in Germany shows a roughly linear relationship between the maximum back-up length and the congestion characteristic (Figure A-12).

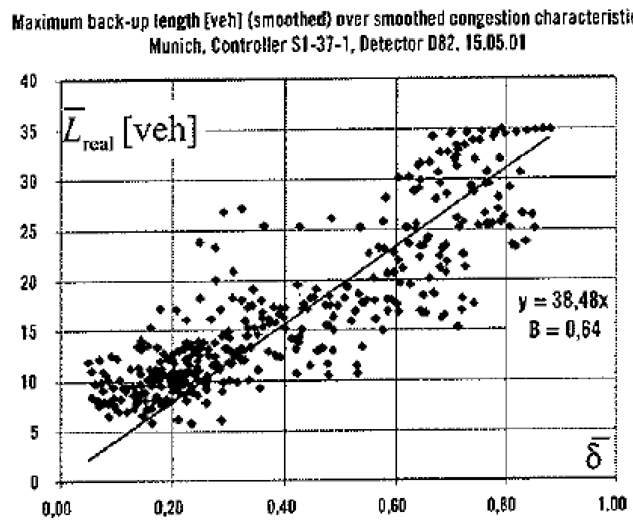


Figure A-12. Maximum back-up length (veh) over smoothed congestion characteristic (Source: Mueck, 2002)

Then, the back of queue length can be estimated by Eq. A-11 based on the congestion characteristic estimate.

$$L_n = m\overline{\delta}_n \qquad \text{Eq. A-11}$$

where *m* is a multiplicative gradient factor, which should be determined for each detector individually based on its location from the stop bar and its level of sensitivity.

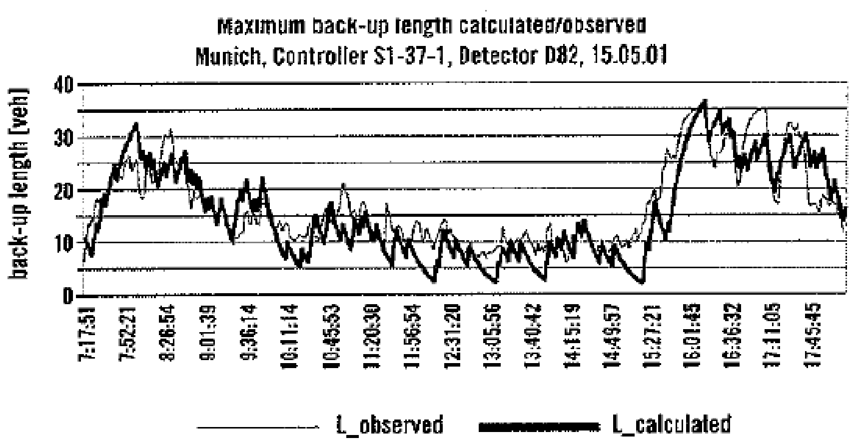


Figure A-13. Evaluation of fill-up time queue estimation (Source: Mueck, 2002)

Figure A-13 shows remarkable results for estimating queues that extend up to 5-10 times further upstream from the actual detector location (Mueck, 2002). However, this heuristic approach is based upon the premise that arrival rate of traffic flow is constant within a cycle; therefore the measured fill-up time can be proportionally inverse to the queue build-up time. Such an approach could be problematic when the arrival rate of traffic flow fluctuates greatly, which is not unusual with the gating effect of an upstream traffic signal.

Link Load in the TUC Adaptive Control System

The TUC (**T**raffic-responsive **U**rban **C**ontrol) adaptive traffic control system (Diakaki et al., 2003; Dinopoulou et al., 2005; Kosmatopoulos, et al., 2006) was developed to provide coordinated, traffic-responsive control in large-scale urban networks, even in cases of saturated traffic conditions. TUC includes four distinct control modules: split control, cycle control, offset control and public transport priority. The system is dependent upon real-time measurement, *i.e.* average number x_z of vehicles within each network link *z* over a cycle; and occupancy measurement o_z measured by a traditional detector in each link *z* is utilized to estimate x_z by a suitable function:

$$x_z / x_{z,\max} = f(o_z, \lambda_z) \qquad \text{Eq. A-12}$$

where $x_{z,\max}$ is the maximum number of vehicles that can be in link *z*; λ_z is the proportion between the detector distance from the stop line and the whole link length (an indicator of detector location); and *f* is defined as a suitable function. The available literature on TUC does not describe such an important function in detail.

The proportion of x_z and $x_{z,\max}$, is defined as the link load in TUC and actually indicates the degree of saturation of link *z*.

NCHRP 3-79

NCHRP 3-79 provided a comprehensive summary of the current state of the practice on urban street performance measurements including delay, queue, travel speed, and travel time. Most of their results are concluded in their interim reports (Bonneson, 2005); and some are published in journals (Sharma et al., 2007; Sharma & Bullock, 2008). The 3-79 project proposed two techniques for measuring queue lengths and corresponding saturation level of an arterial traffic link.

Input-Output and Hybrid Techniques for Delay and Queue Length Measurement

Input-output (I-O) and Hybrid input-output (Hybrid), are proposed for measuring queue lengths and delays. The basic concepts are similar; i.e., by utilizing the information of the arrival and discharge flow rates, delay and queue length at a signalized intersection can be quantified using input-output analysis (Figure A-14). The I-O technique only requires one upstream detector, which is used to measure the arrival flow rate. Combining this information with the signal phase status, estimated turn proportions, and saturation flow rate, this technique can develop a queue polygon and estimate delay for the cycle. The hybrid technique extends this approach to utilize two sensors located upstream and at the stop-line to directly measure the arrival and departure rates, respectively. Figure A-15 describes the process of both techniques (Sharma et al., 2007).

The accuracy of both techniques highly depends on the input flow measured by the upstream sensor and thus the location of such a sensor is crucial. Generally, this sensor should be located sufficiently far back in order to avoid frequent queue spillback, but at the same time, the sensor should not be located so distant that "driveway activity between the sensor and the stop line seriously degrades the accuracy of the predicted arrival flow profile" (Bonneson, 2005).

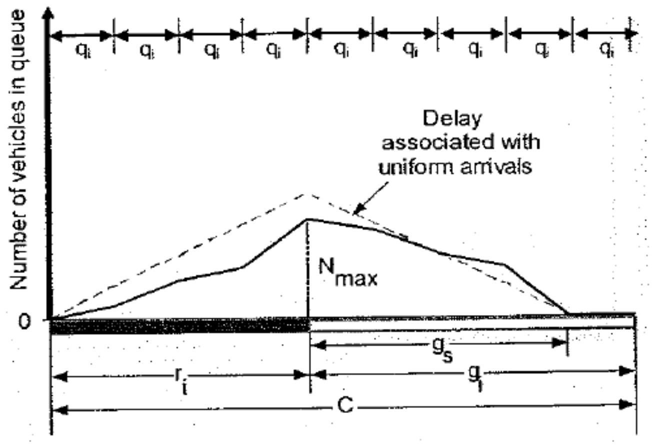


Figure A-14. Delay & Queue Polygon using Input-output Process (Source: Bonneson, 2005)

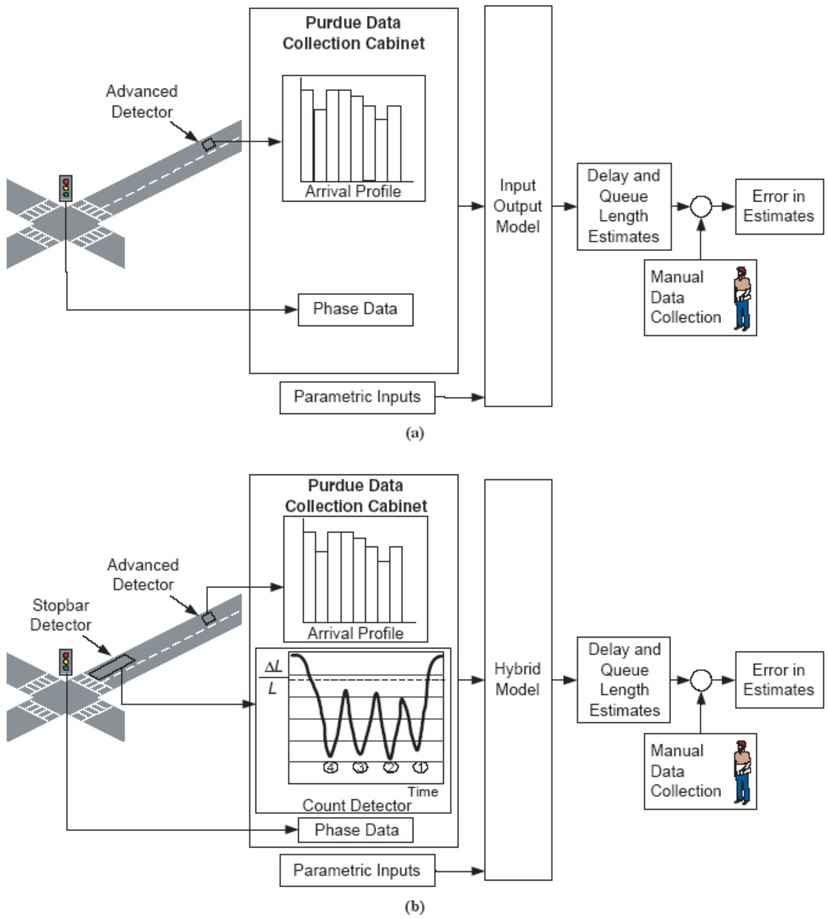


Figure A-15. Process (a) I-O; (b) Hybrid (Source: Sharma et al., 2007)

Non-Intrusive Detection for Queue Length Measurement

The second technology proposed by the NCHRP 3-79 research team to measure queue length is video detection. This approach is based on the use of a series of video detection zones, which monitor a segment of an intersection link, with coverage of all approach lanes. Queue length is indicated by the number of detection zones that are fully occupied during the red interval. Delay can then be estimated from this information on the length of the queue (Bonneson, 2005). The advantage of such a technique is that "they can monitor a length of roadway"; but it is acknowledged that the accuracy of the video detection technology degrades with increasing distance from the sensor (Bonneson, 2005). The placement of most video detectors that are used for phase actuation is such that queues could not be sensed to extend more than 300-400ft from the stop bar.

Bus Probe for Delay and Running Time Measurement

This technique is based on "the use of a transit-based automatic vehicle location (AVL) system for running time and overall travel time measurement" (Bonneson, 2005). The AVL data can provide the arrival time of buses at key points along major urban streets which can then be converted into segment-based travel time. An estimate of delay, and perhaps the length of standing queues could then be made by comparing the running time with the free-flow running time. One major issue related to this technique is the frequency of bus stops since this frequency will affect the accuracy of travel-time estimation.

Traffic Flow Characteristics for Running Time Measurement

This proposed technique uses measured traffic volume to estimate average running speed according to a calibrated relationship of speed-volume. The running speed is then converted to running time by dividing speed into segment length. This technique requires separate detectors for each traffic lane to accurately measure the traffic volume as shown below in Figure A-16.

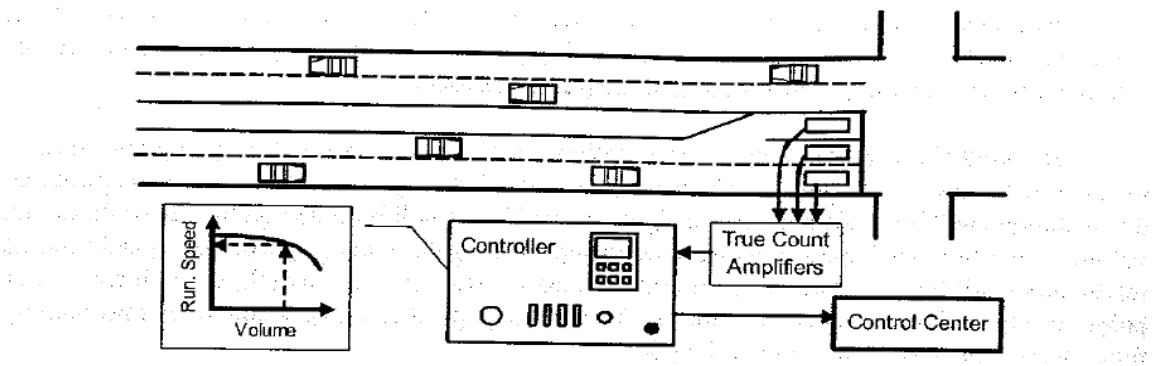


Figure A-16. Framework for estimating running speed using measured traffic flow characteristics (Source: Bonneson, 2005)

Control Strategies for Oversaturated Conditions

Control strategies for managing traffic congestion in oversaturated conditions seek to improve facility capacity or reduce facility demand. Enhancements to the roadway capacity can be achieved by increasing the physical capacity of the roadway system or by maximizing the operational capacity by improving the traffic signal controls. On the other hand, demand reduction might be achieved through restrictive control procedures such as metering or by using macro-level measures aimed at reducing vehicle use or influencing travelers to make significant modifications to their travel mode, departure time, route or destination (Quinn, 1992). This project does not address demand side management strategies such as influencing mode choice or providing ATIS routing messages and, as such, we will not review such strategies.

Several signal control strategies specifically designed for management of oversaturated conditions have been described in the literature. These strategies span a range of characteristics from being static or dynamic, reactive and proactive, having single objectives or multiple objectives, and applied for isolated intersections, closely-spaced intersections, diamond interchanges, arterial networks, and grids. This section focuses on review and critique of the algorithms and strategies that have been previously developed and described in openly-available literature.

Strategy Taxonomies for Managing Urban Congestion

Taxonomies for congestion management strategies have not been explored in much detail in the past. Most of the research that was reviewed focused on specific types of applications without a detailed effort devoted to framing the context of the problem.

Pignataro et al. (1978) classified strategies designed to address urban traffic congestion into:

- Signalized approaches: **minimally responsive** policies
- Signalized approaches: **highly responsive** policies
- Non-signal-related approaches: **demand management**, transit operations, etc.

Huddart and Wright (1989) classified <u>congestion prevention treatments</u> into:

- **Static** protection measures
- **Dynamic** protection measures
- Traffic **input control (**otherwise known as **demand management)** measures.

Pignataro et al. (1978) 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 methods:*
    (a) Intersection: Cycle length, Block length, splits, and extra phases

(b) System: Phases adjusted for progression, equity offsets, splits allocated to available queue storage.

- *Highly responsive methods:*
    (a) Intersection: Maximum queue policy
    (b) System: Accommodate queue spreading from intersection

2. Non-signalized treatment

- *Regulatory measures:*
    (a) Enforcement of exacerbating conditions (e.g., double parking)
    (b) Prohibition of exacerbating traffic movements (e.g., parking, turning)
- *Operations measures:*
    (a) Enhancement of turning facilities (e.g., left bays, right bays, dual turning lanes, right-turn-on-red)
    (b) Enhancement of lane arrangements (e.g., one-way street, reversible lane)
    (c) Strategies for handling disruption (e.g., pedestrian, bus stop, parking/ no-parking, mid block)

No other attempts to categorize types of strategies were found in the literature. Neither of these taxonomies is exhaustive or particularly systematic in its approach to categorizing strategies. This lack of a good taxonomy of types of strategies mirrors the lack of specificity in the qualitative approaches applied to the definition of types of congestion or oversaturation. In the following sections, we detail strategies for managing oversaturated conditions found in the literature in order from simple to more complex. This organization of strategies generally proceeds from approaches that deal with individual approaches or intersections to strategies that address network conditions.

Switching of Green Time

Dunne and Potts (1964) developed an algorithm for controlling an isolated intersection based on a closed-loop feedback control concept. Green phases were changed according to the variation in traffic demand. The concept was further improved by Gazis and Potts (1965) for traffic signal control in oversaturated conditions. Gazis and Potts' control philosophy was to minimize total system delay by maximizing intersection productivity. In this approach, a maximum green time is provided to the approach with the higher saturation flow until the queue on that approach is dissipated, while other approaches receive a minimum green time. The control decisions fluctuate only between these two values, minimum green or maximum green, and the cycle time remains constant. This approach assumes that the traffic demand can be known, and that the demand does not change due to the control decisions (i.e. drivers do not re-route).

Michalopoulos (1975) used a reservoir analogy to improve the policy developed by Gazis by developing an algorithm that determines what the "switch-over" points should be. The "switch-over" concept in this case refers to the points in time at which the green duration of a certain approach should be changed from maximum green to minimum green allowing other approaches' green durations to change from minimum to maximum and vice versa (Gazis et al., 1968). Michalopoulos and Stephanopoulos then later (1977, 1978) proposed an efficient two-stage timing method (known as "bang-bang" control) to find the "optimal" switch-over point during the oversaturated period. Later, Chang et al. (2000) showed that the continuous delay model used by Michalopoulos and Stephanopoulos is inadequate in identification of the optimal cycle length since a penalty for stops is not considered in the model's formulation.

Khakzadi (1980) demonstrated that the optimal control of an intersection cannot be achieved by only switching green times from a maximum to a minimum green splits while the cycle remained constant. By allowing the cycle length to vary within specified upper and lower limits, Khakzadi's improved algorithm identifies the green allocation policy that minimizes delay and limits the queue growth rates. This formulation again assumes exact knowledge of the arrival rates over time.

Chang and Lin (2000) further improved the Gazis model by developing a discrete optimization algorithm that uses the bang-bang-like control method to determine the switch-over point. Unlike in Michalopoulos and Stephanopoulos' continuous model, the Chang and Lin model can allocate the switch-over points exactly at the end of a cycle. According to the authors, the discrete operation provides a smooth, regular, and ordered transfer of control and reliable calculation of delays. The bang-bang-like control operates alternatively and sequentially with given minimum and maximum green times. According to Chang and Lin's study, such a control strategy can significantly outperform a policy that only increases the cycle time and gives each phase the same amount of additional green time. Chang and Lin indicate that the performance of their algorithm is robust even when the input data have some measurement error, but the approach still requires a trajectory of given, known input volumes.

Potential Application to Practice

These approaches that switch green times between minimum and maximum values must also take into account that these times are typically regulated by laws in many agencies (e.g., pedestrian crossing time). Such regulations could be considered as constraints in the mathematical formulations. In general though, the main limitation of such policies is that the optimization algorithm is driven by knowledge of the exact demands during the oversaturated period. However, the policy where the maximum possible green is allocated to the most congested approach is an intuitive one. Such a determination could easily be made with a queue estimation algorithm for each approach or movement at an intersection.

Maintaining Queue Ratio

Longley (1968) introduced the concept of a real-time "queues proportionality" strategy. Longley suggested that green times should be adjusted to balance the queues at the approaches to an intersection during oversaturated conditions. The objective is to minimize the number of intersections that are blocked by queues. Longley first considered the dynamic behavior of an isolated intersection and then extended the analysis 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 the parameters of each controller. The policy, however, requires exact measurements of queue lengths.

Queue-Actuated Control (Extended Green Time)

Miller (1965) 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. This concept appears to have motivated much of how existing modern actuated-coordinated systems operate today in undersaturated conditions.

The queue-actuated control was claimed to be effective in high volume levels in general and, in particular, intersections with high turn-in volumes. Miller's concept was later expanded by Weinberg (1966) to include downstream delays in the computation of total delay. Including the downstream delay in the formulation provides coordination between the critical intersection and its neighboring intersections to ensure that the reduction in critical intersection delay would not come as a cost of increased delay at other intersections. Weinberg's model was modified and evaluated later by Ross et al. (1971) via simulation. The simulation results indicated that the policy is effective to reduce total delay during saturated conditions. The policy implementation required special treatment regarding the allocation of detectors.

Strategies for Coordinated Intersections

For under-saturated traffic conditions the offsets between green phases can be determined based on the objective of maximizing the progression band. As congestion increases and intersections become more saturated, residual queues (and 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-parameter 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 (May, Montgomery and Quinn, 1988). This means that conventional procedures for optimizing fixed-time signal control, such as TRANSYT (Robertson, 1969), deteriorate rapidly when severe congestion persists.

Queue Length Control

Gordon (1969) was the first to introduce the concept of optimizing the green times to utilize 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 by delaying the service of a particular approach as long as possible.

Offset Design Schemes for One-Way Progression

Pignataro et al. (1978) 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 (or "negative") progression, the offsets are set to allow the green of a downstream intersection to starts before the upstream green in order to flush the residual queue. In addition, Pignataro et al. recommended using the following principles to determine the signal split at the upstream intersection on an oversaturated arterial where a "reverse progression" offset is being applied:

1. 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 it should not limit the cross-street through movement. Pignataro et al. recommended physical capacity improvements (e.g. turn lane bays and signalization) in this case.
2. In allowing such cross-street movements during oversaturation, it should be recognized that the arterial movement needs requires only as much green as it can use effectively use at the downstream intersection. Any additional green is, in fact, wasted but could be allocated to the arterial to restrict the cross turning movement (if signalization by movement does not exist).
3. If possible, turn prohibitions should be considered on the cross street so that the crossing through movement is not impacted by turning vehicles that are queued on the side street.

Pignataro et al. (1978) also suggested a queue management strategy that was termed "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 for determining the intersection splits: Where (1) there are negligible turn-ins from the cross (minor) streets at the upstream intersection, and (2) where the split at the intersection can be as commonly determined (e.g. Webster's delay minimization formula). Where there are substantial turn-ins from the cross streets at the upstream intersection, the cross street traffic should be allowed to have 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. (1978) did not provide mathematical equations to cover all combination of approaches 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 (1989) described a procedure that uses simultaneous and negative offsets along the arterial to control signals and manage queues and negative offsets and flaring of green along cross roads.

Recent research by Baird, et al (2007) and Smaglik, et al (2007) 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. Thus it is better to truncate the green time for the current phase and move on to another phase.

Models for Queue Interactions between Closely-Spaced Intersections

Rouphail and Akcelik (1992) 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 due to interference from a downstream queue. 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 (1994) 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 an upstream intersection will discharge when queue spillback occurs downstream.

Messer (1998) 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.

Strategies for 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 (Kovvali et al., 2002). Kovvali, et al developed an extension for 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 agency engineers tend to favor strategies that cause queue formation on the cross streets in order to enhance progression. In the case of diamond interchanges, these cross-streets are the freeway off ramps or frontage roads. Meanwhile, state highway officials prefer strategies that control queue length on the off-ramp so vehicles do not back onto the mainline freeway facilities and impede their operation, possibly at the expense of the arterial street network (Herrick and Messer, 1992). Queue spillback and oversaturated conditions at interchanges can have significant detrimental effects.

Signal control at diamond interchanges has traditionally has been provided by either a three-phase signal sequence or the four-phase, two-overlap signal phase sequence (TTI-4-phase). The three phase sequence favors progression for the arterial through movement. Three-phase control requires sufficient queue storage space for the left-turn and cross-street 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 very limited and thus queue spillback in between the two intersections would be a major concern (Tian, 2006).

Texas Urban Diamond Signal Control was developed by Texas Transportation Institute (Herrick and Messer, 1992). 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 of the through movements on the arterial, typically preferred when the ramp traffic is balanced and the arterial left-turn traffic volumes are low. The four-phase 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 based on a time of day schedule.

The Arlington method was developed by the City of Arlington, Texas, for diamond interchange control under saturated conditions. The Arlington approach uses a 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 Arlington method 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 (1992) indicate that the additional phases initiated by activation of queue detectors 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 (1992) 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 knowledge of the traffic demand profile. Therefore, any significant variation in the actual traffic demand may render the signal timing plans established by this optimization approach ineffective.

Real-Time 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. 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. 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 online decision-making. The main benefit of applying adaptive systems to oversaturated conditions is that they delay the onset of the oversaturation and can more effectively remove the conditions during the recovery regime.

SCOOT

The SCOOT (Bretherton, R.D., 1989) control system includes algorithms for dynamic control of individual intersections, arterials, and grids/networks. The core algorithms of SCOOT use a link flow profile (a composite representation of volume and occupancy) to tune the cycle 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 was not able 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 (Bretherton and Bowen, 1990) many features have been added to SCOOT to tackle the problem of address severe congestion (Martin, 2006):

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

Initially, both SCOOT and SCATS have algorithms that allow 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. In addition, each cycle in SCOOT the split optimizer can increase or decrease split times by a small amount (+/- 4 seconds) to keep all of the splits below 90% saturation (which could be a user-defined value, but not typically modified). If re-allocation of the split time cannot accomplish this, then an increase to the cycle time is recommended. One criticism of this policy is that it tends to increase the cycle time of many intersections for oversaturation on approaches that may have only minor flows (i.e. side-streets). Similarly, balancing saturation levels favors side street delays at the expense of progression on arterials and coordinated operation. To accommodate this, SCOOT includes the concept of split “weighting” that allows the split optimizer to consider certain phases as more important than others when considering adding additional split time. A similar concept was implemented in the ACS-Lite adaptive system (Gettman, et al 2007) as well as RHODES. Split weighting is a preventative measure to delay the onset of congestion on key links.

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.

SCOOT also includes a number of features to improve progression performance during heavy flows. Initially, SCOOT (and other adaptive approaches including ACS-Lite, 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. 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 that offset value to accommodate changes in the traffic flow. Finally, SCOOT includes the concept of a “congestion offset”. A congestion offset is one that is pre-determined to be used when a certain link is congested. Congestion is defined as occupancy over the upstream detector on the link (at the exit point) continuously for several seconds (e.g. 15s). When this condition is satisfied, 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.

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, 2006) 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. For the purpose of 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.

As a natural extension to all of these other 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. 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.

SCATS

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 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 P.M. peak periods on arterials. Pattern matching is used to select which set of offsets is used under any given pattern.

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 (Quinn, 1992).

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:

1- Forced and hold green
2- Gating and metering
3- Maximum capacity flow
4- Negative offset- reverse green waves
5- Green waves with cross streets
6- Flared green with cross street (increases the green time windows at downstream intersections)
7- Diversion away from congestion
8- Shorter cycle length
9- 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 (1992) suggested that the two techniques could be used for improving operations on arterials.

Traffic Metering/Gating

Metering or gating strategies impede traffic input at a suitable point upstream to prevent the volume from reaching critical levels at downstream locations. These strategies can be applied either locally, to protect a particular intersection, or on an area-wide level. The practitioner must determine suitable links with enough spacing to store the metered traffic. The Gated links are those links which have been designated as links that store the queues which would otherwise block the bottleneck link(s). Rathi (1991) discussed queue management control strategies that meter the rate of flow into and within high-traffic density networks (control areas). He broadly categorized metering control strategies as “internal” metering and “external” metering approaches.

Internal Metering strategies include 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, restrict the inflow of traffic along the periphery of a control area, while servicing demand at an acceptable level to improve the overall quality of traffic flow within the control area. The overall performance of the affected traffic should be improved. Rathi and Lieberman (1991) indicate that the external metering control strategies have the potential to improve traffic operations within and on the approaches to a congested control area (Quinn 1992).

Although metering strategies appear to have potential, many factors against implementing metering control schemes should be considered. For example, there are likely adverse impacts on business in the affected area and equity concerns since not all metered road users are necessarily contributing to the congestion. It can also be argued that the effect of metering may simply transfer the congestion from inside to the outside of the control area without any change in overall travel time, air quality improvement, or other benefits. There are other implementation issues with metering, including the existence of sufficient storage, the potential of re-routing of delayed traffic, and the potential for increased red-light violations at the queue storage locations. Quinn also cautions that while traffic metering can be applied to protect a busy area from the sudden influx of morning peak traffic, it is less likely to be effective during the evening peak when much of the traffic originates from within the control area (e.g., parking garage) (Quinn 1992).

Lieberman et al. (2000) developed a real-time traffic control policy for congested arterials based on the concept of gating. The algorithm, denoted 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 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 turning movements at the intersection and precisely controls the duration of each phase 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 (2002) developed a procedure based on genetic-algorithms (GA) 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 online load-balancing mechanism to protect critical intersections from becoming oversaturated. Therefore, positive and/or negative progression strategies can be employed depending on the location of the critical intersections (i.e., entry or exit points). The 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 a longer duration at the entry signals than they would be at the exit locations. This will 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 effective green times that are optimal to provide for the queue-dissipation process. The algorithm requires an efficient use of the green time and avoidance of "de-facto" red. De-facto red exists when the signal is green but traffic cannot proceed because of backed-up traffic on a receiving street (Abu-Lebdeh et al., 2000). 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 apply it in a two-way-street network. 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, but not both. In a later effort, Girianna and Benekohal (2003) 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.

Recovery from Severe Congestion

Most, if not all, traffic signal control strategies were developed to prevent, or at least delay, 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, alternative methods are required. 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 (Quinn, 1992). These recommendations are not followed by detailed analytical algorithms to achieve these objectives.

Daganzo (1995) introduced a cell transmission model (CTM) to capture traffic flow dynamics. CTM is used as a numerical discrete approximation of hydrodynamic traffic flow theory (Lighthill, et al, 1955). Lo et al. (1999, 2001, and 2004) 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 (variable cycle times). 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 (2004) presented a dynamic control method incorporating a bang-bang control strategy to improve TRANSYT-7F’s performance in dealing with oversaturated signalized networks. Their method considers all the over-saturated conditions until all intersections become 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.

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:

1. Identify the queue and the queue discharge time.
2. Identify the downstream storage available for queue discharge.

3. 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. (1997) 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. (2003) 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. (2000) 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. (2002) 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.

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:

a) 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);

b) Short cycles are useful for clearing intersections blocked by turning traffic; and

c) 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).

Multi-Objective Analysis

In undersaturated 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. (1992) 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.

Sayers et al. (1998) 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. (2007) proposed using 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 oversaturated conditions.

Summary of Literature Review

Significant effort has been invested in studying oversaturated conditions in the traffic control and traffic modeling community. Most of the work has been focused on identifying how to measure delay or model the effects of oversaturation. These efforts have marginal utility for this research. It has long been recognized that queue measurement and/or v/c ratio estimation is the key to managing oversaturation. There have been a myriad of approaches for performing queue estimation, most of which are input-output models that can estimate queues downstream of the detection point but have little predictive capability if the queue extends past the point of detector placement. Several encouraging research projects such as (Mueck, 2002) were found that show promise for methods to estimate queues using point detection that do not assume “exit” detection is available. The work begun by Liu (Liu and Ma, 2007) and extended in this project, continues along this direction to provide capability to measure queue length without exit detection. This is important since the typical agency in North America cannot afford to install “exit” detection at every intersection (or even at a subset of critical locations). While adding some detection is unavoidable in most situations, the challenge is that adding additional detection, while helpful, does not inherently solve the problem of not being able to measure the input volumes when the queue has extended past the detection point. Additional detectors are needed upstream of that location, and so on, and so on. Strategically placed detection is important, but we believe that methods such as will be described in Task 2 are necessary (until, at least, we have ubiquitous IntelliDrive technologies on a significant percentage of passenger vehicles and trucks). Approaches based on this type of technology are outside the scope of this research.

Most of the remaining research in estimating oversaturated conditions has occurred within the context of the development of adaptive control systems. These systems must operate in both regimes (over and under-saturated conditions) and thus must at least approximate the effects of oversaturation. Because of the proprietary nature of these commercial applications, only limited open literature is available that describes their methodologies for estimation as well as strategies for control. Both SCOOT and SCATS approximate the degree of saturation with detector occupancy information. SCOOT models a link flow profile which is a combination of the volume and occupancy data at an exit detector location. SCATS models the level of saturation using short stop-bar detectors by counting the “spaces” between vehicles and comparing the amount of unused green to that amount of green that would be unused at saturation flow. Thus, SCATS can determine when a phase is oversaturated and has an estimate of how much, but does not directly model queue length. SCOOT more accurately approximates the queue on a link by modeling the dynamics of flow on the link. When the dynamics are not very dynamic, so to speak, SCOOT can estimate the severity of the congestion as compared to the link performance during saturation flow. ACSLITE and other systems that measure link flows at advance locations can approximate this type of oversaturation estimate, but have not developed algorithms to estimate the level of congestion or strategies to take advantage of the information.

Several theoretical formulations of control approaches ("optimal" controls) for oversaturated conditions have been developed. This includes off-line search algorithms extending popular design tools (PASSER and TRANSYT) for finding cycle, split, and offsets that are "optimal" for oversaturated conditions. The main drawback of most theoretical formulations is the assumption that demand volumes are known and measurable in the real-world, which is simply not the case during oversaturation. Only limited study of the transferability of these formulations is presented in the literature, thus it is not clear how to take advantage of these efforts in this research work. Only a few research projects surveyed in the literature seem to have developed manageable, real-world approaches to control strategies for oversaturated conditions. These projects are:

- Real-time internal metering policy to optimize signal timing (RT/IMPOST) was formulated explicitly for oversaturated arterials. RT/IMPOST controls the queue growth on saturated approaches by metering traffic utilizing the network storage capacity. (Lieberman, E. B., Chang, J. & Prassas, E. S., 2000). Although this method is driven by measurement of volumes, it may have value if it could be modified to accept queue length measurements as input values.
- Diamond interchanges operation strategies in oversaturated conditions, which include Texas urban diamond signal control, the "Arlington" approach, and the approach developed by Kim and Messer. These strategies basically seek to maximize the system throughput by switching between two-phase, three-phase, and four-phase operations based on traffic demand variations through the day. Texas diamond mode in either three-phase or four-phases is the only strategy that is known to have been implemented in the field. The Arlington approach was used many years ago in the City of Arlington and is not known to have been continued operation. Kim and Messer's approach never made it to the implementation stage. (Kim, Y. & Messer, C. J., 1992).
- Rathi et al. (1988) developed a control approach that avoids spillback in urban grid networks by adjusting signal offsets. The control approach was implemented in Manhattan CBD in NY (5th Avenue between 63rd and 54th Streets) and provided 20% reduction in overall travel time. Combinations of strategies were used (i.e. simultaneous offset for N-S streets, negative progression with flared green for E-W streets to avoid spillback and capacity maximization). (Rathi, A. K. et al, 1998). Since this methodology was implemented in the real world, it should be investigated further for inclusion into the list of potential application strategies.

- The i*f…then* queue management concepts in SCOOT and SCATS. These approaches for initiating gating operation and triggering jumps or reduction in cycle times are directly relevant to this research project. Even without the adaptive operations that slowly adjust timings to react to changes in flow, the if…then strategies can be fairly easily implemented by agencies to accommodate specific recurrent conditions with the deployment of just a few queue estimation detectors. The algorithms developed in Task 2 for estimating queues and oversaturation levels combined with the approach discussed for development in Task 7 can leverage the concepts that are used in SCOOT to apply to real-world situations.

Many other recent works have focused on individual intersections and simple rules for truncating phases due to downstream restrictions, or identifying that cycle times that should be just short enough to clear turn bay storage, but no longer. These methods are useful to include into the catalog of strategies, but need more systematic definition. Other research on strategies was found to be rather vaguely descriptive of a general approach such as "use negative offsets" or "use flared green times," but little additional detail is provided as to describe how to mathematically construct such strategies.

Research Directions from the Literature Review

Previous research in *diagnosis* of oversaturated conditions and strategies for addressing these conditions is not extensive. This is mainly due to the challenging nature of the problem and the limited capabilities of existing state-of-the-practice detection systems. Most models for diagnosis and evaluation of oversaturated conditions have volumes as an input variable. This is especially problematic since the limitations of existing detection systems make it rather difficult to measure volumes during oversaturation, and particularly when queues grow past the detection point. Thus, the various "optimal" control formulations that we found in the literature are difficult to apply or use in practice since they need this volume information that is highly unreliable.

We also found little to no research on what is an appropriate time for a queue to be persistent before a corrective action(s) needs to be taken. From these findings, we identified that it is important to develop diagnosis techniques that can identify queues that grow past the detection point (measurement of the extent of oversaturation) and identify quantitative measures for the relationships between queue lengths and green time allocation. Similarly, it will be important to characterize the degree to which a traffic facility is oversaturated, versus simply indicating that it is oversaturated or not. In particular it is important to identify when an oversaturated condition on one traffic facility (movement, link, or intersection) is detrimentally affecting the traffic flow on other facilities. This dovetails with the development of a framework for recommending appropriate strategies to practitioners since certain actions (such as phase truncation, for example) can have a direct and reverse effect on the intended outcome, if applied incorrectly. Similarly, if the bar is set too low to indicate when alternative control measures are required, actions might be implemented by practitioners when the condition is intermittent. Undue delays might be induced when the corresponding improvement to throughput is not significantly realized.

Literature on *strategies* for handling oversaturation could be criticized in several areas:

- not having enough description to replicate,
- requiring information on traffic arrivals and volumes that is not possible to measure, and/or
- not having been supported by sufficient research performed to identify transferability of the results to other networks or traffic scenarios.

Realized performance improvements due to application of strategies for handling oversaturated conditions have not in the past been described adequately. For example, the concept of gating or metering appears promising for further research although existing literature describes the concept very qualitatively. This also indicates that it is critically important to develop a diagnosis methodology that identifies when gating is warranted, and when it is not.

Simultaneous and negative offsets were also identified in the literature as promising strategies for arterials, although there is little research on the implementation issues such as when to implement a different offset pattern and how to mitigate the transition effects. Some recent work has focused on individual intersections and simple rules for truncating phases due to downstream restrictions, or identification that cycle times should be just short enough to clear turn bay storage, but no longer. These methods are useful to include into the catalog of strategies developed in this project, but they need more systematic definition.

Other more dated research on strategies was found to only provide general approach such as "use negative offsets" or "use flared green times", but little additional detail is provided as to describe the details of such approaches. The same is true of the methodologies used by adaptive control systems, notably SCOOT and SCATS and to some extent RHODES.

From these findings, we identified that it will be important in this project to provide quantitative evidence that certain strategies improve total throughput and other performance measures. In addition to this, the research will focus on study of the rationale necessary to identify when it is appropriate to switch objectives from minimizing delay (normal operations) to maximizing throughput, and mitigating the effects of queues.