

1. Report No. SWUTC/10/169113-1		2. Government Accession No.		3. Recipient's Catalog No.	
4. Title and Subtitle ANT COLONY OPTIMIZATION ALGORITHM FOR SIGNAL COORDINATION OF OVERSATURATED TRAFFIC NETWORKS				5. Report Date May 2010	
				6. Performing Organization Code	
7. Author(s) Emily Zechman, Luca Quadrifoglio, Rahul Putha				8. Performing Organization Report No. Report 169113-1	
9. Performing Organization Name and Address Texas Transportation Institute Texas A&M University System College Station, Texas 77843-3135				10. Work Unit No. (TRAIS)	
				11. Contract or Grant No. 10727	
12. Sponsoring Agency Name and Address Southwest Region University Transportation Center Texas Transportation Institute Texas A&M University System College Station, Texas 77843-3135				13. Type of Report and Period Covered	
				14. Sponsoring Agency Code	
15. Supplementary Notes Supported by general revenues from the State of Texas.					
16. Abstract Traffic congestion is a daily and growing problem of the modern era in mostly all major cities in the world. Increasing traffic demand strains the existing transportation system, leading to oversaturated network conditions, especially at peak hours. Oversaturation occurs when queues of vehicles fill the streets approaching intersections and interfere with the performance of adjacent upstream intersections. Traffic conditions, measured based on the overall throughput of vehicles and total travel time, can be improved by an effective employment of intelligent transportation system techniques. While a significant amount of research has been devoted to the development of signal control algorithms under normal traffic conditions, a relatively small number of studies have explicitly considered oversaturated conditions. The overall objective of this study was to investigate the effectiveness of the Ant Colony Optimization (ACO) algorithm in solving the traffic signal control problem under oversaturated conditions. Due to its ability to reach optimality conditions and identify acceptable solutions efficiently, ACO was a good candidate for a practical use. This research compared the performance of ACO to that of another heuristic method, the genetic algorithm (GA). The methods were applied to identify signal control strategies for two example networks. The results demonstrate that ACO was able to identify fit solutions more reliably than the GA-based approach.					
17. Key Words Transportation Planning, Traffic Signal Timing, Ant Colony Optimization, Genetic Algorithm			18. Distribution Statement No restrictions. This document is available to the public through NTIS: National Technical Information Service 5285 Port Royal Road Springfield, Virginia 22161		
19. Security Classif.(of this report) Unclassified		20. Security Classif.(of this page) Unclassified		21. No. of Pages 57	22. Price



# **Ant Colony Optimization Algorithm for Signal Coordination of Oversaturated Traffic Networks**

by

Emily Zechman  
Researcher, Texas Transportation Institute

Luca Quadrioglio  
Researcher, Texas Transportation Institute

Rahul Putha  
Researcher, Texas Transportation Institute

Report SWUTC/10/169113-1

Texas Transportation Institute  
Texas A&M University System  
3135 TAMU  
College Station, TX 77843-3135

May 2010



## ABSTRACT

Oversaturation of a traffic network occurs when the queues of vehicles on a receiving street interfere with the performance of the respective adjacent upstream streets. To improve system efficiency, traffic signals can be timed to maximize the number of vehicles processed by the network and minimize the travel time of the vehicles. While significant research efforts have developed signal coordination for uncongested systems, designs for over-congested systems have not been studied as extensively. Oversaturated policies should prioritize the dissipation of queues and removal of blockages over cost minimization.

A new algorithm for signal coordination along oversaturated arterials has been developed and reported based on a Simple Genetic Algorithm (SGA) approach. The signal coordination problem was formulated as a throughput maximization problem and the development of queues and dissipation was managed to prevent the formation of de facto red. Ant Colony Optimization (ACO) is a heuristic technique for solving computational optimization problems that can be applied to manage oversaturated systems. This paper reports the development and application of ACO to solve the oversaturation traffic network problem. Two traffic networks were considered, and SGA and ACO were applied to both case studies. Several scenarios were considered which varied the computational power allocated to solve the problem, and a comprehensive comparative analysis was performed. The results demonstrate that the performance of ACO compares favorably with that of SGA. It was observed that ACO yielded better results when more computational resources were available, and ACO identified more stable objective function values than SGA.

## TABLE OF CONTENTS

<b>LIST OF FIGURES</b> .....	<b>vii</b>
<b>LIST OF TABLES</b> .....	<b>viii</b>
<b>DISCLAIMER</b> .....	<b>ix</b>
<b>ACKNOWLEDGEMENTS</b> .....	<b>x</b>
<b>EXECUTIVE SUMMARY</b> .....	<b>xi</b>
<b>CHAPTER 1. PROBLEM INTRODUCTION</b> .....	<b>1</b>
<b>CHAPTER 2. BACKGROUND</b> .....	<b>3</b>
<b>CHAPTER 3. MODEL FORMULATION</b> .....	<b>9</b>
3.1 SIGNAL COORDINATION .....	9
3.2 OBJECTIVE FUNCTION.....	9
3.3 CONSTRAINT: IDEAL OFFSETS .....	10
3.4 CONSTRAINT: DE FACTO RED .....	11
3.5 CONSTRAINT: COORDINATED LOOPS .....	12
3.6 CONSTRAINT: QUEUE STORAGE CAPACITY .....	12
3.7 CONSTRAINT: CONTROL VARIABLES .....	13
3.8 NETWORK FLOWS.....	13
<b>CHAPTER 4. SOLUTION METHODOLOGY</b> .....	<b>15</b>
4.1 ANT COLONY OPTIMIZATION.....	15
4.2 GENETIC ALGORITHM .....	17
<b>CHAPTER 5. APPLICATION TO ILLUSTRATIVE CASE STUDIES</b> .....	<b>19</b>
5.1 MODEL I.....	19
5.2 MODEL II .....	20
<b>CHAPTER 6. RESULTS</b> .....	<b>23</b>
6.1 MODEL I RESULTS .....	24
6.2 MODEL II RESULTS .....	27
<b>CHAPTER 7. HYPOTHESIS TEST</b> .....	<b>31</b>
<b>CHAPTER 8. CONCLUSIONS AND RECOMMENDATIONS</b> .....	<b>35</b>
<b>REFERENCES</b> .....	<b>37</b>

## LIST OF FIGURES

<b>Figure 1.</b> One Way Traffic Network with 20 Signals.....	20
<b>Figure 2.</b> Downtown Network of the City of Fort Worth, TX .....	21
<b>Figure 3.</b> 4 × 4 network used for Model II .....	22
<b>Figure 4.</b> Convergence Plot of ACO showing the Fitness Function for the case Ants=400, Trials=50 .....	24
<b>Figure 5.</b> Box plots of objective function values (Eq. 14) found using GA and ACO for Model I.....	24
<b>Figure 6.</b> Box Plots of number of cars processed (first term of Eq. 1) using GA and ACO for Model I .....	25
<b>Figure 7.</b> Box plots of ideal offset (Eq. 2) using GA and ACO for Model I.....	26
<b>Figure 8.</b> Box plots of de facto red processed (Eq. 4) using GA and ACO for Model I .....	26
<b>Figure 9.</b> Box Plots of average ideal offset using GA and ACO for Model I .....	26
<b>Figure 10.</b> Box plots of objective function values found using GA and ACO for Model II .....	27
<b>Figure 11.</b> Box plots of number of cars processed (first term of Eq. 1) using GA and ACO for Model II .....	28
<b>Figure 12.</b> Box plots of ideal offset (Eq. 2) using GA and ACO for Model II.....	28
<b>Figure 13.</b> Box plots of average ideal offset using GA and ACO for Model II .....	28
<b>Figure 14.</b> Box plots of de facto red using GA and ACO for Model II.....	29
<b>Figure 15.</b> Box plots of storage capacity using GA and ACO for Model II.....	29

## LIST OF TABLES

<b>Table 1.</b> Algorithmic settings for four cases tested for the ACO and GA methods .....	23
<b>Table 2.</b> Comparison of ACO and GA for Model I .....	32
<b>Table 3.</b> Comparison of ACO and GA for Model II.....	32

## **DISCLAIMER**

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the Department of Transportation, University Transportation Centers Program, in the interest of information exchange. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

## **ACKNOWLEDGMENTS**

The authors recognize that support for this research was provided by a grant from the U.S. Department of Transportation, University Transportation Centers Program to the Southwest Region University Transportation Center which is funded, in part, with general revenue funds from the State of Texas.

## EXECUTIVE SUMMARY

### CHAPTER 1. PROBLEM INTRODUCTION

Traffic congestion is a daily and growing problem of the modern era in mostly all major cities in the world. The increasing traffic demand puts a lot of strain on the existing transportation system. One of the major problems is the oversaturated network conditions, especially at the peak hours. Oversaturation occurs when the queues of vehicles on a receiving street interfere with the performance of the respective adjacent upstream streets. Strategies for enhancing the mobility and efficiency of the existing traffic system can significantly reduce costs related to infrastructure renewal and expansion. With the traffic volumes rapidly increasing with time and more and more road networks being forced to function at high density, the need to solve the problem of signal coordination for oversaturated traffic networks has never been more significant than it is today. Due to a limited budget, it is very important to sort out ways to enhance mobility and improve the efficiency of the existing infrastructure.

To improve the efficiency of the system, intelligent traffic signal networks can be designed to maximize the number of vehicles processed by the network and minimize the travel time of the vehicles in the system. While significant research efforts have studied signal coordination for uncongested systems and designed traditional control policies to minimize costs, in the case of over-congested systems, a similar handling of the problem yielded undesirable outputs and the dissipation of queues and removal of blockages were prioritized over the minimization of cost. The objective of the problem is therefore to generate a set of green times for the network such that the number of vehicles that are released at every signal of the network during congestion phase has to be maximal.

Existing algorithms for signal coordination in oversaturated networks are too slow or require very high computational power to be effective in a reasonably short time. The Ant Colony Optimization (ACO) algorithm is a fairly novel technique for solving computational problems mimicking the natural behavior of ants trying to identify the shortest path from their colony to a food source. ACO produced good results in various other fields of research and has been applied successfully for transportation planning problems. This paper attempts to implement ACO to solve the oversaturation traffic network problem and compare its

performance against solutions obtained with the Genetic Algorithm, the optimization technique used to solve these problems in the existing literature.

## **CHAPTER 2. BACKGROUND**

A set of studies developed strategies for signal timing for oversaturated conditions (Gazis 1964; Gazis and Potts 1965; Burhardt 1971; Kaltenbach and Koivo 1974; Michalopoulos and Stephanopolos 1977, 1978; Chang and Lin 2000). A few studies proposed adaptive signal control for oversaturated conditions, but this approach is better suited for highly varying traffic conditions, rather than predictably high traffic volumes.

Development of optimal traffic signal timing strategies for oversaturated conditions is a difficult task, as system-level congestion should be controlled by adjusting parameters at individual intersections, such as cycle length. Signal timing optimization for oversaturated networks requires emphasis on traffic throughput and queue management rather than on delay and progression, which are typically optimized for control in unsaturated conditions. Heuristic optimization methods have been applied successfully for unsaturated traffic signal timing problems; and for oversaturated conditions, Park et al. (1999) employed a genetic algorithm (GA)-based approach to reduce queue time, and Lieberman et al. (2000) used a mixed integer linear programming-based approach to control queue growth for oversaturated arterials. To better facilitate use of the GA to solve oversaturated conditions, a study was conducted to reformulate the optimization model (Abu-Lebdeh and Benekohal 1997, 2000). This work formulated an objective function that takes into account the total amount of vehicles that are processed by the network during the oversaturated period. New models for estimating the capacities of oversaturated arterials were developed based on the capacities of individual intersections, vehicle queue lengths and offsets. The GA was applied to coordinate signals to maximize throughput, and results demonstrated a control strategy that avoided queue spillback and de facto red. This strategy was extended to coordinate oversaturated signals along an arterial that crosses multiple, parallel coordinated arterials (Girianna and Benekohal 2002, 2004).

## **CHAPTER 3. MODEL FORMULATION**

This research utilizes the optimization model developed by Girianna and Benekohal (2002, 2004) and compares the use of an ACO algorithm to the performance of a GA-based approach. The traffic signal design problem for oversaturated conditions has been effectively

formulated as a constrained dynamic optimization problem with the objective to maximize the number of vehicles released by a signal network, summarized as follows:

$$\text{Max } Z = \sum_t^T \sum_{(i,j) \in L} \sum_h^H \frac{d_{i,j}}{d_{\max}} D_{i,j}^h(t) - \sum_k^K \sum_{(i,j) \in L_p} \delta_{i,j}(k) \max \left( \begin{matrix} 0 \\ q_{i,j}^{h*}(k) - q_{i,j}^{\max} \end{matrix} \right) \quad (1)$$

$$\phi_{i,j}^{h*}(k) = \frac{d_{i,j}}{v_{i,j}} - \left( \frac{(v_{i,j} + \lambda) l_{veh}}{v_{i,j} \lambda} \right) q_{i,j}^{h*}(k) \quad \forall (i,j) \in L_p, \quad k = 1, \dots, K \quad (2)$$

$$g_i^{h*}(k) \leq g_j^{h*}(k) + \phi_{i,j}^{h*}(k) + \beta_{i,j}(k) \quad \forall (i,j) \in L_p, \quad k = 1, \dots, K \quad (3)$$

$$\sum_{(i,j) \in F(r)} \phi_{i,j}(k) - \sum_{(i,j) \in R(r)} \phi_{i,j}(k) + \sum_{j \in N(r)} (g_{i,r}(k) + \Delta) = \sum_{m=k, j \in N(r)}^{k+n_k} C_j(m) \quad \forall r \in R, \quad k = 1, \dots, K \quad (4)$$

$$q_{i,j}^h(t) \leq \frac{d_{i,j}}{l_{veh}} \quad \forall h \in H, \quad \forall (i,j) \in L_s, \quad t = 1, \dots, T \quad (5)$$

$$g_{\min} \leq g_j^h(k) \leq g_{\max} \quad \forall j \in N, \quad \forall h \in H, \quad k = 1, \dots, K \quad (6)$$

$$I_{i,j}^h(t) = \sum_{p \in U_{h,b} \in B_i} \theta_{b,i} D_{b,i}^p(t) \quad \forall h \in H, \quad t = 1, \dots, T \quad (7)$$

$$A_{i,j}^h(t) = F \theta_{i,j} I_{i,j}^h(t - \tau_{i,j}) + (1 - F) A_{i,j}^h(t - 1) \quad \forall h \in H, \quad \forall (i,j) \in L, \quad t = 1, \dots, T \quad (8)$$

$$D_{i,j}^h(t) = \min \left( \begin{matrix} c_j^h(t) \Delta T \\ A_{i,j}^h(t) \Delta T + q_{i,j}^h(t - 1) \end{matrix} \right) \quad \forall h \in H, \quad \forall (i,j) \in L, \quad t = 1, \dots, T \quad (9)$$

Where:

$G = (N, L, P)$  denotes a traffic signal network

$N$  = set of signals

$L$  = set of directional streets

$P$  = set of coordinated paths  $p_{ij}$  starting form signal  $i$  to signal  $j$

$N_p$  = Set of signals on the coordinated paths

$L_p$  = set of streets along coordinated paths

$K$  = period of oversaturation in a cycle number

$T$  = period of oversaturation in a sample time

$t = 1, 2, \dots, T$  is a discrete time index

$\Delta T$  = sample time interval (say 2, 3, 4 or 5 or more secs)

$H$  = total phase number

$d_{i,j}$  = distance from signal  $i$  to  $j$

$d_{\max}$  = maximum length of streets in the network

$D_{i,j}^h(t)$  = departure flows of phase  $h \in H$  at signal  $j$  serving flows from signal  $i$  over a period of  $[t\Delta T, (t+1)\Delta T]$

$q_{i,j}^{h^*}(k)$  = number of vehicles in queue approaching signal  $j$  coming from signal  $i$  at the beginning of the downstream coordinated green phase  $h^*$  in cycle  $k$ . The star (\*) indicates a coordinated phase.

$\delta_{i,j}(k)$  = non-negative disutility factor

$\phi_{i,j}^{h^*}(k)$  = offset between signal  $i$  and  $j$

$v_{i,j}$  = speed of a released platoon

$l_{veh}$  = average length of vehicles

$\lambda$  = starting shock wave speed

$g_i^{h^*}(k)$  = effective green time at signal  $i$

$\beta_{i,j}(k)$  = time it takes for a stopping shock wave to propagate upstream

$C_j(m)$  = length of  $m^{\text{th}}$  cycle

$N(r)$  = set of nodes on loop  $r \in R$  (number of loop in the network)

$F(r)$  = set of nodes where traffic moves in the same direction as the loop

$R(r)$  = set of nodes where traffic moves in a different direction to that of the loop

$\Delta$  = lost green time

$I_{i,j}^h(t), A_{i,j}^h(t), D_{i,j}^h(t)$  = inflow, arrival and departure flows of phase  $h$  over  $[t\Delta T, (t+1)\Delta T]$

$U_h$  = set of phases at the upstream signal that feeds traffic for phase  $h$  of the downstream signal

$b \in B_i$  = set of upstream intersections connected to intersection  $i$

$\theta_{b,i}$  = % of the departed traffic volume of upstream streets ( $b, i$ ) that enters road section ( $i, j$ )

$\gamma$  = unitless platoon dispersion factor empirically derived (0.5)

$\tau_{i,j}$  = unitless cruise travel time of a released platoon factored by 0.8

$F$  = smoothing factor =  $\frac{1}{(1+(\gamma)(\tau_{i,j}))}$

$c_j^h(t)$  = capacity during effective green interval

Equation 1 is composed of two terms. The first term is the total number of vehicles processed by the network throughout the oversaturation period. Each signal's release is weighted by the ratio of distance travelled to the length of the longest street in the network. The second term is the disutility function penalizing the occurrence of queues at the end of green time along coordinated arterials. Equation 2 is needed to guarantee a coordinated offset between signal  $i$  and  $j$ . Equation 3 is introduced to avoid de-facto Red. The sum of offsets and green times around any loop of the network is equal to an integer multiple of the cycle time. Equation 5 enforces that queues in a non coordinated arterial must not block the traffic movements of the upstream intersections. Equation 6 fixes the range of the control variables  $g_j^h(k)$ . Equations 7, 8, and 9 are the necessary network flow constraints.

The signal coordination model finds optimal signal timing for the entire period of oversaturation. The problem becomes a large combinatorial optimization problem and cannot be efficiently solved using traditional calculus-based optimization techniques. Ant Colony Optimization and Genetic Algorithms are used to solve the problem. Both methods are heuristic optimization methodologies that mimic the mechanisms of nature.

For these optimization techniques, the variables are divided into many intervals based on the amount of accuracy required. Equation 8 simulates the traffic flows and queues for a given set of green times. The range of green times is enforced automatically by the algorithm and does not need to be checked. The minimum and maximum values of the variables are to be inputted and the algorithm divides the range into a finite number of intervals. In this article, an interval size of 5 sec is used. The ants then choose a node number. The value of the variable is then determined as:

$$g = (g_{\min} - \text{interval size}) + (\text{node number} * \text{interval size}). \quad (10)$$

Equations 2–6 need to be checked for violation and a penalty is to be added to the objective function. The constrained signal coordination problem is transformed into an unconstrained problem by associating a penalty to the objective function every time the constraint is violated. The fitness value of an individual  $i$  is defined by extending the domain of the objective function  $Z_i$  using Equation 11, where  $\mu_j$  is a penalty coefficient for constraint  $j$ ,  $m$  is the number of implicit constraints, and  $H_j$  denotes  $j$ 's constraint function (inequality and equality). The fitness that is analogous to the length of the path to be travelled by the ant is then

minimized.  $C_{\min}$  is an input coefficient introduced to overcome the negative value of the augmented objective function.

$$fitness_i = C_{\min} - \left( Z_i - \sum_{j=1}^m \mu_j H_j \right) \quad (11)$$

The constraint in Equation 5 is not active if a coordinated signal is an open-loop system; that is, when multiple coordinated arterials cross a single coordinated arterial. Hence, the augmented objective function becomes as formulated below.

$$\begin{aligned} Max \quad & C_{\min} - \left[ \sum_k^K \sum_{(i,j) \in L} \sum_h^H \frac{d_{i,j}}{d_{\max}} D_{i,j}^h(k) - \sum_k^K \sum_{(i,j) \in L_p} \delta_{i,j}(k) \max \left( \begin{array}{c} 0 \\ q_{i,j}^{h*}(k) - q_{\max} \end{array} \right) \right. \\ & - \mu_1 \sum_{k,(1,j) \in L_p}^K \left( \phi_{i,j}^h(k) - \left( \frac{d_{i,j}}{v_{i,j}} - \frac{(v_{i,j} + \lambda) l_{veh}}{v_{i,j} \lambda} q_{i,j}^{h*}(k) \right) \right)^2 \\ & \left. - \mu_2 \sum_{k,(i,j) \in L_p}^K \max \left( \begin{array}{c} 0 \\ g_i(k) - (g_j(k) + \phi_{i,j}(k) + \beta_{i,j}(k)) \end{array} \right) - \mu_3 \sum_{k,(1,j) \in L}^K \max \left( \begin{array}{c} 0 \\ q_{i,j}^h(k) - \frac{d_{i,j}}{l_{veh}} \end{array} \right) \right] \end{aligned} \quad (12)$$

The variable that is green time is represented for the GA optimization using a four-bit binary string that is decoded using Equation 13.

$$g = g_{\min} + \left( \frac{g_{\max} - g_{\min}}{2^{d-1}} \right) DV \quad (13)$$

Where:

$DV$  = decoded value of a string.

#### CHAPTER 4. SOLUTION METHODOLOGY

Two approaches are used and compared in this study to identify an optimization methodology that could be utilized reliably to solve the traffic signal timing problem for oversaturated conditions. Both methods are heuristic optimization methods based on mechanisms observed in nature. Ant Colony Optimization is a novel technique for solving computational optimization problems. ACO is a heuristic method for identifying a near-optimal solution and mimics the natural behavior of ants as they collectively identify the shortest path to a food source from their colony. Genetic Algorithms are a population-based search methodology. This method uses a large number of individuals and adjusts the vector of decisions values through operators that combine characteristics of very fit solutions and randomly changing a few numbers in the population of individuals. Fit solutions are selected to

survive to the next iteration, or generation, and poorly fit solutions are discarded from the population. Over many iterations, the population converges to a near optimal solution.

## **CHAPTER 5. APPLICATION TO ILLUSTRATIVE CASE STUDIES**

Two test cases are considered: a model network of 20 intersections with simulated traffic data and a realistic application; namely a section of the downtown traffic network of Fort Worth, Texas, for which this study collected and used actual traffic data. Both networks have a total duration of 15 minutes for oversaturated conditions.

## **CHAPTER 6. RESULTS**

Both the GA and the ACO approaches are applied using the above formulation to solve two different case studies. It is observed for both models that the fitness function was improved when the number of executions is increased. For the first network, the ACO performed better than the GA and was able, for the best case, to find a solution that would process a total number of 3000 vehicles through the network, and minimize the ideal offset to an average of approximately 2 seconds for each cycle and at each intersection in the network. For the second network, the ACO performed more reliably, with less variability in a set of results. ACO was able to identify signal timings that processed 4000 vehicles through the network and minimized the ideal offset to approximately one second for each cycle at each intersection.

## **CHAPTER 7. HYPOTHESIS TESTS**

Statistical hypothesis tests were conducted to determine the algorithm with better performance. It was observed for both the models that the fitness function continues to decrease with an increasing number of executions. ACO performs better for the first network, and particularly better than the GA for the largest number of executions. For the second model, GA outperforms ACO in all cases except the one, which is the case with the highest number simulation executions. ACO was found to more consistently improve its performance with an increasing number of executions, compared to the GA.

## **CHAPTER 8. CONCLUSIONS AND RECOMMENDATIONS**

It has been found that for the 30 trials, the results identified by ACO returned reliably low values of the ideal offset, which contributes most significantly to keeping a steady flow of cars through the network and avoiding queues in the streets that are upstream of intersections. For the first network, the ACO performed better than the GA, returning reliably good solutions. For the second network, the GA identifies more fit solutions for lower number of solution

evaluations (e.g., computational power), but for the highest setting, ACO is able to identify more fit solutions. By utilizing increased amounts of computational power, both algorithms were able to improve fitness values of the solutions identified. With improved fitness values, the number of cars passing through the network did not improve, but the ideal offset improved dramatically for the increasing computational power, especially for solutions identified by the ACO algorithm. For the second network, results demonstrated that the two different algorithms identified different types of solutions: GA identified solutions that show considerable variability in the ideal offset and some variability in the number of vehicles processed, but routinely identifies solutions with minimal de facto red time and queue storage time. ACO, on the other hand, identified solutions with variability in the queues storage time and de facto red time, but minimal values for the ideal offset and similar values for the number of vehicles processed through the network. Therefore, ACO identifies solutions that minimize the more significant objective of planning traffic light timing for oversaturated conditions.

Statistical analysis showed that ACO yielded better results when compared to GA for cases having a higher number of executions. As a result, better performing solutions can be identified for the same computational power. This is especially important when the best possible solution should be obtained and the user can afford the computational time to do so. These results indicate that ACO may prove to be a good alternative when trying to solve very complicated networks.

## CHAPTER 1. PROBLEM INTRODUCTION

Increasing areas of urbanization result in increased demands on the transportation system. In rapidly developing areas, volumes of traffic may increase well beyond the original load used to design the system. Increased strain on transportation systems results in traffic networks that are oversaturated, especially during peak flows. Oversaturation of a traffic network is a phenomenon observed when the queues of vehicles on a receiving street of the network are long enough to interfere with the performance of the respective adjacent upstream streets. Due to the limited budgets of municipalities, cost-effective strategies should be designed to enhance the mobility and improve the efficiency of the existing infrastructure. While investments in additional infrastructure may improve conditions, a more cost-effective strategy would improve the efficiency of the system through intelligent traffic signal networks. Intelligent traffic signal networks can be designed to maximize the number of vehicles processed by the network and minimize the travel time of the vehicles in the system.

Signal coordination for uncongested systems can be posed as an optimization model, where the timings for traffic lights are identified that will minimize the time for vehicles to pass through the network. Extending the existing approaches for over-congested conditions, however, does not effectively manage the large numbers of vehicles. Traditional control policies, such as minimization of time, do not yield satisfactory results in oversaturated conditions. For over-congested networks, the dissipation of queues and removal of blockages should be prioritized before the minimization of costs (Roess et al. 1998). The objective of managing congested conditions is to generate a set of green times for the network that would maximize the number of vehicles that are released at every signal of the network (Abu-Lebdeh and Benekohal 1997; Girianna and Benekohal 2002).

At high traffic volumes, the majority of the roads within a network are forced to function at high density, and identification of signal timings that will move vehicles through the network most effectively is a difficult problem to solve. This research develops and applies an ant colony optimization (ACO)-based algorithm to solve the problem of signal coordination for transportation systems in oversaturated conditions. ACO is a heuristic optimization method that probabilistically identifies a near-global optima for solution of non-linear, discrete, and complex problems. This research demonstrates the application of the ACO approach to identify traffic

signal timings for an illustrative case study and a realistic network. Results are compared to another heuristic method, the genetic algorithm.

## CHAPTER 2. BACKGROUND

Considerable research has been done to develop algorithms for controlling traffic signal networks; however, oversaturation conditions remain a challenging area. For example, SCOOT (Split Cycle Offset Optimization Technique) is a tool for managing and controlling traffic signals in urban areas. It is an adaptive system that responds automatically to fluctuations in traffic flow through the use of on-street detectors embedded in the road. SCOOT, however, showed major deficiencies in oversaturated and highly fluctuating conditions (Yagar and Dion 1996). While the adaptive control model that incorporates probabilistic forecasts of individual vehicle arrivals derived works well for high volumes, this framework could not be successfully extended to oversaturated traffic demands (Yu and Recker 2006).

Other algorithms have been developed specifically to address oversaturated conditions. Other algorithms that use a set of heuristic rules, or “rules of thumb,” include

- a hierarchy control based algorithm (Chen et al. 2007);
- a self-tuning system for controlling signalized intersections in undersaturated and oversaturated conditions (Ceder and Reshetnik 2001); and
- a dynamic traffic-control formulation for modeling oversaturated traffic after the cell-transmission model (Lo 2001).

Chang and Sun (2004) proposed a dynamic method to control an oversaturated traffic signal network by coupling a traditional model for under saturated intersections and a “bang-bang” model for oversaturated conditions. The bang-bang method is used to find an optimal switchover point during the oversaturated period to interchange the timing of the approaches, which are based on a set of heuristic rules. Other studies utilize optimization searches to find the best strategies for light timings. Lieberman and Chang (2005) decomposed a grid network into its constituent arterial subsystems that are responsive to user-specified priorities and used a signal control system named Real-Time/Internal Metering Policy to Optimize Signal Timing (RT/IMPOST). This technique is designed to compute signal timing plans for the entire range of operating conditions from under saturation to oversaturation. It was observed that the IMPOST timing plans were better than the fine tuned existing control along a New York State arterial system. Signal timing optimization for oversaturated networks requires emphasis on traffic throughput and queue management rather than on traditional delay and progression. Four new

optimization objective functions were introduced to the popular TRANSYT-7F model version 8. These objective functions were analyzed and compared with the traditional objective functions and were found to be generate superior timing plans (Li and Gan 1999).

Another approach to generate signal timing plans is to apply an optimization methodology to maximize “bandwidth.” Maximizing bandwidth and progression is the oldest type of arterial signal optimization. Results can be viewed graphically very easily, and experienced engineers can identify sufficiently efficient plans through trial-and-error. It is, however, a deterministic model and cannot handle bandwidth under variable traffic conditions and important characteristics in oversaturated conditions, such as the queue information, cannot be utilized in this method. Little et al. (1981) developed a portable FORTRAN IV computer program called MAXBAND that maximizes the bandwidth for arterial signals. This program can

- choose an optimum cycle length,
- select an optimum design speed,
- select the best lead or lag pattern for left turn phases,
- allow queue clearance time for secondary flow accumulated during red and
- accept user-specified weights for green bands in each direction to maximize the weighted combination of bandwidths.

The optimization uses Land and Powell’s MPCODE branch and bound algorithm. As many as 12 signals can be handled by this program efficiently. Chang et al. (1988) extended the MAXBAND program to include the left-turn phase sequence as a variable to optimize the flow of traffic in multi-arterial closed networks apart from the well-known green phase time, offset and cycle length. It was found that optimization of phase sequence can provide a substantial benefit in terms of delay and stops.

Gartner et al. (1990) presented a new optimization approach for arterial progression that incorporates a systematic traffic dependent criterion. The method calculates a bandwidth value for each directional road section and obtains the sum of the weighted bandwidths (hence, the term *multi-band*). Mixed integer linear programming is to optimize the system-level bandwidth. Gartner et al. (1991) developed a computer program named Multiband to implement this new approach to arterial progression optimization and incorporate a systematic traffic dependent criterion. The method generates variable bandwidth progression schemes in which each directional road section is assigned individually weighted band, and it can program the traffic

progression scheme to specific traffic flow patterns on each link of the arterial. Simulation results indicate that this method can produce considerable gains in performance when compared with traditional progression methods. It also lends itself to a natural extension for the optimization of grid networks.

Little et al. (1966) developed a Program called TSS3 written for 20K IBM 1620 to synchronize the signals to produce bandwidths that are equal in each direction and as large as possible. They also tried to adjust the synchronization to increase one bandwidth to some specified, feasible value and maintain the other as large as is then possible. Little (1966) formulated a mixed-integer linear program for the following arterial problem: Given an arbitrary number of signals, the red-green split at each signal, upper and lower limits on signal period, upper and lower limits on speed between adjacent signals, and limits on change in speed, find common signal period, speeds between signals, and the relative phasing of the signals, in order to maximize the sum of the bandwidths for the two directions. Several variants of the problem are formulated, including the synchronization of a network of signals. Branch-and-bound algorithms were developed for solving the mixed-integer linear programs by solving sequences of ordinary linear programs. A 10-signal arterial example and a 7-signal network example were successfully solved.

Stamatiadis and Gartner (1996) developed the MULTIBAND-96 model that optimizes all the signal control variables, including phase lengths, offsets, cycle time, and phase sequences, and generates variable bandwidth progressions on each arterial in the network. It has the ability to adapt to each link in the network and offers considerable advantages compared with existing models. Branch-and-bound search techniques were used to solve the mixed-integer linear program to compute signal timing plans. However, the method is numerically unstable and computationally inefficient which results in suboptimal or no solutions for network problems with a range of variable cycle times. Pillai et al. (1998) developed a fast and numerically stable heuristic approach based on restricted search of the integer variables in the solution space for the maximum bandwidth signal setting problem.

More efficient strategies for signal timing may be identified by reformulating the optimization problem from a throughput maximization problem or a bandwidth maximization problem to a model that incorporates the more significant characteristics of oversaturated conditions. These models can then be solved using heuristic optimization methodologies. Park

et al. (1999) designed an algorithm that coupled a Genetic Algorithm-based (Goldberg 1989) optimizer and a mesoscopic traffic simulator, which could handle oversaturated conditions. The methodology was evaluated based on queue time.

Abu-Lebdeh and Benekohal (1997, 2000) developed optimization models for estimating the capacities of oversaturated arterials based on the capacities of individual intersections, vehicle queue lengths and offsets as the input variables. A GA-based approach was used for coordination of signals along oversaturated arterials, by formulating the problem as a throughput maximization problem. Constraints were enforced to discourage formation of queues, and dissipation was encouraged to prevent the formation of de facto red and to provide a time dependant control measure. The results that were obtained in this study showed a dynamic and responsive traffic progression control that avoided undesirable conditions such as queue spillback and de facto red, and performed better than typical heuristic rules for low and high demand volumes.

The objective function was developed to take into account the total amount of vehicles that are processed by the network during the oversaturated period. A disutility function that takes care of penalizing the queues at the end of each cycle in the network is also added to the objective function. Five different constraints have been suggested:

- Ideal Offsets,
- De Facto Red,
- Coordinated Loops,
- Queue Storage capacity and
- Control Variables.

The constraint that enforces ideal offsets enables the coordination of the green times of two adjacent signals on a coordinated arterial of the signal network and optimizes the offset in such a way that the maximum traffic is processed by the two signals. De facto red exists when the traffic in the upstream cannot advance into the receiving downstream street because of the queue in the receiving street. The corresponding constraint avoids this situation. Coordinated loops constraint enables the condition that the sums of offsets and green times around any loop of the network are equal to an integer multiple of the cycle time (Gartner 1972). Queue storage capacity constraint keeps the queues in the non coordinated arterials in check. This constraint makes sure that in order to keep the coordinated arterials free, the non coordinated arterials are

not congested. It also enforces the queue length of the non coordinated arterial and prevents long queues from forming. The control variables constraint makes sure that all the control variables are within reasonable operating ranges. For example, the possible values of the green times of the different signals of the traffic signal network are bounded by a maximum possible value and a minimum possible value.

Further research has been done to reduce the computation time of the GA-based approach to facilitate more efficient implementation on real-time signal control systems. An extensive study was conducted and a solution for the problem of dynamic signal coordination in over-congested networks was suggested using micro Genetic algorithms (Girianna and Benekohal 2002). The algorithm extended the basic concept of signal coordination applied to oversaturated single arterials for a grid network of arterials. The algorithm was found to manage local queues by spatially distributing them over signalized intersections and temporally spreading them over signal cycles. The algorithm also took into account the traffic demand variation and position of the signals to intelligently generate signal timing for the whole network. In addition, the methodology identified a set of common cycles propagated from upstream signals, thus promoting traffic progression. In order for the Simple Genetic Algorithm (SGA) signal coordination to be applicable to real-time systems, the computational time of the SGA was reduced (Girianna and Benekohal 2004). In order to speed up the computation time, a master slave mechanism was used to distribute the computational work among parallel computing nodes. By better utilizing the available computational power, the solution was obtained more quickly.



## CHAPTER 3. MODEL FORMULATION

### 3.1 SIGNAL COORDINATION

In oversaturated networks, traditional metrics such as cost minimization are considered secondary. Instead, the first priority is given to the removal of queues and blockages. The number of vehicles processed by the network is also considered important. For conditions where queues are cleared, the network operates as an undersaturated network, and signal timings should provide green time offsets to maximize the progression of traffic. The objective function used by Girianna and Benekohal (2002, 2004) has been adopted for use in this study and is described below.

### 3.2 OBJECTIVE FUNCTION

Let  $G = (N,L,P)$  denote a traffic signal network

Where,

$N$  = set of signals

$L$  = set of directional streets

$P$  = set of coordinated paths

$L_p$  = set of streets along coordinated paths

$N_p$  = Set of signals on the coordinated paths

$p_{ij} = \{s_i^h, \dots, s_j^h\}$  = path for signal coordination starting form signal I to signal j

$s_i^h$  = signal I with coordinated phase h

$P = \{ p_{ij} / s_i \in S_o \text{ and } s_j \in S_d \}$

$S_o$  = Set of signals at which traffic starts

$S_d$  = Set of signals at which traffic terminates

For the dissipation of queues along coordinated arterials, the offsets between adjacent signals should allow the queue in the downstream to begin dissipating before the upstream platoon arrives at the intersection. In addition, green time must be used effectively; that is, green time not used to process vehicles through the intersection should be avoided. Equation 1 shows the objective function for signal coordination.

$$Max \quad Y = \sum_t^T \sum_{(i,j) \in L} \sum_h^H \frac{d_{i,j}}{d_{max}} D_{i,j}^h(t) - \sum_k^K \sum_{(i,j) \in L_p} \delta_{i,j}(k) \max \left( \begin{matrix} 0 \\ q_{i,j}^{h*}(k) - q_{i,j}^{max} \end{matrix} \right) \times \delta_{i,j}(k), d_{i,j}, d_{max} > 0 \quad (1)$$

Where:

$D_{i,j}^h(t)$  = departure flows of phase  $h \in H$  at signal  $j$  serving flows from signal  $i$  over a period of  $[t\Delta T, (t + 1)\Delta T]$ .

$H$  = total phase number,

$\Delta T$  = sample time interval (say 2, 3, 4 or 5 or more sec), and  $t = 1, 2, \dots, T$  is a discrete time index.

$d_{i,j}$  = distance from signal  $i$  to  $j$ , and

$d_{\max}$  = maximum length of streets in the network.

$q_{i,j}^{h*}(k)$  = number of vehicles in queue approaching signal  $j$  coming from signal  $i$  at the beginning of the downstream queue.

$\delta_{i,j}(k)$  = non-negative disutility factor whose values are determined based on a queue management strategy.

In Equation 1, the first term is the total number of vehicles processed by the network throughout the oversaturation period. Each signal's release is weighted by the ratio of distance travelled to the length of the longest street in the network. The second term is the disutility function that penalizes the occurrence of queues at the end of green time along coordinated arterials.

### 3.3 CONSTRAINT: IDEAL OFFSETS

Offset is defined as the time difference between the green time initiations of two adjacent signals. The offsets need to be coordinated taking into account the

- distance between the signals,
- released platoon speed,
- platoon dispersion and
- time required for the queue to dissipate.

As it is not possible to coordinate all the phases, only the phase that controls the coordinated movement in both the signals is coordinated.

$$\begin{aligned}\phi_{i,j}^{h^*}(k) &= \tau_{i,j}(k) - \alpha_{i,j}(k) = \frac{l_{i,j}(k)}{v_{i,j}} - \frac{q_{i,j}^{h^*}(k)l_{veh}}{\lambda} = \frac{d_{i,j} - q_{i,j}^{h^*}(k)l_{veh}}{v_{i,j}} - \frac{q_{i,j}^{h^*}(k)l_{veh}}{\lambda} \\ &= \frac{d_{i,j}}{v_{i,j}} - \left( \frac{(v_{i,j} + \lambda)l_{veh}}{v_{i,j}\lambda} \right) q_{i,j}^{h^*}(k) \quad \forall (i,j) \in L_p, \quad k = 1, \dots, K\end{aligned}\quad (2)$$

Where:

$\tau_{i,j}(k)$  = time taken by the first vehicle in the released platoon from signal  $i$  to join the moving tail of the downstream platoon

$v_{i,j}$  = speed of a released platoon, and

$l_{i,j}(k)$  = unoccupied space along street  $(i, j)$  at the beginning of cycle  $k$ .

$l_{veh}$  = average length of vehicles.

$q_{i,j}^{h^*}(k)$  = queue for coordinated movement at the beginning of phase  $h^*$  of cycle  $k$ .

$\alpha_{i,j}(k)$  = time required for the tail to start moving

$\lambda$  = starting shock wave speed

$$\alpha_{i,j}(k) = \frac{q_{i,j}^{h^*}(k) (l_{veh})}{\lambda} \quad (3)$$

### 3.4 CONSTRAINT: DE FACTO RED

De facto red exists when the queue in the downstream is long enough to stop traffic from the upstream from entering it even though the signal is green. To avoid this, the effective green time of the upstream signal should be less than the sum of

- the effective green time for the coordinated downstream signal,
- the offset between two signals,
- and the time it takes for a stopping shock wave to move upstream.

$$g_i^{h^*}(k) \leq q_j^{h^*}(k) + \phi_{i,j}^{h^*}(k) + \beta_{i,j}(k) \quad \forall (i,j) \in L_p \quad (4)$$

Where:

$g_i^{h^*}(k)$  = effective green time for the upstream signal

$g_j^{h^*}(k)$  = effective green time for the coordinated downstream signal

$\phi_{i,j}^{h^*}(k)$  = offset between the two coordinated signals

$\beta_{i,j}(k)$  = time it takes for a stopping shock wave to propagate upstream

### 3.5 CONSTRAINT: COORDINATED LOOPS

Sum of offsets and green times around any loop of the network is equal to an integer multiple of the cycle time.

$$\sum_{(i,j) \in F(r)} \phi_{i,j}(k) - \sum_{(i,j) \in R(r)} \phi_{i,j}(k) + \sum_{j \in N(r)} (g_{j,r}(k) + \Delta) = \sum_{m=k, j \in N(r)}^{k+n_k} C_j(m) \quad \forall r \in R \quad (5)$$

Where:

$C_j(m)$  = cycle length of  $m^{\text{th}}$  cycle

$N(r)$  = set of nodes on the loop  $r$

$F(r)$  = set of nodes where traffic moves in the same direction as the loop

$R(r)$  = set of nodes where traffic moves in a different direction to that of the loop

$\Delta$  = lost green time

### 3.6 CONSTRAINT: QUEUE STORAGE CAPACITY

Queues in the non coordinated arterials can get very long if there is no constraint to guard them. As per this constraint, the length of the queue in a non coordinated arterial must not exceed

$q_{i,j}^{\max}$

$$q_{i,j}^h(k) \leq q_{i,j}^{\max} \quad q_{i,j}^h(k) \leq \frac{d_{i,j}}{l_{veh}} \quad (i,j) \in L_s \quad (6)$$

Where:

$$q_{i,j}^{\max} = d_{i,j} / l_{veh}$$

$d_{i,j}$  = length of the street

$l_{veh}$  = average length of vehicles

### 3.7 CONSTRAINT: CONTROL VARIABLES

All the variables in use must be well within their ranges. In our case, if  $g_{\max}$  and  $g_{\min}$  are the maximum and minimum bounds of the green time,

$$g_{\min} \leq g_j^h(k) \leq g_{\max} \quad \forall j \in N, \quad \forall h \in H, \quad k = 1, \dots, K \quad (7)$$

Where:

$g_j^h(k)$  = possible values of the green time at intersection  $j$  during phase  $h$  and cycle  $k$

### 3.8 NETWORK FLOWS

For a two signal model connecting two intersections  $i$  and  $j$ , we define  $I_{i,j}^h(t)$ ,  $A_{i,j}^h(t)$  and  $D_{i,j}^h(t)$  to be the inflow, arrival and departure flows of phase  $h \in H$  over a period  $[t\Delta T, (t+1)\Delta T]$ . Assume the traffic to be entering from signal  $i$  (upstream) and departing from signal  $j$  (downstream) and that the two signals work on a four phase plan,  $H = 4$ . Then the inflow for phase  $h$  of signal  $j$  moving from phase  $p$  of signal  $i$  during interval  $t$  is given by Equation 8.

$$I_{i,j}^h(t) = \sum_{p \in U_h, b \in B_i} \theta_{b,i} D_{b,i}^p(t) \quad (8)$$

Where:

$U_h$  = set of signal phases at the upstream signal that feeds traffic for phase  $h$  of the downstream signal

$\theta_{b,i}$  = percentage of the departed traffic volume of upstream streets ( $b, i$ ) that enters road section ( $i, j$ )

$b \in B_i$

$B_i$  = set of upstream intersections connected to intersection  $i$ .

The arrival flows are formulated using Robertson's platoon dispersion model shown in Equation 9.

$$A_{i,j}^h(t) = F\theta_{i,j}I_{i,j}^h(t - \tau_{i,j}) + (1 - F)A_{i,j}^h(t - 1) \quad (9)$$

Where:

$\gamma$  = unitless platoon dispersion factor empirically derived (0.5)

$\tau_{i,j}$  = unitless cruise travel time of a released platoon factor (0.8)

$F$  = smoothing factor

$$F = \frac{1}{(1 + (\gamma)(\tau_{i,j}))} \quad (10)$$

The departure flows of phase  $h \in H$  at signal  $j$  over a period  $[t\Delta T, (t + 1)\Delta T]$  are shown in Equation 11.

$$D_{i,j}^h(t) = \min \left( \begin{array}{c} c_j^h(t)\Delta T \\ A_{i,j}^h(t)\Delta T + q_{i,j}^h(t-1) \end{array} \right) \quad \forall i, j \in L \quad (11)$$

Where:

$c_j^h(t)$  = capacity during effective green interval

The queue length during phase  $h$  at the end of the time interval  $t$  is shown in Equation 12.

$$q_{i,j}^h(t) = \max \left( \begin{array}{c} 0 \\ q_{i,j}^h(t-1) + (A_{i,j}^h(t) - D_{i,j}^h(t)) \end{array} \right) \quad (12)$$

## CHAPTER 4. SOLUTION METHODOLOGY

### 4.1 ANT COLONY OPTIMIZATION

The signal coordination model should be solved to identify optimal signal timing for the entire period of oversaturation. The problem is a large combinatorial optimization and cannot be efficiently solved using traditional calculus-based optimization techniques. Ant Colony Optimization can be applied to solve the problem. ACO is an optimization technique inspired from the natural behavior of ants (Dorigo and Thomas 2004). Ants can identify the shortest path from a food source to the nest without using visual cues. A colony of ants works together to identify the shortest path through the use of a pheromone trail. Each ant deposits pheromones while walking, and successive ants will more likely follow a path that is rich in pheromone.

In the optimization technique, the variables are divided into many intervals based on the amount of accuracy required, and each value of the decision variable can be represented conceptually as a node. Artificial ants pass through one node for every variable. The value of the fitness function (i.e., objective function) is determined based on the nodes that were selected by an ant, and this fitness function can then be used to determine the amount of artificial chemical-pheromone that is to be deposited on the path, or at each node. As more ants choose a certain node, or variable value, the concentration of the chemical increases, and other ants will choose that value as part of a solution. Higher pheromone concentrations will attract more ants to that node. In this manner, a set of nodes forming short paths are reinforced as part of highly fit solutions the same way the paths from the ant colony to the food are reinforced in nature.

Variables in the signal control problem represent the length of green times. ACO calls the simulation model (Equations 8–12) to simulate the traffic flows and queues for a given set of green times. The range of green times Equation 7 is explicitly enforced in the optimization model, as the decision variables cannot take values that are outside of the given range. The minimum and maximum values of the variables are to be inputted and the algorithm divides the range into a finite number of intervals. An interval size of 5 sec is used for this application. Ants choose a node number, and the value of the variable is then determined as

$$g = (g_{\min} - \text{interval size}) + (\text{node number} * \text{interval size}) \quad (13)$$

The constrained signal coordination problem is transformed into an unconstrained problem by associating a penalty to the objective function every time the constraint is violated.

Equations 2–6 are checked, and a penalty is added to the objective function to represent any violations. The fitness value of an individual  $i$  is calculated as the objective function  $Z_i$  in Equation 14, where  $\mu_j$  is a penalty coefficient for constraint  $j$ ,  $m$  is the number of implicit constraints, and  $H_j$  denotes  $j$ 's constraint function (inequality and equality). The fitness that is analogous to the length of the path to be travelled by the ant is then minimized.  $C_{\min}$  is an input coefficient introduced to overcome the negative value of the augmented objective function.

$$fitness_i = C_{\min} - \left( Z_i - \sum_{j=1}^m \mu_j H_j \right) \quad (14)$$

The constraint in Equation 5 is not active if a coordinated signal is an open-loop system; that is, when multiple coordinated arterials cross a single coordinated arterial. The augmented objective function becomes as formulated below.

$$\begin{aligned} Max \quad & C_{\min} - \left[ \sum_k^K \sum_{(i,j) \in L} \sum_h^H \frac{d_{i,j}}{d_{\max}} D_{i,j}^h(k) - \sum_k^K \sum_{(i,j) \in L_p} \delta_{i,j}(k) \max \left( \begin{array}{c} 0 \\ q_{i,j}^{h*}(k) - q_{\max} \end{array} \right) \right. \\ & - \mu_1 \sum_{k,(1,j) \in L_p}^K \left( \phi_{i,j}^h(k) - \left( \frac{d_{i,j}}{v_{i,j}} - \frac{(v_{i,j} + \lambda) l_{veh}}{v_{i,j} \lambda} q_{i,j}^{h*}(k) \right) \right)^2 \\ & \left. - \mu_2 \sum_{k,(i,j) \in L_p}^K \max \left( \begin{array}{c} 0 \\ g_i(k) - (g_j(k) + \phi_{i,j}(k) + \beta_{i,j}(k)) \end{array} \right) - \mu_3 \sum_{k,(1,j) \in L}^K \max \left( \begin{array}{c} 0 \\ q_{i,j}^h(k) - \frac{d_{i,j}}{l_{veh}} \end{array} \right) \right] \end{aligned} \quad (15)$$

ACO divides the green times into number of intervals, and ants select one node per variable and form a path. The fitness function is calculated, which is analogous with the distance travelled by the ant. It is possible that early in the search, some paths may be generated which are not optimal, and it is likely that the paths would be reinforced because of the deposition of pheromone. An evaporation function is used to reduce the domination of early un-fit solutions. The pheromone on the entire network slowly evaporates at a constant pace. As a result of this, pheromone on the less often used paths dries up and the paths are not chosen as the search progresses. This enables optimum paths to continue to be selected over nodes that do not comprise fit solutions. For comparison purposes, the ACO algorithm was not stopped based on the functional evaluation or on how close the mean and the best fitness value are. The total number of trials was exhausted in all the cases to ensure generating the best possible values for the given parameters.

## 4.2 GENETIC ALGORITHM

Simple Genetic Algorithm (SGA) is a search technique based on the mechanics of natural selection and natural genetics. SGA transforms the constrained signal coordination problem into an unconstrained problem by associating the penalty with all constraint violations and then maximizes the fitness function (Equation 14). The decision variables, which are green times at each light, are represented using a four-bit binary string that is decoded using Equation 16.

$$g = g_{\min} + \left( \frac{g_{\max} - g_{\min}}{2^{d-1}} \right) DV \quad (16)$$

Where:

$DV$  = decoded value of a string.

We use the same augmented objective function we used in ACO (Equation 15). GA works with individual binary strings that represent the green times. The fitness values are determined and SGA selects the good individuals from the current population and inserts their duplicates into the mating pool. Crossover operation is then executed, which involves randomly choosing two individuals in the mating pool, cutting them at randomly selected site(s) and swapping the parts, thereby creating new individuals. A probability of  $p_c$  is applied to the crossover operation; a sub-set of individuals participate in this operation to preserve some of the good individuals found in the previous population. To escape local optima, mutation operation is carried out. This involves changing a 0 to 1 or vice versa in a binary string with a mutation probability of  $p_m$ . In the successive generations of the SGA, the same process of selection, crossover and mutation is repeated.

For comparison purposes, GA was not stopped based on the functional evaluation or on how close the mean and the best fitness value are. The total number of generations was exhausted in all the cases to ensure generating the best possible values for the given parameters.



## CHAPTER 5. APPLICATION TO ILLUSTRATIVE CASE STUDIES

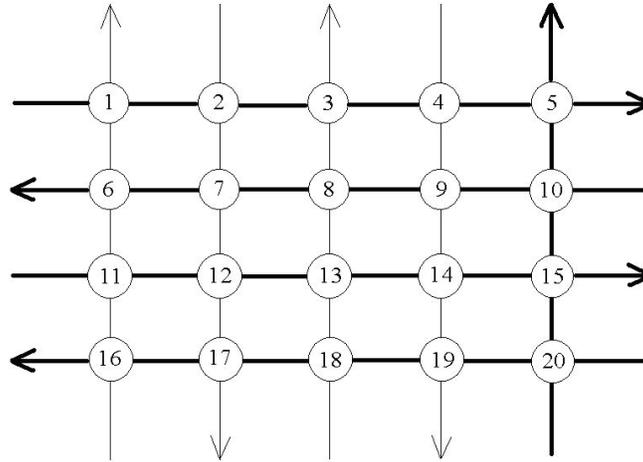
### 5.1 MODEL I

An illustrative network is used to demonstrate the optimization model and its solution using both ACO and GA. The system is a one-way system (Figure 1). Each signal works on a two-phase plan, and no turning movements are allowed. The traffic flows in the system are an eastbound flow from Intersections 1 to 5 and 11 to 15; westbound from Intersections 10 to 6 and 20 to 16; northbound from Intersections 16 to 1, 18 to 3 and 20 to 5; and southbound from Intersections 2 to 17 and 4 to 19. The signals on the northbound arterial from Intersections 20 to 5 and the east-west running arterials are coordinated. The thick lines in the figure represent the coordinated movements.

The northbound arterial from signals 20 to 5 and the westbound arterial from signal 10 to signal 6 are assumed to have a flow rate of 2000 vehicles per hour per lane (vphpl). The westbound arterial from Intersections 20 to 16 and the eastbound arterials are assumed to have a flow rate of 1800 vphpl. The northbound arterials at signals 16 and 18 and the southbound arterials at Intersections 2 and 4 are assumed to have a flow rate of 1500 vphpl. The following parameters are used in modeling the network:

- Lost green time = 5 seconds
- Number of arterial lanes = 2
- Speed limit = 40 ft/sec
- Vehicle deceleration/acceleration = 4 ft/sec<sup>2</sup>
- Saturation flow = 1800 vphpl
- Initial queues = 20 vehicles per lane
- Effective vehicle length = 25 ft
- Starting shock wave speed = 16 ft/sec
- Stopping shock wave speed = 14 ft/sec

Flows and queues are evaluated at sample time ( $\Delta T$ ) of 5 seconds. Total duration of oversaturation is assumed to be 15 minutes. To optimize the network, green times are identified in increments of 5 seconds, and can range between 20 and 80 seconds.



**Figure 1.** One-Way Traffic Network with 20 Signals

## 5.2 MODEL II

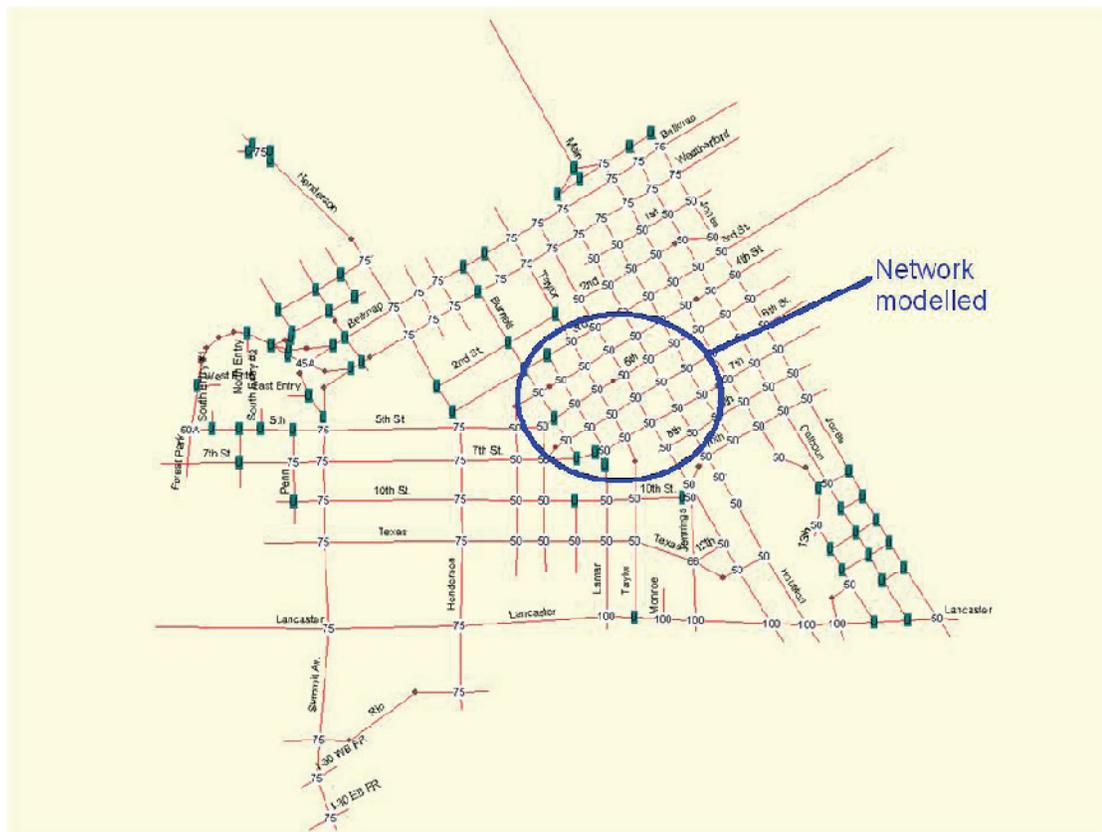
A model of the downtown traffic signal network of the city of Fort Worth was used to demonstrate the algorithm for a realistic traffic network (Figure 2). This model was more complicated than the hypothetical network due to the inclusion of left turning movements, different street lengths, and the varying number of lanes on each arterial.

A  $4 \times 4$  traffic signal network (shown in Figure 3) was selected and the green times were calculated for the grid.

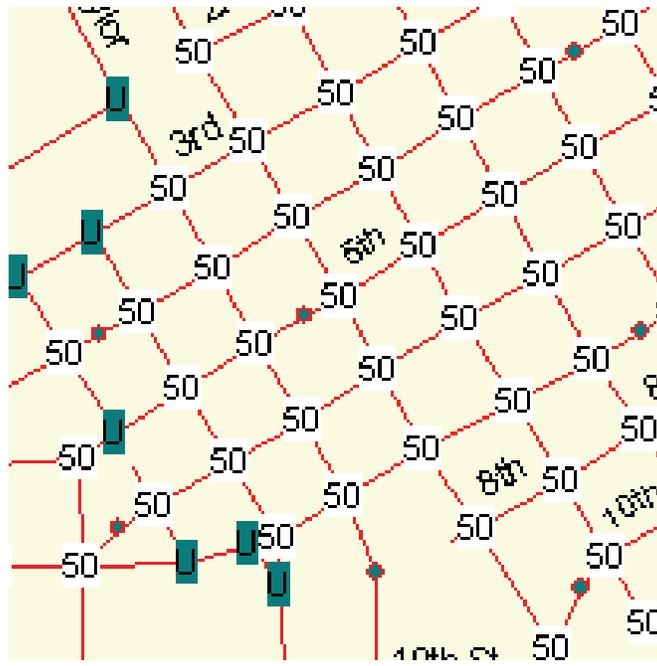
- Lost green time = 5 seconds
- Speed limit = 40 ft/sec
- Vehicle deceleration/acceleration =  $4 \text{ ft/sec}^2$
- Saturation flow = 1800 vphpl
- Initial queues = 2 vehicles per lane
- Effective vehicle length = 25 ft
- Starting shock wave speed = 16 ft/sec
- Stopping shock wave speed = 14 ft/sec

Flows and queues are evaluated at sample time ( $\Delta T$ ) of 5 seconds. The total duration of oversaturation is assumed to be 15 minutes. The volumes during peak hours were used for analysis. In spite of being peak hour volumes, they were not sufficient to emulate oversaturation. They were multiplied with a factor of 10 to simulate oversaturation. To optimize the network,

green times are identified in increments of 5 seconds, and can range between 20 and 80 seconds. Each interval is represented as a node. The number of intervals for each light is  $(80-20)/5=12$ , and the number of nodes is  $12+1=13$ . In a similar fashion, each variable is divided into 13 nodes (in this case because, all the variables are of the same type and have the same ranges).



**Figure 2.** Downtown Network of the City of Fort Worth, TX



**Figure 3.** Traffic Signal Network (4 X 4) used for Model II

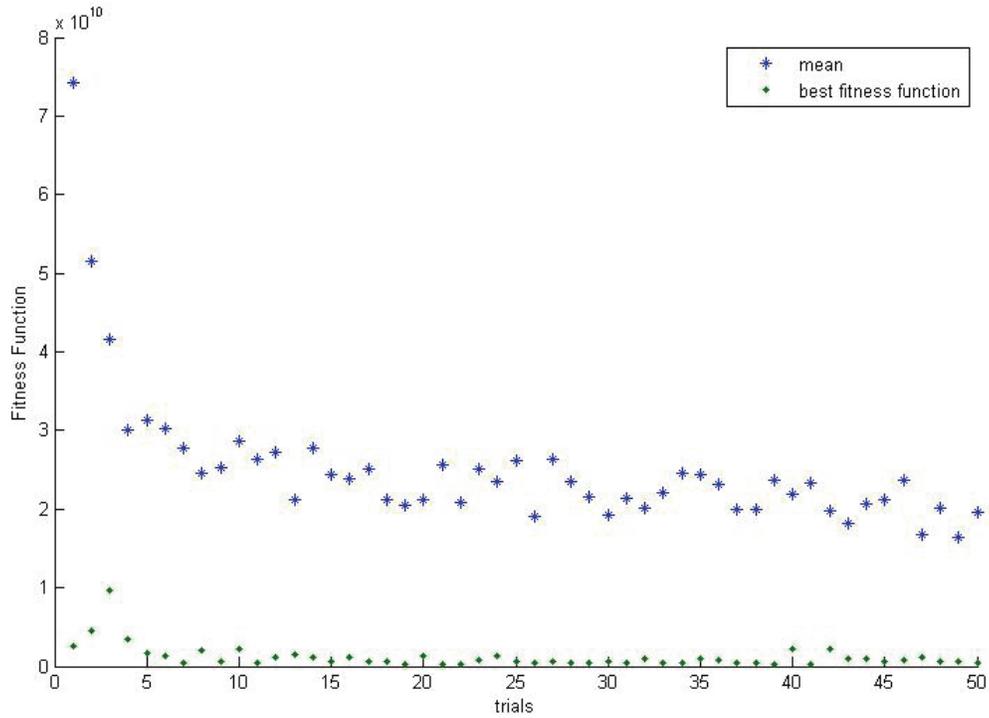
## CHAPTER 6. RESULTS

The network has been analyzed using both Ant Colony Optimization technique and Genetic Algorithms. The MATLAB GA Toolbox has been used for a GA implementation. The crossover factor has been assumed as 0.8, and a Gaussian mutation operator is used. An implementation of ACO was developed in MATLAB, and the evaporation constant has been set as 0.8, based on preliminary testing.

A set of trials were executed to test GA and ACO for different levels of computational power. The population size in GA was analogous to the number of ants in ACO. Similarly, the number of generations was analogous to the number of trials in ACO. The two algorithms were tested for a number of random trials for each case of the same population size/number of ants and number of generations/number of trials. As both algorithms are heuristic methodologies with random components, there is some expected stochasticity in the results. By executing each algorithm for a number of trials with randomized starting points, the reliability of each method to identify a good result can be evaluated. For each case (values for algorithmic parameters), the algorithms were executed 30 times. The cases considered with different settings for the population/ants and generations/trials are listed in Table 1 for both the models. The fitness functions obtained were then compared using Equations 17–19. For example, the convergence plot of ACO for one solution of Model I is shown in Figure. 4 for 400 ants and 50 trials.

**Table 1.** Algorithmic settings for four cases tested for the ACO and GA methods

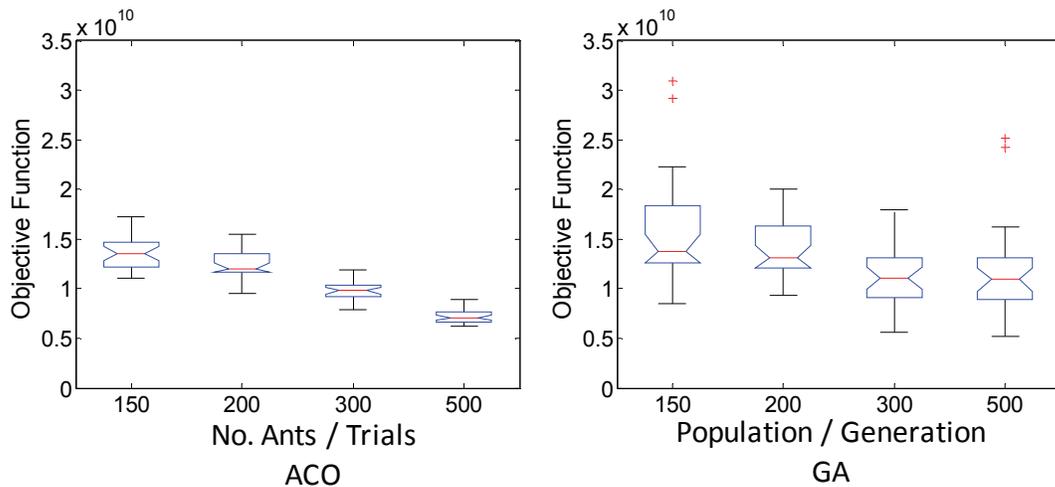
Case	Population/Ants	Generations/Trials
1	150	150
2	200	200
3	300	300
4	500	500



**Figure 4.** Convergence Plot of ACO showing the Fitness Function for the case Ants=400, Trials=50.

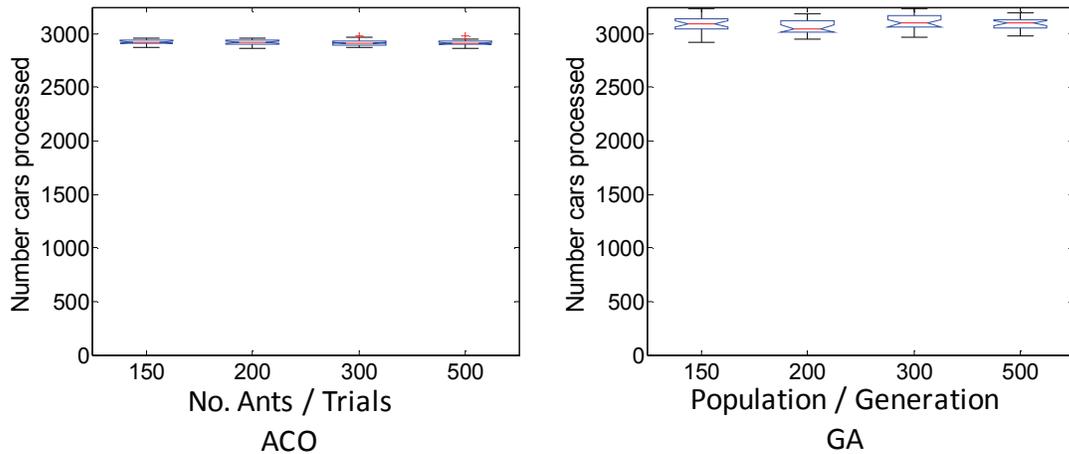
### 6.1 MODEL I RESULTS

The box plots comparing the objective functions (Equation 14) of the best solutions found using ACO and GA with different settings are shown in Figure 5. ACO had a lower objective function than the GA and showed smaller variability in the results.

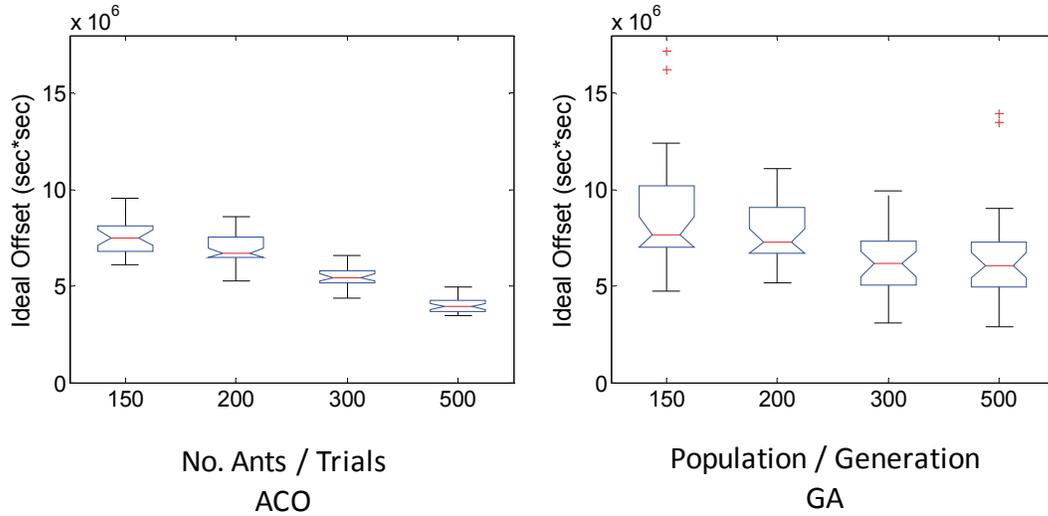


**Figure 5.** Box plots of objective function values (Eq. 14) found using GA and ACO for Model I.

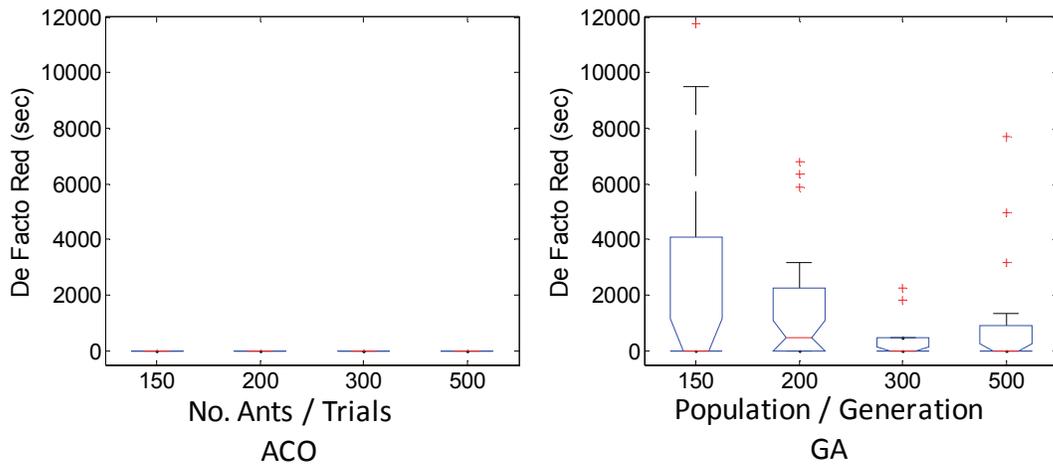
As the objective function represents several different terms to describe different characteristics of the network performance, the terms are separated in the following figures. Figure 6 plots the number of cars processed in the network (first term of Equation 1); Figure 7, the ideal offset (Equation 2) associated with the solutions; and Figure 8, the de facto red time (Equation 4). For all solutions, the queue storage capacity constraint (Equation 6) was satisfied; that is, the penalty was equal to zero. These figures demonstrate that the number of vehicles that pass through the network does not significantly improve as more fit solutions are identified through higher levels of computational power (Figure 6). Instead, the ideal offset is more dramatically decreased (Figure 7). The ideal offset (Figure 7) is shown in units of sec\*sec to represent the deviation from the ideal offset over all intersections in the network. An average ideal offset can be calculated roughly to represent the deviation from the ideal offset at each intersection and at each cycle. This is shown in Figure 9. By applying more computational power, the average ideal offset was decreased from approximately three to two seconds for the ACO, and the GA was not able to minimize this offset as well. In addition, the GA returned solutions with significantly more variability in their fitness achievements. Many solutions identified by the GA had high values for de facto red times, while the solutions identified by ACO had zero de facto red times.



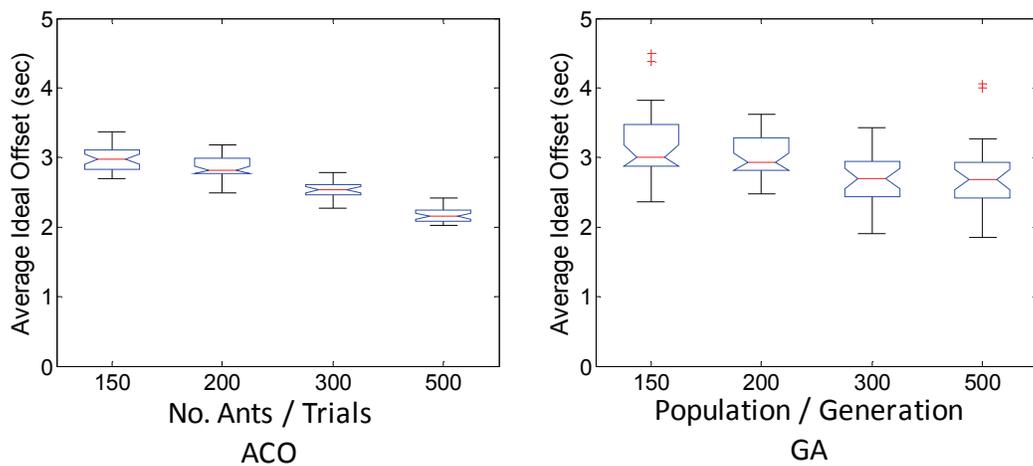
**Figure 6.** Box Plots of number of cars processed (first term of Eq. 1) using GA and ACO for Model I.



**Figure 7.** Box plots of ideal offset (Eq. 2) using GA and ACO for Model I.



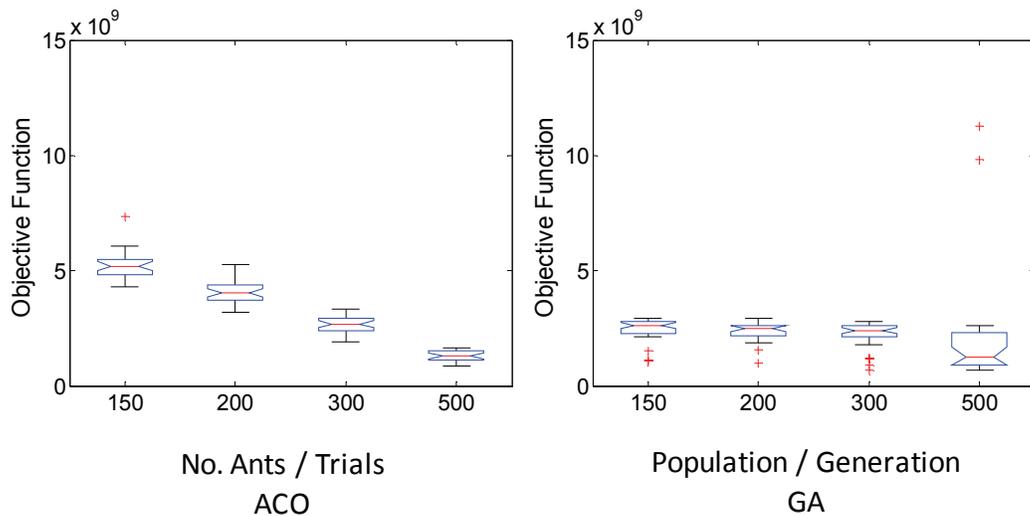
**Figure 8.** Box plots of de facto red processed (Eq. 4) using GA and ACO for Model I.



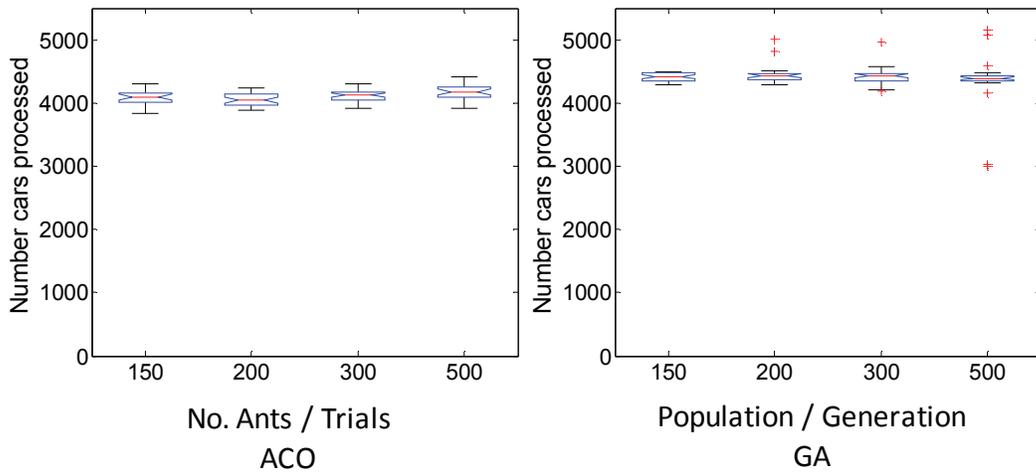
**Figure 9.** Box Plots of average ideal offset using GA and ACO for Model I.

## 6.2 MODEL II RESULTS

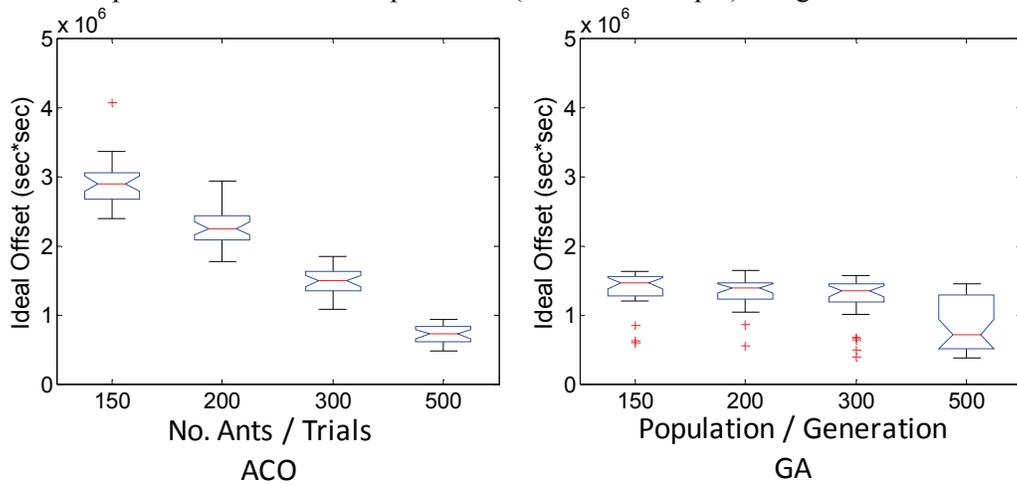
The results for Model II do not result in a significantly different comparison between the GA and ACO approaches. The box plots comparing the objective functions (Equation 14) of the best solutions found using ACO and GA with different settings are shown in Figure 10. For this network, the GA identified more fit solutions for lower number of solution evaluations (e.g., computational power), but for the highest setting, ACO was able to identify more fit solutions. Again, the number of cars passing through the network does not improve with improved fitness values (Figure 11), but the ideal offsets improved dramatically for the increasing computational power allowed for the ACO algorithm (Figure 12). The average ideal offset can be approximated as decreasing from two seconds to one second as the number of ants and trials was increased for the ACO methodology; for the GA, the average ideal offset was improved only when the settings were changed to a population of 500 and 500 generations (Figure 13). GA identified solutions that show considerable variability in the ideal offset and some variability in the number of vehicles processed, but routinely identified solutions with minimal de facto red time (Figure 14) and queue storage time (Figure 15). ACO, on the other hand, identified solutions with variability in the queues storage time and de facto red time, but minimal values for the ideal offset and similar values for the number of vehicles processed through the network.



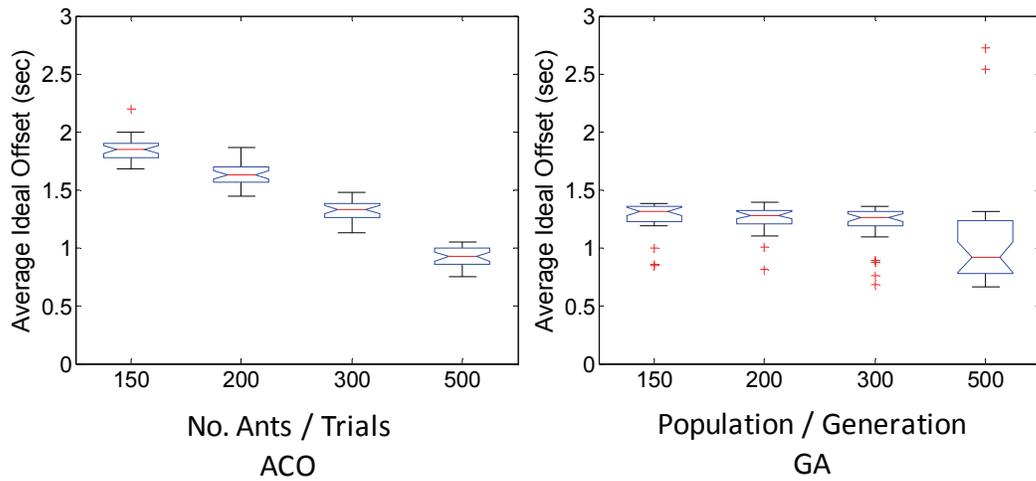
**Figure 10.** Box plots of objective function values found using GA and ACO for Model II.



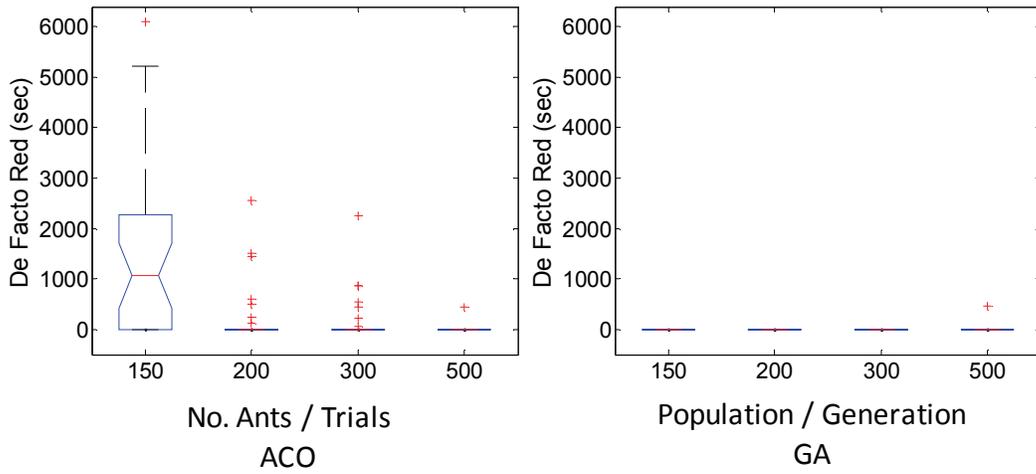
**Figure 11.** Box plots of number of cars processed (first term of Eq. 1) using GA and ACO for Model II.



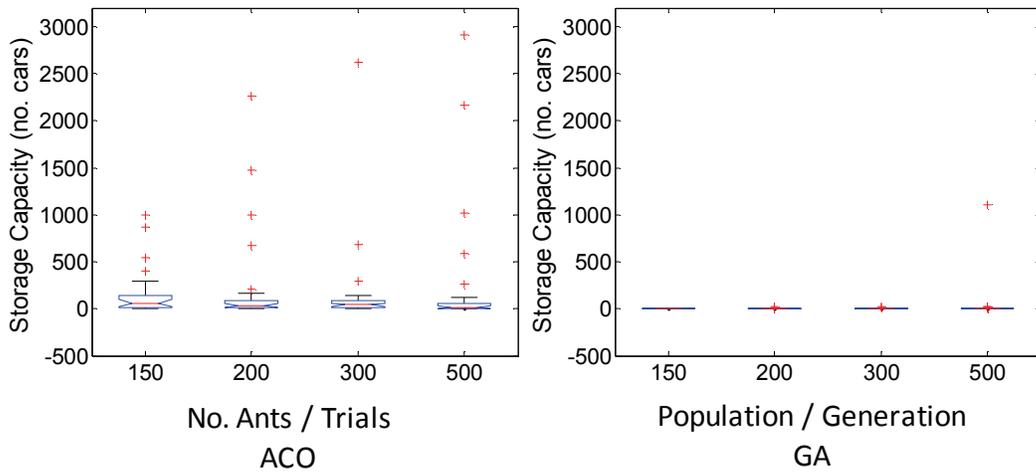
**Figure 12.** Box plots of ideal offset (Eq. 2) using GA and ACO for Model II.



**Figure 13.** Box plots of average ideal offset using GA and ACO for Model II.



**Figure 14.** Box plots of de facto red using GA and ACO for Model II.



**Figure 15.** Box plots of storage capacity using GA and ACO for Model II.



## CHAPTER 7. HYPOTHESIS TESTS

Hypothesis tests were conducted statistically to compare two different systems, GA and ACO, competing to provide the same objective, based on  $n$  simulation runs with the same random number sequences for both systems. The performance measures were  $Y_i$  and  $Y_i'$ , where  $i=1$  to  $n$ , and  $n$  is the number of trials of each algorithm. For this study,  $n = 30$ .

Define  $Z_i = Y_i - Y_i'$  the performance difference at replication  $i$ .

$$Z(n) = \frac{\sum_i Z_i}{n} \quad (17)$$

$$S^2(n) = \frac{\sum_i [Z_i - Z(n)]^2}{n-1} \quad (17)$$

The  $1-\alpha$  confidence interval of the true mean of  $Z$  is contained in the interval

$$Z(n) \pm t_{n-1, 1-\alpha/2} \sqrt{\frac{S^2(n)}{n}} \quad (18)$$

One of the systems is better than the other with  $1-\alpha$  confidence only if the interval in Equation 18 does not contain zero.

The fitness function of ACO was assumed to be  $Y_i$  and the fitness function of GA was assumed to be  $Y_i'$ . The fitness function of GA was subtracted from the fitness function of ACO to give  $Z_i$ , and the number of iterations was 30 for each case. The upper and lower limits of  $Z_i$  were then computed. The smallest value shows the best performance. Positive values for the upper and lower limits of  $Z_i$  indicate, therefore, that the fitness function of ACO was larger (worse) than the fitness function of GA with a confidence interval of 95% ( $\alpha=0.1$ ). This means that GA was better in this case. On the other hand, if the upper and lower limits were negative, it meant that the fitness function of GA was larger than the fitness function of ACO with a confidence interval of 95%, or that ACO was better than GA with 95% confidence.

**Table 2.** Comparison of ACO and GA for Model I

populatio n/ants	generatio ns/trials	Z(n)	n	S <sup>2</sup> (n)	LL	UL	Conclusion
150	150	-2E+09	30	3.1E+19	-3.6E+09	-2E+08	ACO is better
200	200	-2E+09	30	8.54E+18	-2.7E+09	-8E+08	ACO is better
300	300	-1E+09	30	1.084E+19	2.5E+09	-5E+08	ACO is better
500	500	-4E+09	30	1.932E+19	-5.6E+09	-3E+09	ACO is better

**Table 3.** Comparison of ACO and GA for Model II

population/ ants	generation /trials	Z(n)	n	S <sup>2</sup> (n)	LL	UL	Conclusion
150	150	2.7E+09	30	5.5E+17	2.5E+09	3E+09	GA is better
200	200	1.7E+09	30	4.2E+17	1.5E+09	1.9E+09	GA is better
300	300	4.8E+08	30	4.8E+17	2.6E+08	6.9E+08	GA is better
500	500	-8E+08	30	5.7E+18	-2E+09	-8E+07	ACO is better

It was observed for both the models that the fitness function continued to decrease with an increasing number of executions. The case of 500 population/ants and 500 generations/trials resulted in the best performance. ACO performed better for Model I, and particularly better than the GA for the largest number of executions (ants/population=500, generations/trials=500). In Model II, GA outperformed ACO in all cases except the one, which was the case with the highest number of ants and trials (ants/population=500, generations/trials=500). ACO was found to more consistently improve its performance with an increasing number of executions compared to the GA.

As both methods utilize a large number of simulations to solve the problem, the problem of speeding up the computation time should be addressed for both algorithms in further research. Parallel computing can be introduced to shorten the computation time. In ACO, the ants in each trial are independent. A master slave algorithm can be implemented with a master processor keeping track of the pheromone levels on the tracks and the slave processors can run the ants. In ACO, each ant chooses its own path based on the pheromone levels maintained by the master processor. In GA, there may be more dependence among individuals in a population, and the master processor is tasked with more calculations. The master processor has to handle the selection, crossover and mutation operations. Therefore, in using the GA, the master processor must generate new individuals. For ACO, generation of new individuals can be completed by the slave processors, leading to an implementation that would be parallelized to a higher degree than the GA. This facilitates further distribution of the computation time for the ACO.



## CHAPTER 8. CONCLUSIONS AND RECOMMENDATIONS

An algorithm has been developed to solve an oversaturated traffic signal network using Ant Colony Optimization. The algorithm has been compared with Genetic Algorithms to solve the same problem. It was found that for the 30 trials, the results identified by ACO returned reliably low values of the ideal offset, which contributes most significantly to keeping a steady flow of cars through the network and avoiding queues in the streets that are upstream of intersections.

The networks that have been solved using the GA and ACO represent both an illustrative case and a realistic municipality. ACO performed well in comparison with the GA. For the first network, the ACO performed better than the GA, returning reliably good solutions. For the second network, the GA identified more fit solutions for lower number of solution evaluations (e.g., computational power), but for the highest setting, ACO was able to identify more fit solutions. By utilizing increased amounts of computational power, both algorithms were able to improve fitness values of the solutions identified. With improved fitness values, the number of cars passing through the network did not improve, but the ideal offset improved dramatically for the increasing computational power, especially for solutions identified by the ACO algorithm. For the second network, results demonstrated that the two different algorithms identified different types of solutions: GA identified solutions that show considerable variability in the ideal offset and some variability in the number of vehicles processed, but routinely identified solutions with minimal de facto red time and queue storage time. ACO, on the other hand, identified solutions with variability in the queues storage time and de facto red time, but minimal values for the ideal offset and similar values for the number of vehicles processed through the network. Therefore, ACO identified solutions that minimized the more significant objective of planning traffic light timing for oversaturated conditions.

Statistical analysis showed that ACO yielded better results when compared to GA for cases having a higher number of executions. As a result, better performing solutions can be identified for the same computational power. This is especially important when the best possible solution should be obtained and the user can afford the computational time to do so. These results indicate that ACO may prove to be a good alternative when trying to solve very complicated networks.

For the Fort Worth system (Model II), the data did not introduce sufficiently oversaturated conditions, so the traffic congestion was simulated at higher volumes to necessitate efficient planning. In the future, this methodology should be applied to develop traffic signal timing for congested networks across the U.S. In addition, an adaptive strategy may be useful for conditions that change rapidly and unpredictably. An ACO-based approach can be used in a real-time manner to adapt traffic light timings to incoming traffic loads. To better facilitate an adaptive planning process, an ACO methodology can be distributed across cluster computing nodes and utilize high performance computing capabilities. The nature of ACO is inherently parallel and relatively easy to distribute across clusters of computing nodes.

## REFERENCES

- Abu-Lebdeh, G., and Benekohal, F. (1997) Development of traffic control and queue management procedures for oversaturated arterial. *Transportation Research Record*, No. 1603, pp. 119–127.
- Abu-Lebdeh, G., and Benekohal, R.F. (2000) Signal coordination and arterial capacity in oversaturated conditions. *Journal of the Transportation Research Board*, No. 1727, TRB, National Research Council, Washington, D.C., pp. 68–76.
- Burhardt, K.K. (1971) Urban traffic system optimization. Ph.D. Dissertation, University of Minnesota, Minneapolis, Minnesota.
- Ceder, A., and Reshetnik, I. (2001) An algorithm to minimize queues at signalized intersections. *The Journal of the Operational Research Society*, 52(6), pp. 615–622.
- Chang, E.C.P., S.L. Cohen, C.C. Liu, N.A. Chaudhary, and C.J. Messer. (1988) MAXBAND-86: Program for optimizing left-turn phase sequence in multiarterial closed networks. *Transportation Research Record*, No. 1181, pp. 61–67.
- Chang T.-H., Lin J.T. (2000) Optimal signal timing for an oversaturated intersection. *Transportation Research Part B*, 34(6) pp. 471–491.
- Chang, T.-H., and G.Y. Sun. (2004) Modeling and optimization of an oversaturated signalized network. *Transportation Research Part B*, 38(8), pp. 687–707.
- Chen, J., Lihong Xu, Xiaoguang Yang, and Changliang Yuan. (2007) A hierarchy control algorithm and its application in urban arterial control problem. *Proceedings of the 2007 Institute of Electronics and Electrical Engineering Intelligent Transportation Systems Conference*. Seattle, WA, USA, Sept. 30-Oct. 3, 2007, pp. 373–378.
- Dasgupta, D., and Michalewicz, Z. (1997) *Evolutionary Algorithms in Engineering Applications*, Springer-Verlag, Berlin, Germany.
- Dorigo, M., and Thomas, S. (2004) *Ant Colony Optimization*, Cambridge, Mass., MIT Press.
- Gartner, N.H. (1972) Constraining relations among offsets in synchronized signal networks. *Transportation Science*, 6, 88–93.
- Gartner, N.H., S.F. Assman, F. Lasaga, and D.L. Hou. (1990) MULTIBAND - A variable-bandwidth arterial progression scheme. *Transportation Research Record*, No. 1287, pp. 212–222.

- Gartner, N.H., S.F. Assman, F. Lasaga, and D.L. Hou. (1991) A multi-band approach to arterial traffic signal optimization. *Transportation Research Part B*, 25(1), pp. 55–73.
- Gazis, D.C. (1964) Optimal control of a system of oversaturated intersections. *Operations Research*, 12, pp. 815–831.
- Gazis, D.C., Potts, R.B. (1965) The oversaturated intersection. *Proceedings of the Second International Symposium on the Theory of Road Traffic Flow*. Organization for Economic Cooperation and Development, Paris, pp. 221–237.
- Girianna, M., and Benekohal, R.F. (2002) Dynamic signal coordination for networks with oversaturated intersections. *Transportation Research Record*, No. 1811, pp. 122–130.
- Girianna, M., and Benekohal, R.F. (2004) Using genetic algorithms to design signal coordination for oversaturated networks. *Journal of Intelligent Transportation Systems*, 8, pp. 117–129.
- Goldberg, D.E. (1989) *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley.
- Li, M-T; Gan, A. C. (1999) Signal timing optimization for oversaturated networks using TRANSYT-7F. *Transportation Research Record*, No. 1683, pp. 118–126.
- Lieberman, E., and Chang, J. (2005) Optimizing traffic signal timing through network decomposition. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1925, Transportation Research Board of the National Academies, Washington D.C., pp. 167–175.
- Lieberman, E. B., Chang, J. N., Prassas, E.S. (2000) Formulation of real-time control policy for oversaturated arterials. 79th Annual Meeting of the Transportation Research Board, No. 1727, pp. 77-88.
- Little, J.D.C., M.D. Kelson, and N.H. Gartner. (1981) MAXBAND: A program for setting signals on arterials and triangular Networks. *Transportation Research Record*, No. 795, pp. 40–46.
- Little, J.D.C., B. Martin, and J.T. Morgan. (1966) Synchronizing traffic signals for maximal bandwidth. *Highway Research Record*, No. 118, pp. 21–47.
- Little, J.D.C. (1966) The synchronization of traffic signals by mixed-integer linear programming. *Operations Research*, 14, pp. 568–594.

- Lo, H. A. (2001) Cell-based traffic control formulation: strategies and benefits of dynamic timing plans. *Transportation Science*, 35(2), pp. 148–164.
- Kaltenbach, M., Koivo, H.N. (1974) Modeling and control of urban traffic flow. *Proceedings of the Joint Automatic Control Conference*. University of Texas, Houston, Texas, pp. 147–154.
- Michalopoulos, P.G. and Stephanopoulos, G. (1977) Oversaturated signal system with queue length constraints-I. *Transportation Research Record*, No. 11, pp. 413–421.
- Michalopoulos P.G. and Stephanopoulos G. (1978) Optimal control of oversaturated intersections: theoretical and practical considerations. *Traffic Engineering and Control*, 19, pp. 216–222.
- Milton, J.S., and J.C. Arnold. (1995) *Introduction to Probability and Statistics*. McGraw-Hill, Inc., New York.
- Park, B., Messer, C.J., and Urbanik II, Thomas. (1999) Traffic signal optimization program for oversaturated conditions: genetic algorithm approach. *Transportation Research Record*, No. 1683, pp. 133–142.
- Pillai, R.K., A.K. Rathi, and S.L. Cohen. (1998) A restricted branch and bound approach for generating maximum bandwidth signal timing plans for traffic networks. *Transportation Research Part B*, 32(8), pp. 517–529.
- Roess, R.P., McShane, W.R., and Prassas, E.S. (1998) *Traffic Engineering*, Prentice Hall, Upper Saddle River, NJ.
- Stamatiadis, C., and N.H. Gartner (1996) MULTIBAND-96: A program for variable bandwidth progression optimization of multiarterial traffic networks. *Transportation Research Record*, No. 1554, pp. 9–17.
- Yagar, S., and Dion, F. (1996) Distributed approach to real time control of complex signalized networks. *Transportation Research Record*, No. 1554, pp. 1–8.
- Yu, X.-H., and W. Recker. (2006) Stochastic adaptive control model for traffic signal systems. *Transportation Research Part C*, 14(4), pp. 263–282.