A KNOWLEDGE-BASED TRAFFIC SIGNAL CONTROL APPLICATION

A Paper
Submitted to the Graduate Faculty
of the
North Dakota State University
of Agriculture and Applied Science

By

Mohammad Ghaleb Smadi

In Partial Fulfillment of the Requirements
for the Degree of
MASTER OF SCIENCE

Major Department:
Computer Science

December 2001

Fargo, North Dakota

ABSTRACT

Smadi, Mohammad Ghaleb, M.S., Department of Computer Science, College of Science and Mathematics, North Dakota State University, December 2001. A Knowledge-Based Traffic Signal Control Application. Major Professor: Dr. Ahmed Kamel.

This study examines the use of a knowledge-based expert system for traffic signal control. It illustrates a new technique for achieving adaptive traffic signal control which would reduce vehicle delay and travel times. The proposed system was tested on a case study corridor in Fargo, North Dakota, using a traffic simulation program (VISSIM). The results of the testing are presented. Conclusions and suggestions for future research are also offered.

ACKNOWLEDGMENTS

Thanks go to

My major adviser, Dr. Ahmed Kamel, for all the help, guidance, support and patience. My committee members: Dr. Paul Juell, Dr. Ken Magel, and Dr. Don Andersen.

All the folks at the Upper Great Plains Transportation Institute for providing the friendliest and most understanding work environment for the past three years. It was and still is a great experience.

My family that provided me with all the love, encouragement, and support in the world. And always believed in me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure Page

CHAPTER 1. INTRODUCTION

1.1. The Traffic Control Problem

Today's traffic on roadways in urban areas all around the world suffers from congestion and delays. The continuous increase in traffic levels results in heightened demands on the infrastructure of roads, bridges, and highways. Many of the roads operate at full capacity for hours each day during the peek periods. Construction of new roads and highways, and expansion of existing ones can potentially increase the capacity and alleviate the congestion problem, and helps in the reduction of delay. However, new construction projects are not always feasible for reasons of high cost or not being able to accommodate the new expansions. Even when construction projects are feasible, they require closing of roads and detouring traffic away from the construction site for the duration of the project, which disrupts traffic on the road network throughout the entire area.

Since the construction of new infrastructure and the expansion of the existing one do not solely solve traffic problems and are not always feasible, there is a need for better management of existing resources. Better management can be achieved through efficient traffic control. This paper addresses traffic control at intersections using Traffic Signal Control Systems.

Historically, traffic signals have been used at intersections to direct traffic alternately to stop and to move. This sequence of directions is referred to as the signal plan. Traditional traffic signal control systems operate using a pre-specified signal plan based on historical information, regardless of the traffic conditions that are currently occurring on the road. The plan used may not be the best suited to deal with the current traffic conditions, which leads to inefficient traffic control.

To deal with the problem of fixed signal plans in traditional traffic signal control systems, more advanced systems were developed, systems that respond to changing traffic conditions on the roads and, in some cases, predict how the conditions will change. A suitable signal plan to deal with these conditions may then be implemented. These advanced systems are called Adaptive Signal Control Systems.

In the following chapters, when discussing road networks, we will be referring to two types of road networks: arterial networks and grid networks. Arterial networks can be defined as signalized streets that primarily serve through traffic; they are a length of at least 2 miles, with a signalized intersection spacing that ranges from as little as 400 feet to as long as 2 miles, and with turning movements at intersections that usually do not exceed 20% of total traffic volume (TRB 1998). The term “grid network” refers to the system of links and nodes that describes the physical network of roads and rail, where the “physical network” is defined as the spatial configuration of the transportation system which includes the layout of streets and highways, and the layout of bus routes for transit (UC-Berkeley 2001). A grid network is, therefore, made of intersecting roadways where the nodes often represent signalized intersections.

1.2. Description of the Research

In this paper, a new methodology for performing adaptive traffic signal control is researched. The idea is to have a system that observes the traffic conditions on the roads and watches for major changes. Current adaptive signal control systems operate on a high resolution, watching for traffic changes every second and assessing whether a new signal plan needs to be applied at the end of every second. This high-resolution level of operation increases the complexity and the cost of the system. The system proposed in this paper observes the roads for large traffic changes, changes that exceed a predetermined threshold for each approach of the intersection, and can have a significant impact on the signal plan.

The system will have a set of signal plans; each of these plans will be optimized to deal with the traffic conditions at a certain threshold value of traffic change. The signal plans are obtained from an available signal plan optimization program (SYNCHRO; Trafficware 2000). The traffic changes are observed using a Generic Task Hierarchal Classified expert system (Chandrasekaran 1985). The expert system also determines which signal plan should be implemented to deal with the observed traffic change. The performance of the system is tested on a case study network in Fargo, North Dakota; the testing was done using a microscopic traffic simulation program (VISSIM; PTV 2000).

1.3. Importance of the Research

Traffic congestion and delay are serious problems in many metropolitan areas around the world. The current situation strains the road infrastructure, resulting in substantial operations. Construction of new infrastructure to increase capacity is very costly and, in some cases, unfeasible. When new construction projects are undertaken, they often disrupt the traffic flow for the entire road network with road closure and detours. In fact, road work is among the leading causes of delays in the United States.

If congestion and delay can be reduced, great savings can be achieved, such as reductions in the trip times of commuters. Savings can be achieved in the time of transmitting goods as well as shipment costs. Also, reducing congestion will have a positive impact on the environment, with the reduction of vehicle emissions of harmful gases. When considering the improvement of traffic conditions using better traffic management, adaptive signal control systems evolve as the most efficient strategy for controlling traffic at intersections. Updating the signal plans in traditional traffic signal control systems in order to try to keep up with traffic growth is a costly and time consuming process. Even if the signal plans are updated periodically, traditional traffic signal control systems are still incapable of dealing with the short-term changes of traffic flow, such as changes that occur day to day on the intersections due to incidents or other traffic-changing events.

Adaptive traffic signal control systems have been developed, tested, and used for the last three decades, but due to their complexity, high cost, and varying results, they are still not completely accepted in the United States. Another reason for not fully accepting adaptive signal control systems is that the early development of these systems, which resulted in the two most well-known systems (SCOOT and SCATS), was done in the United Kingdom and Australia, respectively. Those systems were developed to deal with the traffic patterns and conditions that are local to these countries, and many believe that SCOOT and SCATS are not best suited to control traffic in the United States.

The system proposed in this paper is less complex than existing adaptive traffic signal control systems. It provides a simplified approach to real-time traffic control, hence achieving the benefits of a real-time adaptive traffic signal control system without having to deal with the high cost and complexity issues. The system, however, is intended for significant changes in traffic (surges) such as those associated with incidents or special events.

1.4. Organization of the Paper

This paper consists of five chapters, including this one. Chapter 1, the Introduction, provides a general introduction to the paper, and describes the research being done and its importance.

Chapter 2, the Literature Review, consists of the following sections:

1. Traffic Signal Control includes definitions and descriptions of the basics and terminology of traffic signal control systems. It also introduces the concept of adaptive signal control systems.
2. Review of Current Adaptive Signal Control Systems includes a description of three of the most used and well known adaptive traffic signal control systems: the SCOOT, SCATS, and RT-TRACS.
3. Knowledge-Based Systems provides a definition and introduction to the principles of expert systems. It introduces Generic Task Expert Systems and contains a discussion of using hierarchal classification systems for traffic control.
4. Artificial intelligence (AI) in traffic introduces possible fields of application for AI in traffic control and the AI techniques that can be applied. It also describes a prototype system that uses a rule-based expert system for adaptive signal control.

Chapter 3, the Methodology, consists of the following sections:

1. Problem Definition: this section provides a description of the problem that the paper is addressing; it also describes the state-of-practice in performing traffic signal control.
2. Approach: this section provides a description of the components of adaptive signal control systems; it goes on to explain the technique that this paper proposes for traffic signal control. The process of signal plans selection is also described along with the system design.

Chapter 4, the Case Study and Testing, consists of the following sections:

1. Description of the Case Study provides a detailed description of the case study network of 25th St. South in Fargo, North Dakota. The section also explains why this particular network was used and the data used in the case study.
2. System Testing describes the testing process of the proposed system and the details of the test case.

Chapter 5, the Results, Conclusion, and Future Work, consists of the following sections:

1. Results provides the results of the simulation testing, along with explanation of these results.
2. Conclusion and Future Work provides conclusions and offers possibilities of future work.

CHAPTER 2. LITERATURE REVIEW

This chapter provides a brief discussion of traffic signal control and reviews three traffic signal control systems. It also introduces the basic concepts of knowledge-based systems.

2.1. Traffic Signal Control

2.1.1. Traffic Signal Control Basics and Terminology

Road traffic on intersections is controlled using Traffic Signal Control Systems; these systems determine which direction (approach) of the intersection has the right of way and for how long. There are mainly two types of signal controllers: pretimed and actuated. Both types deal with a set of parameters for traffic control: cycle length, phase sequence, and phase length. A fourth parameter is offset, which is used for coordinating the signal operation at an intersection with other adjacent intersections.

These parameters can be defined as follows (McShane et al. 1998):

1. Cycle length is the time in seconds it takes a signal to complete one full cycle of signal indications.
2. Interval is a period during which none of the lights at a signalized intersection changes. In a signal cycle, the following intervals can be found:
    - change interval: the yellow indication for a certain movement.
    - green interval: the green indication for a particular movement or set of movements.
    - red interval: the red indication for a particular movement or set of movements.

- clearance interval (all red): a period after each yellow interval during which all signal faces show a red indication.

3. Phase is a set of intervals that allows a designated movement or a set of movements to flow, and to be halted safely before release of another set of movements, such as green interval plus the change and clearance intervals that follow it.
4. Offset is the time difference between the start of the green phase at an upstream intersection as related to the start of the green phase at an adjacent downstream intersection.

In pretimed controllers, these parameters are constant all the time. The signal functions without regard to demand patterns and requires no detection. In the case of actuated controllers, the signal operation is affected by current demand in each cycle. Therefore, the signal parameters change according to the demand within set maximums. Hence, actuated controllers require detection to provide the controller with current information on vehicle arrivals.

Both pretimed and actuated controllers can provide several different timing patterns or plans, usually referred to as time-of-day plans. Time-of-day plans normally include an AM peak plan to deal with the morning rush hour, a midday plan for the lunch rush hour, a PM peak plan for the evening rush hour, and a non-peak plan for the rest of the day. The aim of having these different plans is to try to account for varying traffic patterns that occur throughout the day. Signal plan and cycle length remain constant during a particular plan. Traffic signal control systems usually consist of three components:

1. Traffic controller: this device contains the logic, or rules, for performing traffic control and information like the signal parameter values. In early stages of traffic signal control, the traffic controllers were electro-mechanical devices. Nowdays, traffic signal controllers house computers that perform the signal operations.
2. Display devices: the signal heads with the red, yellow, and green lights for phase display at the intersection.
3. Detection devices: devices that detect arrival events at a certain point, such as an intersection. Variations of detection devices include loop detectors embedded in the pavement, video detection, as well as laser and radar detection.

2.1.2. Adaptive Signal Control

The traditional traffic signal control system with the time-of-day plans do not provide enough flexibility to deal with all traffic variations. Signal plans are applied solely depending on time of day and do not take the traffic that is actually on the roads at that time into account. In addition, these plans quickly become outdated because of increases and decreases in traffic volumes due to construction projects, incidents, sporting events, or other reasons. Knowing that a signal that is operating on an outdated plan could worsen delays and traffic congestions calls for the expensive process of periodically updating traffic plans for the signal controllers in order to keep the plans as optimal as possible. Updating the signal plans involves the timely and costly process of data collection, running optimization software for the new plans, and usually testing the plans prior to implementation. The issue of when to carry out these traffic plan updates then arises.

As a solution, a smarter kind of controller is promised, one that adapts to the traffic patterns currently occurring on the roads by first detecting them and then applying the plan which best deals with them. Such a system is called an Adaptive Signal Controller.

Adaptive signal control systems are traffic control systems that adapt to the traffic patterns currently existing on the roads with the plans best suited for dealing with those particular patterns. Such systems require extensive detection, complex estimation and prediction components, and a strong communication network for transfer of data between intersections controlled by the system and, in some cases, between a central computer and the intersections and vice versa.

2.2. Review of Current Adaptive Signal Control Systems

During the last three decades, there have been several attempts to develop an effective adaptive signal control system. The following sections provide a brief description of some of the most well-known and most applied systems, how they work, their applicability, and their performance.

2.2.1. Split, Cycle, and Offset Optimization Technique (SCOOT)

The SCOOT system was developed in the UK in the mid 1970s by collaboration between the Transportation Research Laboratory (TRL) and the UK traffic system suppliers. SCOOT employs detectors at all controlled intersections; the data produced by these detectors are processed by a central coordinating computer which may decide to change the green split, cycle length, or offset at any of the controlled intersections. These changes to the signal plans are made frequently (typically every phase or cycle) and gradually (usually a few seconds at a time).

2.2.1.1. System Architecture

The SCOOT system has a centralized architecture that consists of the following components:

1. Central computer: the central computer contains the timing algorithms and is where all the analysis is done.
2. Local controllers: the local controllers at the intersection only deal with clearance intervals and minimums; they receive plans from the central computer and apply them at the intersection. SCOOT can operate with most modern general-purpose traffic controllers and does not require special controllers.
3. Detectors: the detectors capture traffic patterns in the network; SCOOT requires a special detector-placing scheme. Good detector data are crucial for the system to operate effectively.
4. Communication system: the communication system relays data among the above three components; there is a need for a reliable, dedicated communication medium, such as a leased line, copper, or fiber optic cable.
5. Data requirements
    - Detection is needed on every link in the network that requires optimization.
    - Detectors need to be located about 30 feet from the upstream end of the link.
    - Connection to the central computer is made through the upstream intersection.

- Links with no detection run fixed time plans or can have data derived from upstream links.
- The network is divided into regions, each containing a number of nodes, or signalized intersections, that run at the same cycle length to allow coordination.

2.2.1.2. How SCOOT works (Bretherton 1999)

The following list describes the operation of SCOOT:

1. The detectors are checked for vehicle occupancy four times every second.
2. Occupancy data are relayed every second to the central computer in the form of four-bit streams for each detector.
3. The detector data are processed in two ways:
    - Congestion measure: a count of the number of 4-second intervals per unit time that a loop is occupied. This measure is used for offset optimization.
    - Traffic flow measure: internally, SCOOT measures flow in a unit called a Link Profile Unit (LPU) which are hybrid of link flow and occupancy. The process of converting occupancy information into flow values is a commercial secret of SCOOT.
4. LPU flow information is stored in histograms of flow in LPUs against time in seconds for one cycle. These histograms are called Cyclic Flow Profiles (CFP).
5. The CFP indicates vehicle arrival time at the stop line and is used to predict traffic delays, queues, and stops.

2.2.1.2.1. Delay, stops, and queue prediction

SCOOT uses a model that assumes all detected LPUs at the loop will proceed along the link at the cruise speed (defined as the average speed that vehicles travel between intersections (SCOOT and UTC 2000); when an LPU reaches the downstream end of the link, it may join the back of the modeled queue or proceed through the intersection if the signal is green. The delay of an LPU is the time spent in the modeled queues. According to these predicted values of delay and queue lengths, SCOOT uses its three optimization procedures to evaluate the current signal plan and suggest changes to the plan if any are needed.

2.2.1.2.2. SCOOT optimizers

SCOOT uses the following optimizers:

1. Split optimizer: it operates on the phase level; before every phase change, it analyzes the red and green timings to determine whether to leave the phase start as it is, advance it, or delay it by a few seconds. The split optimizer works in increments of 1 to 4 seconds.
2. Offset optimizer: it operates on the cycle level at a predetermined phase. In every cycle, the offset optimizer uses the CFPs to determine whether the existing offset should remain the same, or be increased or decreased in 4-second increments.
3. Cycle length optimizer: it operates on the region level, and runs every 2.5 or 5 minutes. The cycle length optimizer identifies the “critical node” (most heavily trafficked intersection) in the region and attempts to adjust the cycle length to maintain this node at 90% link saturation at each phase; it can increase or decrease the cycle time in 4-, 8-, or 16-second increments.

2.2.1.3. Applicability and Performance

SCOOT can be applied both to arterial streets and grid networks with a size up to 1000 intersections. It is used in about 170 towns and cities around the world, with a large application base in the United Kingdom, and is considered the most widely used adaptive signal control system.

SCOOT has the following advantages:

1. SCOOT is based on TRANSYT, which is considered a rigorous traffic optimization model that gives SCOOT strength.
2. SCOOT performs well in heavy, close to saturation flows; complex flow patterns; and unpredictable variations.

SCOOT has the following disadvantages:

1. SCOOT’s performance is based on good detection data. SCOOT can accommodate up to 15% detector failure, but if the faults are not detected and rectified, the performance degrades to fixed time plan levels.
2. SCOOT’s scoop is urban; network freeway interaction is still unknown.
3. SCOOT has a high level of sophistication, and expertise is needed before being able to set and change the options in the system. SCOOT’s internal operations cannot be altered since the kernel is impenetrable, but access to the kernel is available with a special $150,000 license.

2.2.2. Sydney Coordinated Adaptive Traffic System (SCATS)

The SCATS (Martin 2001) system was developed over a 20-year period by the Roads and Traffic Authority of New South Wales, Australia. SCATS uses data from detectors at the stop line of the intersection to calculate the degree of saturation. SCATS then uses that measurement to minimize stops, delay, and travel time.

2.2.2.1. System Architecture

The SCATS system has a centralized architecture but is more distributed than SCOOT. It uses a hierarchy of computer systems to perform signal control; SCATS consists of the following components:

1. Central management system: the central management system consists of the SCATS Management Computer and several operator workstations. The management computer handles communication and database functions.
2. Regional computers: the regional computer handles strategic control and data accumulation; all detector and intersection performance data are stored on a day-by-day basis. The SCATS system can have up to 32 regional computers controlled by 1 management computer; each regional computer handles the traffic controllers in a certain region.
3. Traffic controllers: the local traffic controller handles tactical control, such as phase demand and extension, as well as data collection and fault reporting. The SCATS system can have up to 250 controllers per regional computer. SCATS requires a traffic controller with SCATS functionality; some general-purpose controllers can

be upgraded and equipped with a relay module to be able to function with the SCATS system.

4. Communication system: SCATS requires a local area network of extensive point-to-point and multi-drop communication.
5. Data requirements
    a. Stop line detection on each lane of every link in the network.
    b. The controller collects the number of spaces and total space time during green time of each phase and cycle for use in the SCATS adaptive algorithm. Spacing is defined as the distance between successive vehicles on a traffic lane.
    c. The stop line detection produces accurate turning movement data, which allows for accurate split determination.

2.2.2.2. How SCATS works

The network of intersections is divided into regions based on flow characteristics; intersections with similar flow characteristics are grouped into a region; each region is divided into links and nodes. SCATS performs the following calculations to determine cycle length, split plans, and offsets:

1. SCATS constantly calculates the degree of saturation (DS) for all nodes; DS is defined as the ratio of efficiently used phase time to the available phase time.

$$DS = \frac{\text{green - unused green}}{\text{green}},$$

where green is available phase time,

unused green is a measure of efficiency (zero at saturation flow, positive at under-saturation, and negative at over-saturation flow), and

unused green = ((total space time from controller) – (number of spaces) ∗ (standard space time at max flow)).

The unused green parameter is used for cycle length and split plan calculation.

2. SCATS calculates a third parameter for offset selection; this parameter is Car Equivalent Flow (VK). VK is derived from the DS and saturation flow for each lane.

   VK = DS x (green time) x (vehicles per second at max flow).

3. SCATS identifies the most critical node for each region, which is the node with the highest degree of saturation. The region boundaries can change according to flow characteristics, so nodes can be added and removed to a region dynamically, and also, regions can merge.

2.2.2.2.1. Cycle length selection

SCATS has user-defined equilibrium DS values to determine the relationship between measured DS and the cycle length; this relationship is then used to select a target cycle length that the actual used cycle length moves toward. The cycle length can approach the target length by increments of up to +/- 6 seconds. The calibration and choice of the equilibrium DS value and other user-defined values are essential to effective performance, and these skills are achieved through years of experience on the user side.

2.2.2.2.2. Split plan selection

SCATS has a library of split plans that it examines each cycle to determine the most "equisat" plan for the next cycle. Equisat is the state where the DS at the approaches of critical nodes is equal. The maximum projected DS values are calculated for each plan using the last cycle DS values, and then, the plan with the lowest DS maximum is selected. That plan will be the one that provides the minimal delay.

2.2.2.2.3. Offset selection

SCATS has a library of offset options driven by the cycle length. The engineer or operator is allowed a high degree of adjustments. SCATS calculates a parameter for different offset plans for each link; this parameter is calculated by the multiplication of the equivalent flow value (VK) by the directional bias value (DB) for each offset plan. Based on this parameter, eventually, the plan that results in the highest value is chosen.

2.2.2.3. Applicability and Performance

SCATS can be applied to arterials as well to grid networks. SCATS is used in about 5000 intersections worldwide; about 80% of these intersections are in Australia (Piotrowic 2001).

SCATS has the following advantages:

1. SCATS performs well in heavy, close to saturation flows; complex flow patterns; and unpredictable variations.
2. SCATS is very good with split optimization thanks to turning movement data from the stop line detection.

3. SCATS recognizes faulty detectors.

SCATS has the following disadvantages:

1. SCATS needs intimate network knowledge for installation.
2. SCATS is not model based, and it needs trained specialist staff.
3. SCATS is controller specific; it cannot modify most existing controllers.
4. SCATS does not do good at optimizing offsets, which impacts progression and coordination.

2.2.3. Real-Time Traffic Adaptive Control System (RT-TRACS)

RT-TRACS (Andrews 1997) is a project sponsored by the Federal Highway Administration (FHWA). It includes five agencies working on five different algorithms for adaptive traffic control with the goal of creating a more proactive approach that other existing adaptive traffic control systems, such as SCOOT and SCATS, lack. The reason for having five different algorithms is to address different geometric and traffic conditions with the appropriate control strategies; the algorithms are to be incorporated into an overall RT-TRACS logic with a comprehensive system that includes tools for database management. The RT-TRACS system will also include an expert system that will be able to determine, in real-time, what strategies are best suited for current traffic conditions (Andrews 1997). At this point, two of the five systems that will eventually make up RT-TRACS are being developed and have been initially tested. The following is a description of these two systems.

2.2.3.1. Optimized Policies for Adaptive Control (OPAC)

The OPAC (Pooran 2001) system is being developed by PB Farradyne's special group for intelligent transportation systems; the development started in the late 1970s; first versions of OPAC were used for traffic control on isolated intersections. OPAC IV or Virtual Fixed Cycle OPAC (VFC-OPAC) is developed for real-time adaptive traffic signal control.

OPAC is a distributed real-time traffic control system that is designed for arterials under saturated conditions. OPAC continuously adapts signal timings to minimize a performance function of total intersection delay and stops.

2.2.3.1.1. System Architecture

The OPAC system control strategy was developed primarily for both isolated and arterial control. For isolated intersection control, the system is fully distributed (Pooran 2001). For arterial coordinated system control, the system is distributed except for the following tasks:

1. Cycle length determination, which is made at the central system.
2. Peer-to-peer information can be communicated through the central system if adjacent intersection controllers are not linked physically.

The system consists of the following components:

1. Central system: the central system is used for cycle length determination when OPAC is used for coordinated arterial control and for communication between controllers.

2. Local controllers: OPAC operates with existing controllers that are used in the United States; the controllers need to be equipped with a VME or Pentium processor to run the OPAC software.
3. Detectors: OPAC requires extensive detection to capture traffic patterns.
4. Communication system: the communication system is used to relay data between the different system components.
5. Data requirements
    - Detection on each lane of through phases located about 10 seconds upstream of the stop line or upstream of the worst queue.
    - OPAC also requires one count detector on each lane of left turn pockets as far upstream as possible.
    - OPAC will measure volume, occupancy, and speed from detector data.

2.2.3.1.2. How OPAC works

The following list describes the operation of OPAC:

- For split optimization, OPAC optimizes a weighted performance function of a control variable subject to the constraints of the minimum and maximum green time. The control variable of this function can be total intersection delay, total intersection stops, or both variables.
- When OPAC is used for network coordination optimization, the signal timing optimization function is additionally constrained by the current cycle length.
- For cycle length optimization, OPAC identifies the critical intersection in the network. OPAC uses the Virtual Fixed Cycle (VCF) concept where the network

cycle length meets phase switching timing that is determined locally at each of the controllers depending on the conditions at that intersection; at the same time, the cycle length is chosen to provide the capability for coordination with adjacent intersections. Using a cycle length constraint, the cycle length can start or terminate only within a prescribed range.

2.2.3.1.3. Applicability and Performance

OPAC was developed for optimization of isolated intersections and arterial coordination. OPAC is still under initial testing; it is being tested in New Brunswick, New Jersey, in an arterial of 15 intersections; 4 intersections were chosen for isolated intersection optimization testing, and the remaining segment of 11 intersections was used for arterial coordination optimization testing.

More testing is needed to evaluate OPAC's performance, but results from this initial study in a before and after comparison show the following:

1. OPAC performed best under over-saturated conditions (PM peak south bound movement on the arterial).
2. OPAC did not achieve improvements under light demand and moderately saturated conditions (AM peak on the arterial).
3. OPAC improved the performance of the isolated intersection under the moderate traffic condition.
4. While OPAC achieved significant improvements on the southbound approach of the arterial in the PM peak, the performance degraded on the northbound

approach and the side streets throughout the network; PB claims the improvements made up for the degradation.

2.2.3.2. Real-time, Hierarchical, Optimized, Distributed, Effective, System for Traffic Control (RHODES)

The RHODES (Head and Mirchandani 1998) system is being developed by the University of Arizona and Gardner Systems; the development began in 1991. RHODES is designed to deal with both arterial and network grids; it is intended to respond proactively to the natural stochastic variations in traffic flows. RHODES operates within the framework of North American traffic signal controllers.

2.2.3.2.1. System Architecture

The RHODES system has a hierarchal distributed architecture; it allows for local intersection control. RHODES has been designed to operate as an extension of existing Advanced Traffic Management Systems (ATMS) with requirements for additional communication, detection, and processing. The ATMS architecture includes a network of servers and workstations, with field communication between the traffic controllers and the servers.

2.2.3.2.2. Data Requirements

The RHODES system has the following data requirements:

1. Detection: detection for RHODES requires two types of detectors:
    - Passage: located upstream; used for volume counts, flow prediction, and queue estimation.

- Presence: located at the stop line; used to determine if there is queuing at the intersection. The state of the detector is pulled every second if the presence detector is not occupied, then there is no queue.
- RHODES can accommodate a range of detector locations.

2. Prediction: for RHODES to perform properly, it needs to be able to predict vehicle flows over a certain horizon, or time period, of interest. RHODES uses detector data to perform these predictions.

2.2.3.2.2. How RHODES Works

RHODES' hierarchal design decomposes the traffic control problem into three sub-problems: network loading, network flow control, and intersection control. RHODES has algorithms that are used in real time to solve the problems at each level of the hierarchy based on the inputs from the detectors throughout the network (Lucas et al. 2000).

RHODES has a set of control variables that feed input into its algorithms; these variables can be divided into the following categories:

1. Structural parameters: geometric description of lanes, turning pockets, and detector locations.
2. Traffic dynamics parameters: turning percentages, queues discharge rates, and free flow speed on links.
3. Signal control parameters: the phase with the allowable movements, the minimum and max green values, and the yellow and red clearance times.

4. Optimization parameters
    - Phase order.
    - Length of the prediction horizon, user defined (currently, 45 seconds horizon is used).
    - Resolution: how often are the detectors being polled for data? (Can be as high as one second).

RHODES’ algorithms use the following set of measures of effectiveness (MOE) to evaluate the network performance:

1. An estimate of the queue size in number of vehicles.
2. Predicted link flow profiles, based on detector data and the free flow speed.
3. Predicted delay, based on current queue and predicted arrivals.

The actual algorithms and how the prediction itself works are internal to RHODES.

2.2.3.2.3. Applicability and Performance (Head et al. 1998)

RHODES is designed for both arterials and network grids. Currently, most of the experience and testing of RHODES is carried out on arterials. Field testing of RHODES is still underway; initial testing is being done on two intersections in Tempe, Arizona. However, RHODES has been tested using a traffic simulation program, CORSIM. As part of the RT-TRACS prototype evaluation process, ITT Systems and Sciences Corporation used an FHWA CORSIM test case based on an 11-mile stretch of Tara Boulevard in Atlanta, Georgia, consisting of a 9-intersection arterial. Tests have been done in a software-only approach where no actual controller hardware is used. Instead, hardware-in-the-loop simulation was used where a traffic controller is interfaced with the simulation program for testing. Results of the evaluation indicated that RHODES was able to reduce the average vehicle delay without negatively affecting the throughput (Head et al. 1998).

2.2.4. Discussion

With increasing traffic volumes and constantly varying traffic patterns, effective traffic signal control requires adaptive traffic control systems. The systems that are currently in use have different degrees of success and, in some cases, no success at all. The systems are complicated and require efforts to understand them and have them operating properly. The cost of these systems runs high. The SCOOT and SCATS systems have a centralized architecture with special hardware requirements for the central computer system. They also require specific detector configuration schemes where the detectors need to be properly located and sized in each link of the controlled area, or else the performance of the systems is strongly affected.

Existing adaptive traffic control systems operate on a high resolution; in most cases, the detectors are polled for data every second. A high resolution might be a good way for staying on top of any changes that the traffic in the network might be undergoing; in fact, SCOOT has a proactive approach where it predicts how the traffic is going to change and suggests appropriate signal plans for these changes, however, operating on a high resolution increases the complexity of the system and the computational overhead.

A system can be proposed where it observes the network for large, significant traffic surges based on predetermined threshold values, and then, the system can employ a suitable signal plan for that level of traffic demand increase from a library of signal plans that have been pre-developed to accommodate all, or most, possible traffic changes.

2.3. Knowledge-Based Systems

2.3.1. Expert Systems: Definition and Principles

Expert systems can be defined as practical computer programs that use heuristic strategies developed by humans to solve specific classes of problems (Luger and Stubblefield 1998). Expert systems are developed through cooperation between an expert who provides expert knowledge in a problem field and a knowledge engineer who codes the expert knowledge into a form that a computer program can utilize to solve problems in that field. Expert system programs should have the following dimensions (Chandrasekaran 1985):

1. Expertise: expert knowledge in a certain problem domain.
2. Search: a search engine to search through the knowledge base for solutions.
3. Uncertainty: the ability to deal with uncertainty where a parameter cannot be assigned the value of true or false.
4. Symbolic knowledge structure: the domain knowledge representation in a symbolic form as a collection of facts, rules, or cases.
5. Explanation capability: explanation of the reasoning process through presenting intermediate solution steps and answering questions about the solution process.

Expert systems can be applied more successfully to well-studied domains that have clearly defined problem solving strategies rather than problems that rely on the loosely defined notion of common sense.

First-generation (rule-based, model-based, and case-based) expert systems focus on representation and separate control from the knowledge. They emphasize the power of knowledge itself over the problem-solving method, relying on the inference engine to search through the knowledge base and find a solution.

The lesson that was learned from first-generation expert systems was that it is difficult to separate knowledge from its use. Second generation expert systems, such as the generic task approach of expert systems, have frameworks set where different types of problem-solving methods can be tied to the types of knowledge organizations required for them. This way, there is no need to separate the knowledge from how it is used. The result will be a number of techniques that are known to work for problem solving in a certain domain.

2.3.2. Generic Task Expert Systems

The idea behind the generic task approach evolved from a theory that proposes that there exists a number of well-defined generic tasks where each calls for a certain organizational and problem solving structure (Chandrasekaran 1985). Several generic tasks were identified through work in problem-solving in the domains of medicine and reason engineered systems.

By identifying such generic tasks, a framework to identify the capabilities and applicability of expert systems can be achieved. If a real-world problem can be reduced to a number of generic tasks and if, for each generic task, the knowledge of how to build a reasoning system exists, then it can be concluded that the problem domain can be dealt with using an expert system. Examples of these generic tasks include:

1. The classificatory task: the task that classifies a complex case description as a node in the system hierarchy. This task is generic because it is a component of

many real-world, problem solving situations. Often in problem solving, there is a need to classify a situation into a case or number of cases that we know how to solve and then apply the solution method to the cases that result from this classification.

2. The What Will Happen If (WWHI-type) task: the task that attempts to predict the consequences of an action that might be taken on a complex system. This task is useful as a subtask in an expert system where the consequences of suggested actions from another system, or subsystem, can be evaluated to determine whether an action should be taken.
3. The knowledge-directed retrieval task: the task that, through associative memory, retrieves information by reasoning about other related information. This task was used to create an intelligent database system.

In the classification task, the knowledge structures and corresponding inference methods are closely intertwined and combined together in the concept of the "specialist." The task is reduced into a hierarchy of specialists with establish-refine reasoning. Each specialist tries to establish or reject itself; if it succeeds in establishing itself, the refinement process consists of seeing which of its successors can establish itself. The way in which a specialist attempts to do the establish-refine reasoning process may vary in different domains; some specialists may accomplish the reasoning by using knowledge in the form of rules; others represent the establish-refine activity in the form of functional knowledge about specific modules.

2.3.3. Hierarchal Classification for Traffic Control

The real-world task of traffic signal control requires having signals with signal plans that are suitable for the traffic conditions currently existing on the roads; signal plans need to change as the traffic patterns on roads change. Traffic can increase or decrease in a certain direction; controlling traffic with new demands using an outdated signal plan causes a drastic degradation in system performance (Park et al. 2000). To keep the signal plans up to date, there is a need to identify the traffic changes and then to find the signal plan best suited for the new condition. The process of identifying the patterns that traffic changes into can be done using a hierarchal classification generic task expert system.

The classification task is the identification of a case description with a specific node in a predetermined diagnostic hierarchy (Chandrasekaran 1985). At each node of the classification hierarchy, a specialist can be identified; the diagnostic knowledge of the system can be distributed through the conceptual nodes of the hierarchy. The problem solving for the classification task will be performed in a top-down approach; the top-most specialist will first get control of the case; then control is passed down to an appropriate sub-specialist; etc. With this structure, more general classificatory specialists are located higher in the hierarchy, and more particular ones are lower in it. The problem-solving, or the classification, that goes on in such a structure is distributed; the problem-solving regime that is implicit to the structure is the establish-refine type.

Each specialist has several clusters of rules: confirmatory rules, exclusionary rules, and possibly recommendation rules. The evidence for confirmation and exclusion is weighed to arrive at a conclusion to establish or reject the specialist.

Hierarchal classification has been used in diagnostics for the design of medical expert systems. By analogy, this method can be applied to a traffic control system where there is a need to “diagnose,” or identify, the traffic pattern change on the roads and to prescribe the suitable signal plan to deal with that change. In a traffic control system, there is a hierarchy of intersections, approaches, and movements. At the movement level of the hierarchy, the system can contain several cases of traffic volume increments to deal with the possible traffic increases. The volume increment at a movement in a certain approach will then trigger the signal plan that is embedded in the specialist and is best suited to deal with that traffic change.

2.4. Artificial Intelligence Applications to Traffic Control

The use of artificial intelligence (AI) techniques in traffic signal control has been the subject of study and investigation over the last two decades. AI programs can be used in a variety of problem areas; there exists generic problems where AI systems have already demonstrated a particular degree of success in finding a solution. Some of these generic problem fields include the following areas (Bielli et al. 1994):

1. Interpretation and analysis: interpretation is an analytical problem where conditions and descriptions are given as part of the solution description; the task is to complete the solution description by applying the available knowledge so that the available data and conditions are consistent with the solution.
2. Diagnosis: in this area, the causes of some recognized condition are determined. This task implies relating symptoms to their possible causes in order to find reasons for faults in the system.

3. Monitoring: monitoring means the recognition of situations by continuously observing incoming data. It implies data analysis, data interpretation, and diagnosis.
4. Prediction: prediction is forecasting of likely consequences and future developments given information about the present and past state of the system.
5. Planning: in the planning task, the creation of a plan is done in term of a sequence of steps to achieve some predefined goal.
6. Design: design can be viewed as the generation of specifications for creating concrete or abstract objects which meet some desired requirements.
7. Control: control involves monitoring system behavior in order to reach some known objective. Control problems imply a combination of tasks and capabilities of monitoring, interpretation, diagnosis, predictions, and planning.
8. Advising: advising provides decision support when looking at selection problems. Advising refers to both regulation advising and service advising.

When looking into the traffic control problem as a whole, we find that it consists of many of the generic problems identified in the list above. In traffic control, many systems, such as variable message signs, ramp metering, and demand management, need to be integrated with traffic signal control to work together to reach the goal of efficient traffic control. The integration of AI-based functionalities in a traffic management system needs to be done broadly and applied to each of these systems that offer a promising application base for AI techniques. In the following section, we introduce a prototype that uses an AI technique in the problem of adaptive signal control.

2.4.1. Generically Adaptive Signal Control Algorithm Prototype (GASCAP)

GASCAP (Owen and Stallard 1999) is a rule-based approach for real-time distributed adaptive signal control. The prototype was tested using the traffic simulation program CORSIM.

2.4.1.1. The GASCAP methodology

The GASCAP methodology is based on three elements.

2.4.1.1.1. Queue estimation

There are three basic responsibilities for the queue-estimation algorithm:

1. Predict the number of vehicles in queues, and the content of vehicles for each lane. (The content is the total number of vehicles approaching the intersection.)
2. Analyze the gaps of opposing vehicles for permitted left turners. (A permitted left turn is made across an opposing through vehicle flow; the driver is permitted to cross the opposing flow but must select an appropriate gap in the opposing stream through which to turn. (McShane et al. 1998))
3. Estimate the turning percentage for each approach.

GASCAP records the number of activations in upstream detectors to estimate the queue. The queue estimation algorithm is called every second to determine if any of the vehicles that were created from the detector activations will arrive in queue. For the permitted lift turn prediction, stop bar detection of the opposing approach is analyzed; if a gap between detector activations is long enough for the vehicle making the left turn to complete its movement, that vehicle is no longer counted as in queue for that approach.

2.4.1.1.2. GASCAP rules for un-congested intersections

The queue and content estimates that have been computed for each lane are translated into queue and content data for a particular movement. For each intersection, there is a group of allowed movements (left turn, right turn, and through movements). The number of vehicles in queue and content requesting a certain movement are computed every second, and a set of rules uses this information to determine the signal state at the intersection. These rules that GASCAP uses are

1. Demand rules: demand rules correspond to control of isolated intersections. Demand is determined by checking if the queue for the current movement is less than a constant threshold. Movements with the highest queues are given the right of way.
2. Progression rules: progression rules are used for coordination of the green time at adjacent intersections. Whenever an upstream intersection changes to a phase that will supply vehicles to a downstream intersection, the downstream intersection will schedule phases to accommodate the incoming vehicles.
3. Urgency rules: urgency rules are used when traffic demand increases. Whenever an upstream detector at any intersection has been continuously occupied for 15 seconds, a phase that serves that approach is submitted.
4. Cooperative rules: cooperative rules are used when conditions start to move from saturated to congested. If the approach between two intersections experiences spillback, then the upstream intersection will not select a movement that contributes to the spillback. Spillback happens when the queue of stopped vehicles at an intersection extends into another intersection and blocks it.

5. Safety rules: the safety rules are used in congested conditions. The safety rules prevent the GASCAP prototype from selecting an unsafe signal state, like green indication for conflicting movements, and violation of clearance intervals and minimum green times.

2.4.1.1.3. GASCAP for congested intersections

For intersections that are experiencing congestion, GASCAP uses information from the upstream detectors to construct a fixed-time signal plan. From the activations and deactivations at the upstream detectors, GASCAP computes occupancy for the approaches over a 10-15 minute period; volumes are computed from occupancy information. GASCAP creates a timing plan for the congested intersection based on these volumes.

2.4.1.2. Simulation results

GASCAP was tested using simulation on three different arterial networks. GASCAP performance was evaluated against the RHODES, OPAC systems, and baseline (the existing traffic control strategy on these networks). The networks were chosen with different geometric properties and different levels of saturation. Simulation results showed that GASCAP effectively reduced the delay and increased the throughput.

CHAPTER 3. METHODOLOGY

This chapter provides a definition of the problem addressed in this research and the approach used for solving this problem.

3.1. Problem Definition

This paper addresses the problem of traffic signal control. In the previous chapter, different signal control techniques were discussed. While different signal control methods are suitable for different situations, adaptive signal control was shown to be the best suited to deal with dynamic traffic conditions, conditions where changes in traffic patterns occur suddenly. In this paper, a new methodology for achieving adaptive signal control is proposed. The suggested technique offers a level of simplicity that the existing adaptive signal control methods lack while maintaining benefits that can be expected from an adaptive signal control system. The proposed system deals with large traffic fluctuations rather than small changes in traffic. The system will wait until a certain threshold of traffic increase has been reached and then apply a signal plan that is suitable for controlling the new traffic level.

In order to be able to perform efficient traffic signal control, the signal plans in the traffic signal control systems need to be modified and updated constantly. A signal plan reflects the traffic control strategy that is suitable for controlling a certain traffic condition. Traffic conditions change in terms of volumes and turning movement percentages. (Turning movements can be defined as the percentage of vehicles that make a right or a left turn at an approach of an intersection to the total number of vehicles traveling on that approach.) Based on the traffic conditions, the signal plan determines how the green time in the cycle is divided between the different approaches. If these traffic conditions change, resulting in different volumes and turning percentages, the signal plan becomes inappropriate for control of that traffic condition which may, in turn, result in increased delay and congestion.

The current approach for developing signal plans for traffic signal control works as follows:

1. Traffic data collection: traffic volumes and turning percentages are collected for the intersection or the network that needs to be optimized. The data collection is done by performing traffic counts at the intersections. The traffic counts are carried out three times a day during the AM, PM, and midday peak periods. Counts for each peak period are usually performed for an interval of two hours. Peak hour factors are then calculated. The peak hour factor is defined as the relationship between hourly volume and the maximum rate of flow, in vehicles per hour, within the hour.

   $$PHF = \frac{\text{hourly volume}}{\text{maximum rate of flow}}$$

   The PHF is calculated because the signal plans need to be designed to accommodate the maximum rate of flow within the peak period.

2. Signal plan optimization: the basis for developing signal plans is laid out in the Highway Capacity Manual (HCM; TRB 1998), which is developed and published by the Transportation Research Board. It contains formulas for calculating signal plan parameters such as cycle length and splits. Several software programs were developed largely based on the HCM methods to

facilitate signal timing. These optimization programs take the traffic data as an input and provide the optimized signal plan; the resulting signal plan is suitable to control traffic with the observed conditions at the time the data collection was performed.

3. Signal plan deployment: the signal plans that were suggested by the optimization program are then applied in the signal controllers at the intersections. Deployment of signal plans requires a technician to enter the plans into the controllers located at each of the optimized intersections. Some controllers offer the capability of accepting a connection from a remote computer through telephone lines, or other means of communication, and allow the change of the signal plans. Usually, after a new signal plan is deployed, the intersections or the network are observed to ensure that the plan is working properly. Today, with the availability of sophisticated traffic simulation programs, the signal plans can be tested using simulation to analyze their effectiveness before they are deployed.

The points above represent the common state-of-practice in traffic signal control; throughout the paper, this approach is referred to as traditional traffic signal control. There are two major deficiencies in this approach. The first deficiency is that updating signal plans manually is a costly and a time-consuming process; this often results in signal plans being left for years without being updated, especially in areas with limited financial and personnel resources. The second deficiency lies in the fact that the plans are developed to perform traffic control for a certain traffic condition that was observed when traffic data were collected. The traffic signal control system will use these signal plans regardless of the real traffic conditions. Adaptive signal control offers a solution for both of these problems. The following section introduces the method this paper proposes for achieving adaptive signal control.

3.2. Approach

Adaptive signal control systems consist of three major components. These components are

1. Detection system: detection is necessary to recognize changes in traffic patterns. Detector readings are turned into traffic volume and turning movement information which, in turn, will be used to determine a suitable signal plan to control the current traffic conditions on the intersection. Detector data can also be used to predict traffic changes before they occur. The prediction of traffic changes is done through performing detection upstream of the intersection to try to identify the vehicles' arrival patterns at the intersection.
2. Solution system: this component provides an optimized signal plan for the observed, or predicted, traffic conditions at the intersection.
3. Communication system: communication between the intersections of the network is necessary to share the network optimization parameters; intersections on a coordinated arterial need to operate on the same cycle length. The offset is another parameter that needs to be shared for the intersections of an arterial to be coordinated.

In this paper, only the solution component of traffic adaptive signal control is addressed. Volumes and turning movement information are assumed to be available from the detection component. Since this research is a test in a simulation environment, communication between the intersections of the network is done through the simulator.

3.2.1. Signal Plan Selection

The adaptive signal control system proposed in this paper offers a simplified approach for providing optimized signal plans that are suitable to be used for controlling different traffic conditions and dealing with traffic changes. The approach is to have a set of previously developed signal plans that are suitable for an array of traffic conditions. Please note that micro changes in traffic levels have little impact on the signal plans. When a change in traffic occurs, the system selects a plan that is appropriate to control the new traffic conditions from the plan library.

For such a system to be successful, the signal plans that will make up the library of plans available for the system need to be comprehensive to deal with different traffic changes. In other words, the probable traffic changes need to be identified before the system is deployed. Once the scenarios of traffic changes are identified, signal plans can be developed for these conditions. The signal plans can be developed using a signal plan optimization program. The optimization program that was chosen in this research is SYNCHRO (Trafficware 2000). SYNCHRO is a commercial software package developed by Trafficware which provides single intersection signal plan optimization as well as network coordination optimization.

3.2.1.1. Possible traffic changes

Traffic conditions change when traffic volumes on the network change. The volumes could increase or decrease on different intersections throughout the network. These traffic changes can occur on a certain approach or movement of an intersection. The traffic changes that the system will need to accommodate with suitable signal plans depend on the network. For testing the system, a case study was selected, and traffic changes were identified in accordance with this case study. A description of the case study will follow in the next chapter.

The possible fluctuations in traffic were identified as discrete increments in traffic volume on the intersections of the case study. The increments were chosen to cover the different possible saturation and congestion conditions. The chosen volume increment levels are 25%, 50%, 75%, 100%, and 200% increase in traffic volume from the original base case; the base case was considered to be the existing traffic levels on the case study network at the time the data were collected. The decision of choosing these particular volume increment percentages was made based on support from three different sources; these sources are

1. The signal plan optimizer SYNCHRO: the network was built in SYNCHRO with the base case volumes; the volumes were varied with 10% increments at each designated approach of the intersection. For each volume increment, the signal plan was re-optimized for the volume change. The changes between the optimized signal plan for the base case and the optimized signal plan for the incremented traffic case were noted. It was found that SYNCHRO starts providing significantly modified signal plans when changes in traffic volume

are made at the 25% increment level. Another parameter that was monitored throughout these experiments was the performance of the signal plan. SYNCHRO provides several performance measures for signal plans. Intersection delay was monitored while the traffic volume was increased. Using the intersection delay as a measure of the signal plan performance, it was determined that updating the signal plan at 25% volume increments is a suitable resolution.

2. The traffic simulator VISSIM: the data of the traffic volumes and signal plans from SYNCHRO were entered into the case study network in the simulator VISSIM. Traffic simulations were carried out with varying traffic conditions. VISSIM visual output was monitored for queuing of vehicles on the different intersections of the case study network. VISSIM also provides numerical data on the network performance and the performance of each intersection in the network. The results showed the selection of the 25% volume increment value to be suitable. The 200% increment level was used for the system to be able to deal with highly congested situations. At 200% increase of original traffic, the road capacity is used to its limits, and problems such as spillback start to occur. Anything beyond 200% would cause a total breakdown of the system and lead to a gridlock situation.

3. This increment level selection along with the SYNCHRO performance measures, and the VISSIM simulation results were shared and discussed with traffic engineers at the Advanced Traffic Analysis Center, North Dakota State

University, Fargo, North Dakota. The selected increments were considered to be sufficient to cover the different possible traffic variations.

3.2.2. System Design

For the implementation of the solution component of the proposed adaptive signal control system, an expert system was used. The expert system knowledge base contains the signal plan library as well as threshold information to determine when to change the signal plan according to changes in traffic. One particular class of generic task expert systems was used, hierarchal classification. Hierarchal classification systems are described in Chapter 2, Section 2.2.3, along with an explanation of why a hierarchal classification system is suitable for the traffic control problem.

3.2.2.1. Details of the system

In hierarchal classification, the system consists of a hierarchy of interacting specialists, each carrying out a specific task. In the software package that is used for building the expert system (Generic Task Toolset developed at Intelligent Systems Laboratory, Michigan State University), a database can be constructed. The database contains a set of data types. Variables can be defined within these data types and given legal values depending on the type of the variable. Variable definition also contains a comment or description of the variable, along with a question to be asked to the user of the expert system for the input process to get a value for that variable. The data types that were used in this system are

1. One Of variable: the One Of variable type allows for the definition of a variable that can have a number of legal values. This variable type was used in the definition of the Intersection, Approach, and Movement variables. In the Intersection variable, the legal values are the intersections of the case study network. For the case of the Approach variable, the legal values are the approaches of the intersection: northbound, southbound, eastbound, and westbound. Finally, for the Movement variable, the legal values are the legal movements at that approach of the intersection: left turn, through movement, and right turn.
2. Numerical variable: the Numerical variable type allows for the definition of a variable that has a numerical value. This variable type was used in the definition of the Volume variable. The volume is the number of the vehicles per hour that travel on a certain movement.

The system is built as a tree consisting of a number of nodes; each node represents a specialist. The system has an option to edit the “top table” of a node. In the top table, it can be specified how the node will deal with the input data. The specialist could carry on a comparison process and decide whether to pass the control to a specialist further down in the hierarchy tree. The specialist can also give the result in the form of advice if certain variables match.

The specialists form a tree structure where input data go in at the highest level of the tree and the final output is produced at the leaf level of the tree. In the traffic control problem, the input is traffic volume at a certain movement, approach, and intersection of the network. The output is an optimized signal plan for that level of traffic volume. Since the proposed adaptive signal control system is traffic network specific, a case study was needed to implement the system for testing and analysis purposes. As a case study, a network of five intersections, which is a sub network of the 25^{th} St. South corridor in Fargo, North Dakota, was used. The case study is described in detail in the following chapter.

The structure of the hierarchy of the system is as follows:

1. At the first level of the hierarchy tree, there is a number of Intersection specialists, one for every intersection of the traffic network. This specialist simply determines whether the volume input data belong to its intersection or not. If the volume is for the specialist's corresponding intersection, the specialist returns a match and passes the control down the hierarchy structure.
2. At the second level of the hierarchy tree, there is a group of Approach specialists; depending on the specific geometry of the intersection, there exists a number of approaches where traffic travels into and out of the intersection. The typical four approaches of an intersection are northbound, southbound, eastbound, and westbound. This specialist determines if the volume input data belong to its corresponding approach. If the volume belongs to this specific approach, the specialist returns a match and passes the control down the tree.
3. At the third level of the hierarchy tree, there is a set of Movement specialists. Again, depending on the geometry of the intersection, there exists a number of movements the traffic can travel into at each approach. Typically, there are

three movement types: a left-turn movement, a through movement, and a right-turn movement. The movement specialist determines if the volume input belongs to its specific movement. If it does belong, the specialist will return a match and pass control to the next level of the hierarchy.

4. At the leaf level of the hierarchy tree, there is an array of Increment Level specialists. Depending on the value of the volume input, these specialists decide on the increment level that fits this volume. It is at this level of the hierarchy where the system returns the advice or the optimized signal plan that is appropriate for the input volume level.

This tree structure makes the system flexible and allows adding new nodes at any of the hierarchy levels easily. Adding new nodes or specialists permits the expansion of the system to deal with new intersections, or adding more optimized plans for new approaches and movements that were not initially included in the system as the need for that arises.

The hierarchy tree of specialists is shown in Figure 1. There it can be seen how the specialists interact and how the control is passed down from one level of the hierarchy to the next. We start with the intersection; move on to the different approaches of that intersection; and move on to the legal movements of that approach, or the movements for which we are interested in optimizing the signal plan. Finally, the control is passed to the last level of the hierarchy where the volume input is analyzed to decide the volume increment level to which it belongs, and based on that increment level, an optimized signal plan is returned.

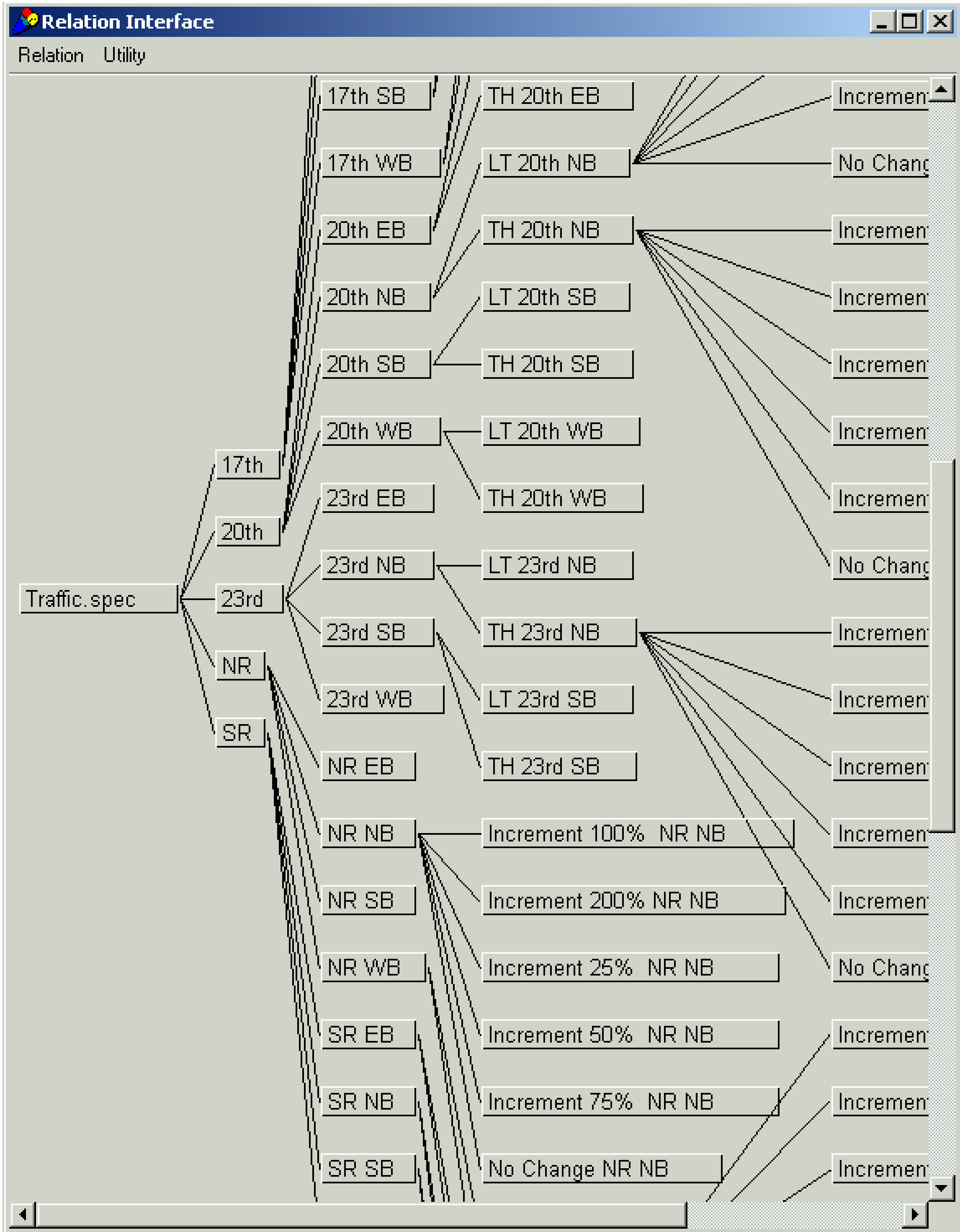

Figure 1. A portion of the hierarchy tree of the system.

The input process to the system is shown in Figure 2. In the input window, all the variables that the system uses are shown; a variable can be clicked, and the desired value of the variable can be entered. A sample of the input for the Intersection variable is shown in Figure 3. This window shows the variable type which is a One Of variable in the Intersection case, and it shows the list of choices for the value of that variable which are the list of the intersections in the network. If the button that says written help is clicked, another window that contains a description of the purpose of this input will appear. The help window is shown in Figure 4.

The output that the system provides is the optimized signal plan. The output consists of the optimized offset of the intersection in seconds and the optimized splits of the intersection. The splits are given in the green time for each of the approaches and movements of the intersection in seconds. The yellow times and red clearance times are left unchanged. Figure 5 shows a sample output of the system.

At this point, the system is used in this offline fashion, where input is done through windows, and the output is produced in windows. However, the input and output process can be modified to be done through files. The input can be done through a file containing intersection, approach, and movement identification data along with volume. This input will eventually be produced by the detection component of the system which turns detector readings into volume and turning movement data. The output can be directed to a file that contains the signal plan data. The file can be in a format that a traffic simulation program can utilize to change the signal plans of the simulation. In the future, the output can be made into an actual signal controller to change the signal plans at an intersection.

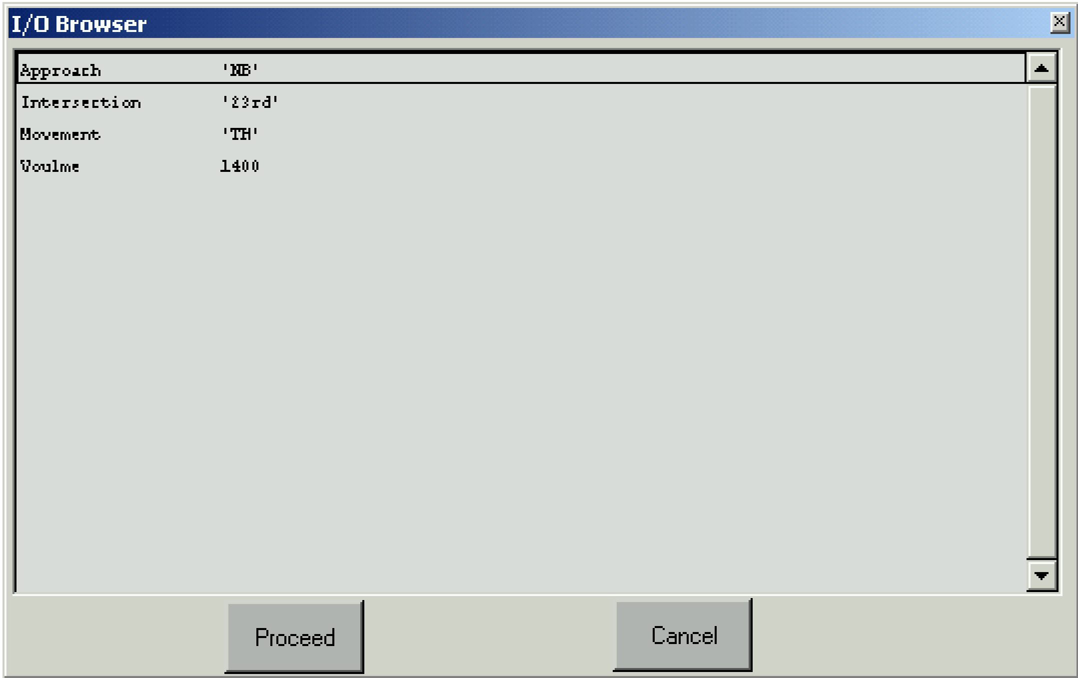

Figure 2. A view of the input window.

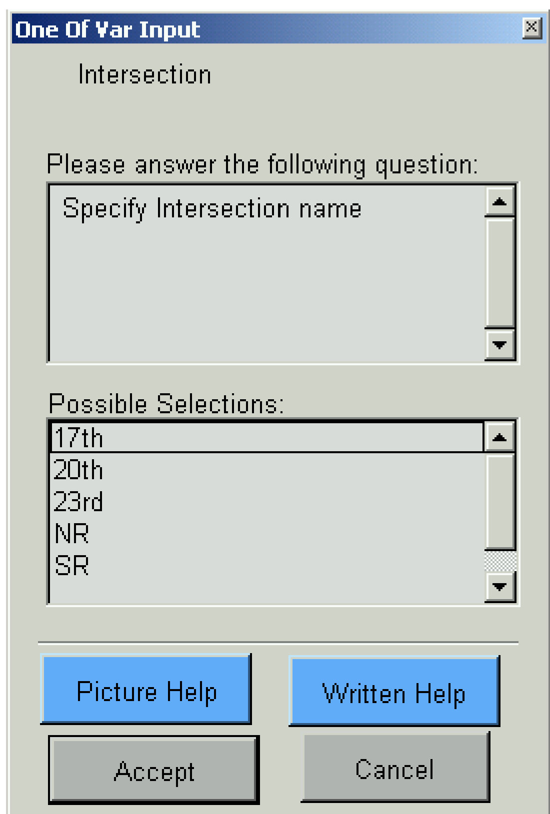

Figure 3. A view of the input for the Intersection variable.

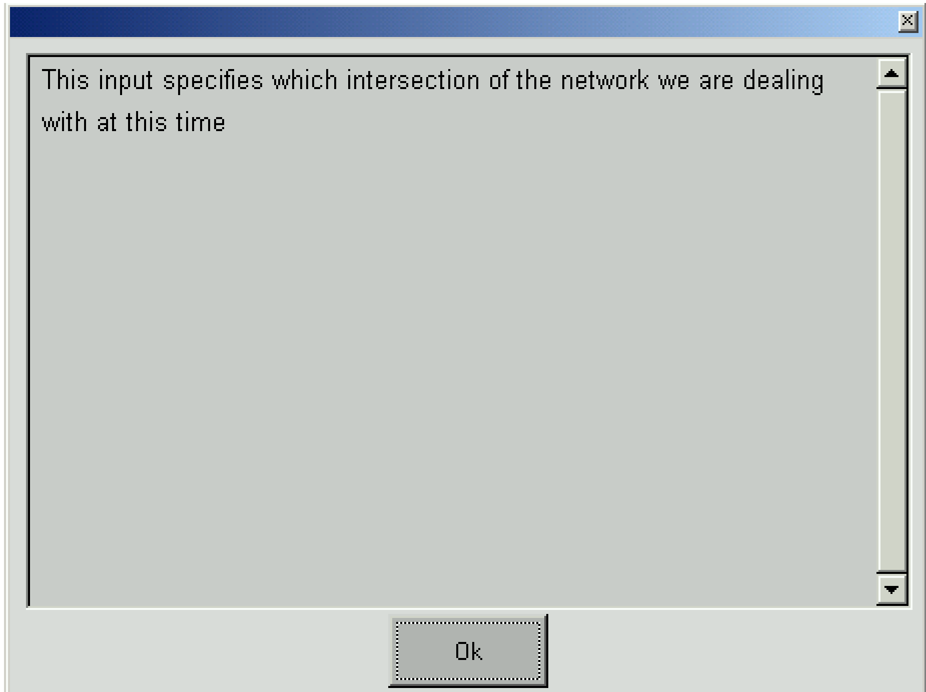

Figure 4. Written help window for the variable Intersection.

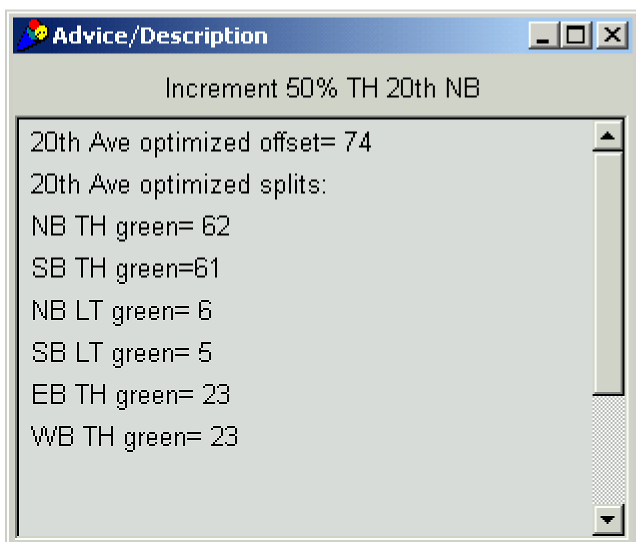

Figure 5. Sample output at the leaf level of the system.

At this level of the research, the objective is to test the performance and effectiveness of this simplified approach to adaptive signal control. All the testing was performed offline. Test data were entered into the system, and the optimized signal plans that the system suggested were collected. Afterwards, the test volume data along with the suggested signal plans were entered into the traffic simulation program, VISSIM, to test their performance.

Figure 6 suggests an architecture for online testing of the system. The system can be interfaced with a traffic simulator such as VISSIM. Detector data can be collected from a VISSIM simulation while traffic volumes can be varied during the simulation. Detector data will then be turned into volume and turning movement information in the detector interface, and fed into the knowledge-based system. The resulting signal plans from the knowledge-based system can then be formatted into signals plans that are compatible with VISSIM in the plans' interface. VISSIM can then change the signal plans in the simulation. This way, the performance of the system can be tested online since online testing is necessary if the system is to be applied with an actual traffic controller.

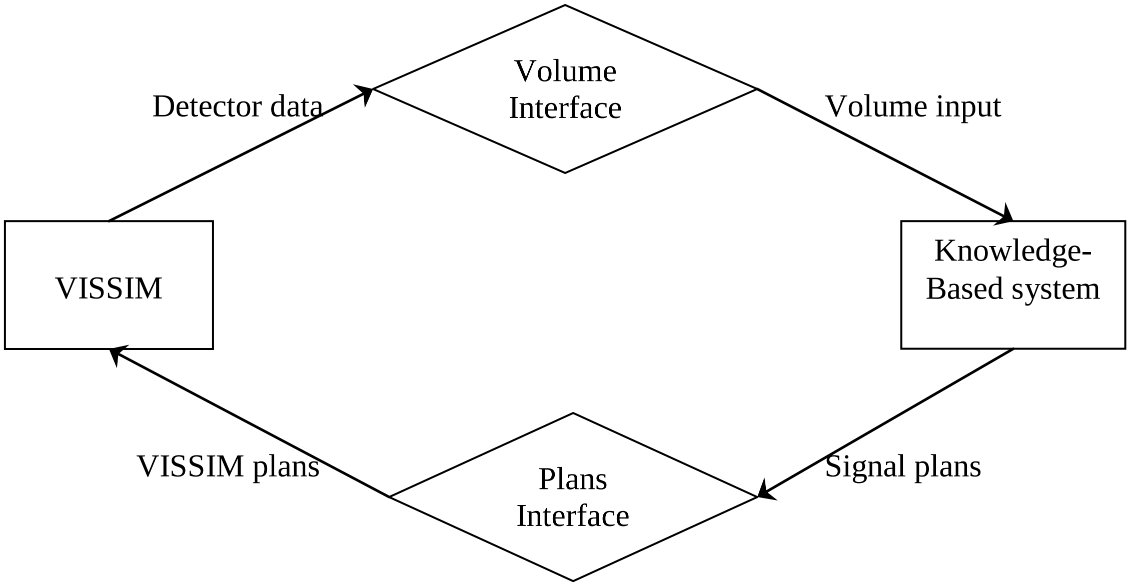

Figure 6. Suggested architecture for online system testing.

CHAPTER 4. CASE STUDY AND TESTING

This chapter provides a description of the case study arterial that was used for the implementation of the system along with details about the system performance testing.

4.1. Case Study

4.1.1. Description of the Case Study

To test the performance of the proposed system, a case study was used. The selected case study is a section of a main arterial network with five intersections. The intersections where chosen from the 25th St. South corridor in Fargo, North Dakota. This street is considered a major north south arterial in Fargo. It consists of two travel lanes in each direction and a center lane for left turns throughout the length of the arterial. The arterial experiences heavy traffic movement in the northbound direction during the AM peak period and heavy southbound traffic in the PM peak period. Several major east-west roads in Fargo cross the 25th St. arterial and contribute to its heavy traffic conditions.

The intersections that were selected for the case study offer different traffic and geometric conditions. The five selected intersections from north to south are

1. 25th St. and 17th Ave. South: Fargo's 17th Ave. is one of the major roads that intersect 25th St. It has moderate to heavy traffic conditions. The road consists of one travel lane, with right-turn and left-turn pockets in the eastbound approach, and a left-turn pocket in the westbound approach.
2. 25th St. and 20th Ave. South: Fargo's 20th Ave. is a road that serves the residential neighborhoods in that area; 20th Ave. consists of one travel lane, with left-turn

pockets for the eastbound and westbound approaches. The road has light to moderate traffic conditions.

3. 25th St. and I-94 North Ramp: this intersection is the northern part of the I-94 interchange with 25th St. It serves vehicles getting off I-94 from the westbound direction and going into the city network. The I-94 North Ramp consists of one travel lane in the westbound direction, with a right-turn pocket and a left-turn pocket. This intersection experiences heavy right-turn movement during the AM peak period.
4. 25th St. and I-94 South Ramp: this intersection is the southern part of the I-94 interchange with 25th St. It serves vehicles getting off I-94 from the eastbound direction and going into the city network. The I-94 South Ramp consists of one travel lane in the eastbound direction and a right-turn pocket.
5. 25th St. and 23rd Ave. South: Fargo's 23rd Ave. is a road that serves the residential neighborhoods in the area. It consists of one travel lane with a left-turn pocket in the eastbound approach; 23rd Ave. experiences light traffic conditions with moderate left-turn movement in the eastbound direction.

The network was built in the signal plan optimization program SYNCHRO (Figure 7) and in the traffic simulation program VISSIM (Figure 8). Traffic volumes and turning movement data were obtained from a study that was performed on the 25th St. corridor in the summer of 2000 by the Advanced Traffic Analysis Center, North Dakota State University. The purpose of the study was to monitor how the closing of one travel lane on.

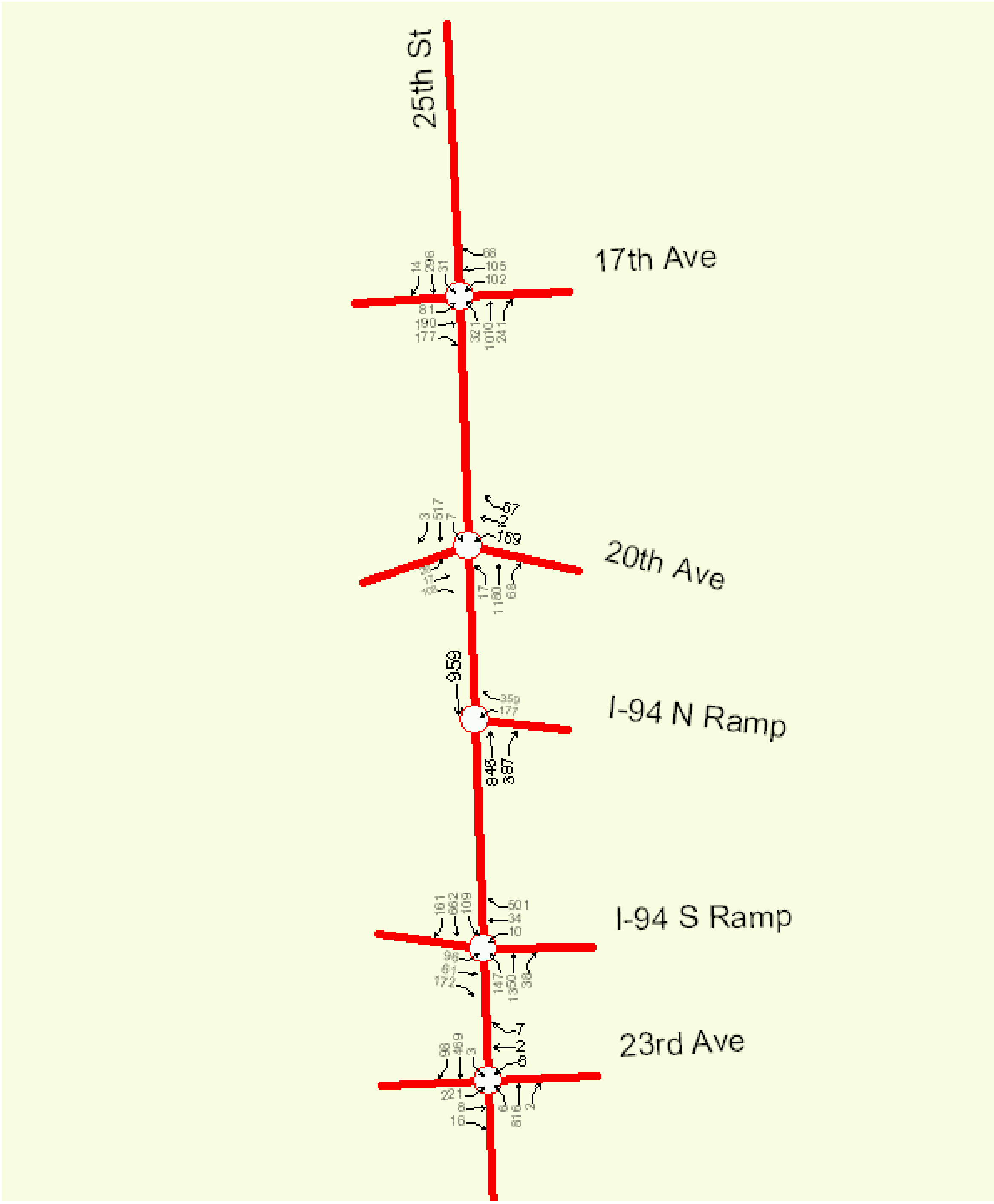

Figure 7. Case study network in the traffic optimizer SYNCHRO.

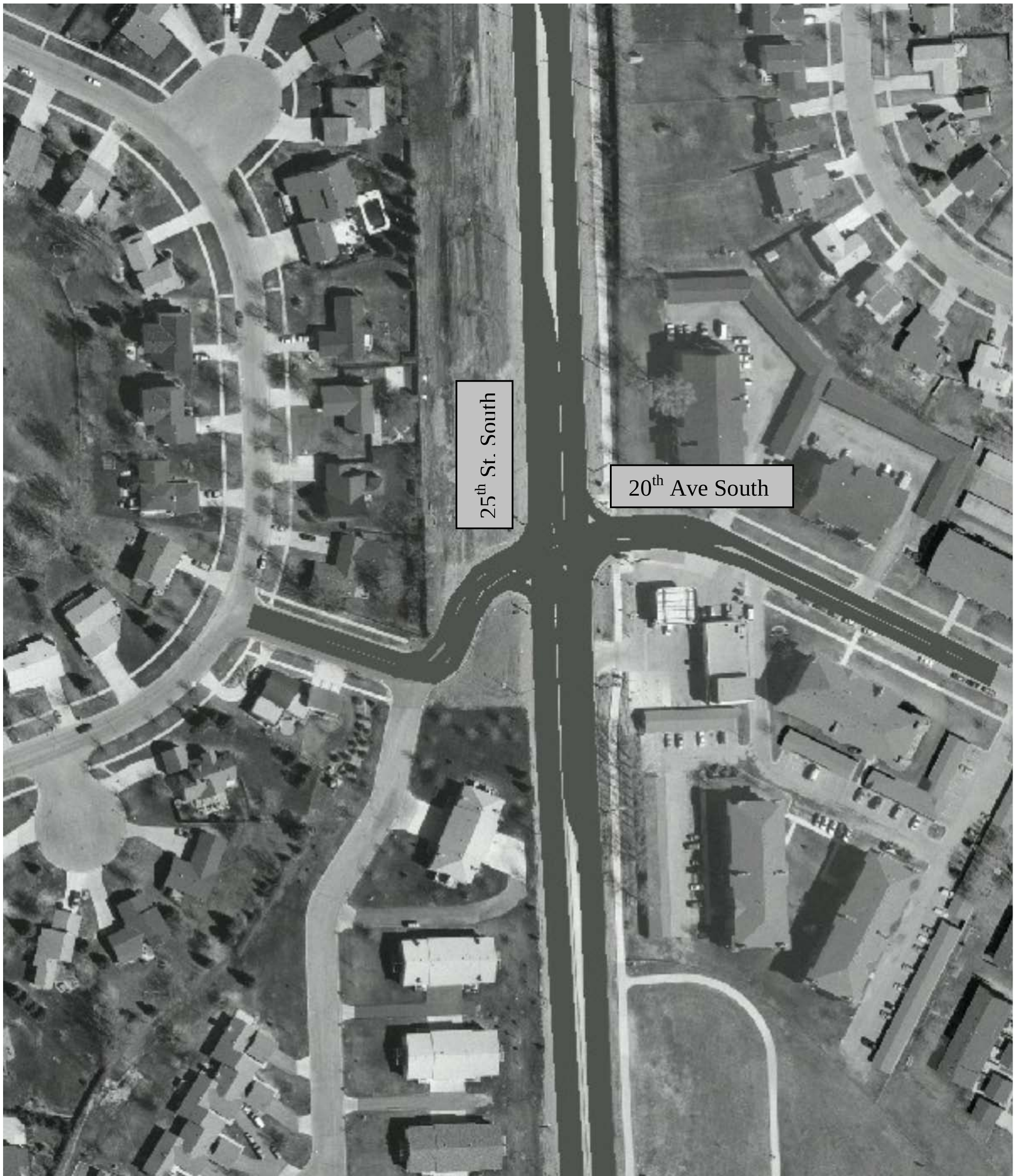

Figure 8. The intersection of 25^{th} St. and 20^{th} Ave. South from the case study network in the traffic simulator VISSIM. VISSIM allows the traffic network to be built over an aerial photograph.

Interstate Highway 29 (I-29, a major north-south corridor) would affect traffic conditions on other roads in the city

4.1.2. Reasons for Selecting This Case Study

The selected arterial and set of intersections offer different geometric conditions. The case study intersections include basic four-approach intersections, as well as the I-94 north and south ramps which have different geometric properties. The case study offers a variation in the number of travel lanes available for traffic, and in the use of right- and left-turn pockets. More importantly, the intersections offer different traffic conditions through varying traffic flows at different intersections. The selected case study contains intersections that can change the traffic flow over the entire arterial. The intersections of 17th Ave. South and the I-94 north and south ramps have moderate and heavy traffic flow which, if they changed for some reason, could impact the traffic conditions along the entire arterial.

During the 2000 study of I-29 construction project impact, 25th St. experienced a large increase in traffic due to its use as an alternate route. All intersections experienced traffic increases. At some intersections, the traffic volume doubled, creating very congested conditions and large delays.

The arterial of 25th St. offers dynamic traffic conditions that can be impacted by several major roads that cross it. Since 25th St. already experiences delays with current traffic conditions, any changes in traffic will impact the traffic plans severely, which leads to more delay and congestion. The signal plans along the arterial's intersections require frequent updates to keep up with the traffic changes. These factors make the 25th St. arterial a candidate for real-time adaptive signal control and a valid testing ground.

4.1.3. Description of Data Used

The data used in the 2000 study of the I-29 construction project impact were in the form of turning movement counts. Counts were carried out in 15-minute intervals on intersections throughout the arterial. AM peak counts were performed between 7:00 AM and 9:00 AM while PM peak period counts were performed between 4:00 PM and 6:00 PM. The peak hour factors (PHF), a measure of traffic intensity, were calculated for both peak periods. The PHF is calculated because the signal plans need to be designed to accommodate the maximum rate of flow within the peak period. Geometric data were also collected, such the number and width of travel lanes, and the use of right-turn and left-turn pockets.

For this case study, the AM peak period counts were chosen because they presented heavier traffic flows than the PM peak period. With heavier traffic flow, the true capabilities of the adaptive signal control system could be tested.

4.1.4. Description of the Traffic Scenarios

For the purpose of this research, several scenarios of how the traffic conditions can change on the five selected intersections of the case study were considered. The scenarios were chosen based on the traffic changes that could happen in the case of real-life events on intersections of the case study and the surrounding roads that could impact traffic in the case study area.

The two major scenarios that were chosen are as follows:

1. An increase in the northbound traffic along the five intersections of the case study: such an increase could result from a traffic surge from the south, similar to what happened with the I-29 construction project.
2. An increase in the right turn movement on 25th St. and the I-94 North Ramp intersection: this traffic increase can occur if there is an increase of westbound traffic on I-94 that gets off the interstate and goes into the city network. This increase could be caused from diverting traffic from I-94 due to an incident west of the interchange.

For these two scenarios, traffic changes in the form of volume increments in these particular approaches were considered. The traffic increment levels that were chosen are 25%, 50%, 75%, 100%, and 200% increases of traffic volume from the original base case. The base case was taken to be the volumes that were collected for the I-29 construction impact study. The decision of choosing these particular volume increment percentages was explained in the previous chapter, Section 3.2.1.

4.2. System Testing

For testing the system, it was decided to perform a comparison study using traffic simulation. Two cases were created. In the first case, referred to as the Existing Case, the simulation was performed with the original signal plans that were optimized for base case volumes. In the second case, referred to as the Optimized Case, the simulation runs were performed with the intersections controlled by signal plans that the knowledge-based system proposed for new traffic levels.

The traffic simulator that was used for the testing is VISSIM. VISSIM is a microscopic simulation model that provides a high level of detail in the representation of the traffic network. VISSIM addresses all aspects of the interaction between vehicles in the network and the control of traffic on intersections. Microscopic simulation models are typically stochastic in nature; therefore, factors such as vehicle position, acceleration, speed, lane changing maneuvers, number of stops, etc., are assigned based on a probability function (PTV 2000). The stochastic characteristics of microscopic simulation models and those of VISSIM, in particular, offer a more realistic representation of real-life traffic conditions. Performing simulation studies is a common practice when testing new traffic signal control strategies, especially for determining the performance and the effectiveness of traffic signal plans.

4.2.1. Test Case Details

To build the two cases, existing and optimized, that will be the basis for the comparison study, sample traffic volumes were needed. It was decided to test the performance of both cases on a set of 10 different scenarios of traffic volume levels. These volumes were selected by generating random numbers which were used to suggest increments of northbound traffic on the intersections of the case study network. The random numbers were generated in a way that allows the traffic volumes to vary between the base case level volumes and the 200% increment level volumes.

Twenty VISSIM input files were created, 10 for the existing case and 10 for the optimized case. Each of these ten files was built with traffic volumes taken from one of the increment scenarios for testing. In the existing case files, the signal plans were the same for each of the files while, in the optimized case files, a signal plan that was suggested by the knowledge-based system was used for each file.

Before simulation runs could be started, one issue had to be addressed. Since VISSIM is a stochastic model, it uses a seed number that initializes the random number generators internal to the simulator. The seed number can be changed by the user. Simulation runs with identical input files and seed numbers generate identical traffic conditions. However, using a different seed number results in a stochastic variation of traffic flows; therefore, traffic volumes and the results of the simulation can change (PTV 2000). In order to normalize this randomness, multiple simulation runs with different seed numbers are needed. Each VISSIM input file from both cases was run with 10 different random seed numbers. The results of the simulation were averaged for the 10 runs.

VISSIM output provides several measures for evaluating the performance of the traffic network; the parameters that were chosen for this case study are the travel time and total delay of the network. These measures are normally used for assessing traffic performance. VISSIM allows the user to define sections of the network where performance data are to be collected. For this case study, these sections were defined along 25th St. on both the northbound and southbound directions. The performance data collection sections covered the entire length of 25th St. through the case study. Data were also collected for the right- and left-turn movements along 25th St. when right-turn pockets existed. The data collection sections are defined with an entry point and an exit point. For the travel time measure, VISSIM calculates the time it takes vehicles to travel from the entry point to the exit point of each previously defined section.

Delay in VISSIM is defined as the time when a vehicle is traveling at a speed below the desired speed distribution. Speed distributions take the speed limit on the road, as well as the acceleration and deceleration process of stopping and going, into account. The delay measure in VISSIM also provides two other parameters: stopped delay, which is defined as the time the vehicle spends at the speed of zero, and number of stops, which is a count of the number of times a vehicle reaches the speed of zero. Stopped delay is included in the total delay measure since a speed of zero is less than the desired speed distribution.

CHAPTER 5. RESULTS, CONCLUSION, AND FUTURE WORK

This chapter will provide the results of the simulation testing of the system and offer conclusions and future work.

5.1. Results

After the simulation runs were completed, output data were extracted from the simulation output files. The travel time and delay measures were compiled from the different data collection sections. The collected parameters were formed into travel time and delay measures for the northbound and southbound approaches of 25th St.
For the measure of travel time, VISSIM returns the average travel time for the vehicles that pass through a designated data collection section. Summing up the values for all the sections on each approach, it returns the average travel time for all vehicles along the northbound and southbound approaches of 25th St. throughout the length of the case study arterial.

Table 1 shows results of the travel time comparison between the northbound (NB) and southbound (SB) approaches of 25th St. for the existing and optimized cases. The "Scenario #" column shows the number of the test scenario; the scenarios are arranged in ascending order based on the level of traffic volume increment. The "Percentage difference" column shows the percentage of savings in travel time that the optimized case had over the existing case. When the percentage value is negative, it shows that the existing case travel time was less than the travel time for the optimized case. Only one negative value occurred in the comparison, and it was in scenario # 4 for the southbound direction of the arterial. This negative value can be explained by the nature of the case study. Since traffic increments were enforced on the northbound direction, northbound movements were assigned more green time in the signal plan splits. Since the cycle length is constant, the additional green time was taken from the other approaches of the intersection. In addition, the original base case AM signal plan favors the northbound approach since it experiences heavier traffic. In the following scenarios when larger volume increments were used, the system achieved savings for both the northbound and southbound directions of 25th St. These savings are a result of the elimination of the problem of spillback in the optimized case. When spillback occurs in the northbound direction of the arterial due to large volume increments, the opposing left-turn movement of the southbound direction gets hindered, which leads to large delays on both northbound and southbound approaches.

Table 1. Total average travel time per approach in minutes.

Scenario #	**Total average vehicle travel time in minutes**					
	Existing Case		Optimized Case		Percentage difference	
	NB	SB	NB	SB	NB	SB
1	5.90	6.66	5.82	6.57	1%	1%
2	5.94	6.77	5.89	6.75	1%	0%
3	6.51	6.86	6.06	6.87	7%	0%
4	6.08	6.78	6.02	6.85	1%	-1%
5	6.88	7.29	6.13	7.22	11%	1%
6	6.96	8.11	6.35	7.60	9%	6%
7	7.26	7.95	6.35	7.48	12%	6%
8	8.72	8.47	7.57	8.44	13%	0%
9	10.14	8.99	9.23	8.23	9%	8%
10	10.99	9.82	9.59	7.99	13%	19%

For vehicle delay, the value that VISSIM calculates is the average total delay per vehicle (in seconds). By multiplying that value by the number of vehicles that were generated in that particular section, the total average vehicle delay for the section can be found. When total vehicle delay values are summed up for all the sections in both northbound and southbound approaches, the total vehicle delay for the entire corridor can be found. Total average vehicle delay is shown in Table 2.

Table 2. Total average vehicle delay per approach in minutes.

Scenario #	**Total average vehicle delay in minutes**					
	Existing		Optimized		Percentage difference	
	NB	SB	NB	SB	NB	SB
1	998.42	242.05	967.12	240.48	3%	1%
2	1128.88	246.14	1101.05	256.82	2%	-4%
3	1499.60	234.79	1257.73	261.77	16%	-11%
4	1339.93	246.59	1296.25	264.45	3%	-7%
5	2309.75	268.68	1770.15	287.90	23%	-7%
6	2370.38	348.65	1918.12	303.87	19%	13%
7	2750.03	298.95	2089.13	298.73	24%	0%
8	4862.18	373.31	3757.47	369.28	23%	1%
9	6306.33	391.28	5601.35	353.98	11%	10%
10	7225.42	459.06	6146.33	337.91	15%	26%

Table 2 shows the results of total vehicle delay comparison between the northbound (NB) and southbound (SB) approaches of 25^{th} St. for the existing and optimized cases. Similar to Table 1, the "Percentage difference" column shows the percentage of reduction in delay between the optimized case and the existing case. When the percentage value is negative, it shows that the existing case delay was less than the delay for the optimized case. The explanation of the negative values is in line with the explanation that was provided for Table 1. Overall, the reduction of the delay that was achieved on both the northbound and southbound directions of 25th St. surpasses the three scenarios where delay was increased on the southbound direction.

5.2. Conclusion and Future Work

In conclusion, this research suggested a simplified methodology to achieve better traffic signal control through the use of a knowledge-based system that suggests an appropriate optimized signal plan to deal with large traffic changes. The knowledge-based system serves as a simplified adaptive signal control system. The results of the simulation testing show that this knowledge-based application is an efficient method for reducing travel time and delay on the arterial of the case study. The savings in travel time and delay are in line with the savings achieved by existing sophisticated adaptive signal control systems.

Future work can involve the following three areas:

1. Expand testing of the system to a larger, more realistic arterial and to a grid network.
2. Investigate possible integration with a traffic simulator as described in Chapter 3. Further testing can be done through integrating an actual signal controller with the knowledge-based system and the simulator. The technique of using a traffic controller with a simulation program is called hardware-in-the-loop simulation; the simulator provides traffic and detector reading while the controllers manage the signal at the intersection.
3. Finally, if the system proves its efficiency in further testing, there is a potential for real-world application. One candidate location for real-world testing is the

Fargodome, in Fargo, North Dakota, where different levels of heavy traffic conditions are created as a result of different events held at the dome.

BIBLIOGRAPHY

Andrews, C. (1997). RT-TRACS: Evaluation of Optimized Policies for Adaptive Control (OPAC). PB Network. XI: 32-33.

Bielli M., G. Amberosing, and M. Boero (1994). Artificial Intelligence Applications to Traffic Engineering, Coronet Books. Utrecht, The Netherlands.

Bretherton, D. (July 1999). The "SCOOT" Urban Traffic Signal Control System. Traffic Advisory Leaflet. The Department of the Environment, Transport and the Regions. London, The United Kingdom.

Chandrasekaran, B. (1985). Expert Systems: Matching Techniques to Tasks. Columbus, Ohio, The Ohio State University Department of Computer and Information Science Laboratory for Artificial Intelligence Research: 41-63.

Head, L., and P. Mirchandani (1998). The RHODES Prototype: A Description and Some Results. Transportation Research Board Annual Meeting. Washington, DC.

Head, L., and P. Mirchandani (1998). The RHODES Prototype: A Technical Review. Transportation Research Board Annual Meeting. Mid Year Meeting and Adaptive Traffic Signal Control Workshop. Washington, DC.

Lucas, D., E. David, P. Mirchandani, and K. Head (2000). Using Remote Simulation to Evaluate Real-Time Traffic Control Strategies. Transportation Research Board Annual Meeting. Washington, DC.

Luger, George F., and William A. Stubblefield (1998). Artificial Intelligence Structures and Strategies for Complex Problem Solving, Addison-Wesley. Reading, MA.

Martin, P. (2001). SCATS, an Overview. Transportation Research Board, Traffic Signal Committee. Washington, DC.

McShane, William R., Elena S. Parassas, and Roger P. Roess (1998). Traffic Engineering, Prentice Hall. Englewood Cliffs, NJ.

Owen, Larry E., and Charlie M. Stallard (1998). A Rule Based Approach to Real-Time Distributed Adaptive Signal Control. 78th Transportation Research Board Annual Meeting. Washington, DC.

Park, B., N. M. Rouphail, and J. Sacks (2001). Assessment of a Stochastic Signal Optimization Method Using Microsimulation. Transportation Research Board Annual Meeting. Washington, DC.

Piotrowic, G. (2001). SCATS Operational Experience at the Road Commission for Oakland County. Transportation Research Board, Traffic Signal Committee. Washington, DC.

Pooran, F. (2001a). Optimized Policies for Adaptive Signal Control (OPAC) Equipment Requirements. Transportation Research Board, Signal Systems Committee. Washington, DC.

Pooran, F. (2001b). Optimized Policies for Adaptive Control (OPAC) Principle Theories. Transportation Research Board, Traffic Signal Committee. Washington, DC.

PTV Planung Transport Verkehr. (2000). VISSIM 3.50 User Manual. PTV. Karlsruhe, Germany.

SCOOT. (Accessed in October 2000). SCOOT Glossary, http://www.scoot-utc.com/frame3/glossary.htm.

Trafficware (2000). SYNCHRO Users Manual. Trafficware. Berkeley, CA.

Transportation Research Board. (1998). Highway Capacity Manual. Transportation Research Board National Research Council. Washington, DC.

UC-Berkeley. (Accessed in October 2000). General Intelligent Transportation Systems Terms, Learn-net, http://learn-net.its.berkeley.edu/glossary.