QUEUING MODELS AND ANALYSES OF TRAFFIC CONTROL

By

Ning Zou

A Dissertation Submitted to the Faculty of the

DEPARTMENT OF SYSTEMS AND INDUSTRIAL ENGINEERING

In Partial Fulfillment of the Requirements
For the Degree of

DOCTOR OF PHILOSOPHY
In the Graduate College

THE UNIVERSITY OF ARIZONA

2007

THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Dissertation Committee, we certify that we have read the dissertation prepared by Ning Zou
entitled Queuing Models and Analyses of Traffic Control
and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy

_____ Date: April 20, 2007
Pitu B. Mirchandani

_____ Date: April 20, 2007
K. Larry Head

_____ Date: April 20, 2007
Wei H. Lin

_____ Date: April 20, 2007
Srinivasan Ramasubramanian

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copies of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

_____ Date: April 20, 2007
Dissertation Director: Pitu B. Mirchandani

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at the University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgment of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: _____

Ning Zou

TABLE OF CONTENTS

TABLE OF CONTENTS — Continued

TABLE OF CONTENTS — Continued

LIST OF ILLUSTRATIONS

LIST OF ILLUSTRATIONS — Continued

LIST OF TABLES

ABSTRACT

This dissertation investigates queuing modeling and related analyses of traffic control at signalized intersections and ramp meters. A primary goal in this dissertation is to apply queuing theory to vehicular traffic control systems, including adaptive traffic signal control, fixed-time traffic signal control, actuated traffic signal control and ramp metering control.

First, a simple two-phase traffic adaptive control scheme is studied, where the traffic signal serves one queue until it dissipates and then serves the accumulated queue in the other direction. Distribution of cycle length and average delay at steady-state is obtained by an iterative algorithm.

Second, a two-phase fixed-time signal control for an isolated intersection with Poisson arrivals is modeled and analyzed. Distribution of residual queue length and average delay is calculated cycle by cycle until the steady state is reached.

Third, an eight-phase actuated signal control is analyzed. Distribution of actuated phase length, corresponding residual queue and delay at steady-state is obtained by iterative algorithm. The other direction, whose phase length is determined by the actuated phase, is then analyzed analytically.

The last model analyzes a ramp metering system. By developing a state transition matrix and corresponding steady-state equations, the steady-state queue length

distribution can be calculated analytically. Comparing the model results with an occupancy measure of an upstream ramp detector allows one to estimate the arrival rate at the ramp meter.

In all of the queuing models developed for the above applications, iterative algorithms are developed to calculate steady probabilities for various measures. The results from the algorithms are compared with those obtained from simulations. In all cases they matched well.

CHAPTER 1 INTRODUCTION

Traffic delay resulting from congestion on urban streets and freeways continues to grow in the U.S. For example, in 2003, congestion caused 3.7 billion hours of travel delay and wasted 2.3 billion gallons of fuel for a total cost of more than $63 billion, this means an increase of 79 million hours delay and 69 million gallons of fuel from 2002 [1]. The most straight-forward approach to alleviate the congestion is to construct more roads and to add more lanes on existing streets and highways, which is not practical in many situations. Also, it has been proven that capacity increases from new construction will never catch up with the increase in traffic demand.

Another approach is to develop high-capacity transit in dense urban areas. A third approach is to reduce congestion by optimizing the traffic control system and utilizing the available system capacity more efficiently. Traffic control systems include traffic markings, traffic signs, and traffic signals.

This dissertation investigates traffic control signals at isolated intersections and freeway ramp metering systems. The performance of different traffic signals options is analyzed and compared using queuing models and other mathematical methods, along with some simulation methods. A methodology is also presented for estimating the arrival rate at ramp meter entrances.

1.1 History of Traffic Control Signal

Manually operated semaphores first used in London in 1868 can be viewed as the origin of the traffic control signals [2]. When the city realized the need to control its traffic (wagons, buggies and pedestrians), they implemented the first traffic signal in the world [3]. The signal device was operated by a policeman; it had semaphore arms and a revolving gas lantern with red and green lights for night use. Unfortunately, a month after the installation, the officer was injured by the signal light because of an explosion [3].

James Hoge developed the first electrical signal in the United States, which was installed in Cleveland, Ohio, in 1914. Interconnected signals followed shortly in 1917 in Salt Lake City, Utah [2]. During this period (from 1908 to 1927), the Ford Motor Company produced the Model T, the first affordable automobile and the car that "put America on wheels" [4]. Automobile ownership and usage expanded rapidly at this time and the need for effective vehicular traffic control was recognized. This led to the development of actuated signals, which were implemented in 1928.

Traffic signal technology has developed rapidly since this early stage. Now the traffic signal is one of the most important control elements in the traffic system, and it can greatly improve the safety and efficiency of the system if properly installed and used. The Manual on Uniform Traffic Control Devices (MUTCD) provides this definition of traffic signals [5]: "any power-operated traffic control device other than a barricade

warning light or steady burning electric lamp, by which traffic is warned or directed to take some specification." When installed in accordance with warrants and standards, the MUTCD indicates that traffic control signals have one or more of the following benefits [5]:

(1) They facilitate orderly movement of traffic.

(2) They can increase the traffic-handling capacities of the intersections where proper physical layouts and control measures are in place.

(3) They can reduce the frequency of certain type of accidents, especially right-angle crashes.

(4) They can be coordinated under favorable conditions to facilitate continuous, or nearly continuous, movement of traffic flowing at a definite speed along a given route.

(5) They can be used to interrupt heavy traffic at intervals to permit other traffic, vehicular or pedestrian, to cross.

1.2 Terms and Definitions of Urban Street Traffic Signal Control Systems

Traffic signal control systems are complicated devices with numerous modes and settings. The principal terms and definitions that follow provide guidance for understanding and analysis of the systems [6]:

(1) **Cycle**. With respect to the traffic signal, a cycle is the rotation of the

controller through all its states. Normally, if the signal timing is fixed, all approaches will receive a green light once during each cycle. If the signal timing is actuated, some movements' green phases might be skipped if there is no vehicle waiting for service.

(2) **Cycle length**. The cycle length is the total time needed for the controller to progress through an entire cycle. For fixed-time traffic signals, the cycle length is constant. With actuated and adaptive signals, the cycle length varies with the changing real-time traffic conditions. Cycle length is usually recorded in seconds; this dissertation will let C to denote cycle length.

(3) **Interval**. The interval is a time period during which signal lights, facing all directions, remain in one state. In one signal cycle, there are many intervals. Intervals can be classified into the following three types:

(a) A **green interval** indicates the corresponding movement has right of way. For efficiency, movements that do not conflict may share the same kind of interval. The following sections describe different combinations available for specific intersection geometries; the final design usually depends on the engineering consideration. Symbol G_i usually denotes green interval for movement i.

(b) A **changing interval** is the yellow indication that occurs for one or more designated movements. This interval occurs between the green

and red intervals and is used to alert the driver to prepare to stop because the traffic light for the corresponding approach will change to red shortly.

(c) A **clearance interval** begins after the changing interval. The corresponding traffic light changes from yellow into red, while all other signals stay on red. This interval serves as a safety precaution, allowing vehicles in the intersection to exit before conflicting traffic enters. Symbol Y_i will denote a changing interval plus clearance interval.

(d) A **red interval** indicates that the corresponding movement does not have the right of way. In some situations, drivers are permitted to turn right during the red interval. This interval will be represented by R_i— for movement *i*.

The intervals of different approaches overlap with each other. For example, the green interval, changing interval, and clearance interval of a movement overlap with the red interval (whole or part of it) of the intersection's conflict movements.

(4) **Phase**. A phase comprises three consecutive intervals at one approach. During each phase, one approach will receive the green interval first, then the changing interval and the clearance interval. These three intervals

consist a whole phase.

(5) **Effective green time**. Effective green time is used for system modeling and analysis. Consider the real situation, vehicles depart much slower in the beginning of the green interval due to a "start lost" time. On the other hand, the yellow interval might be used by some moving vehicles to pass the intersection if they are too close to the stop bar to make a stop. In order to ensure that the "capacity" used in a mathematical model reflects the real system accurately and simply, the model usually assumes that each vehicle departs at a constant rate during an "effective" green time. Using an appropriate effective green time for the model, one can represent capacity of the real system accurately. Effective green time is denoted by g_i. The relationship between effective green time and intervals can be presented by the following equation:

$$g_i = G_i + Y_i - t_L , \tag{1.1}$$

where t_L is total lost time per phase.

(6) **Time lost.** L denotes the time lost in the mathematical model, which includes the clearance interval plus the start time lost.

(7) **Phase diagram** (e.g. see Figure 1.1). A phase diagram depicts all the movements that can be made in each single phase (including protected or permitted left turn movements, where a dashed arrow means the

movement is permitted, and a solid arrow means a protected movement and has no conflicted traffic).

(8) **Ring diagram** (e.g. see Figure 1.2). Multiple movements are made in each phase. Usually the movements in the same phase are grouped into different "rings" in the ring diagram. The predominant ring diagram has two rings and is called a dual ring. Each ring shows the particular movements that can be made by a phase. The traffic signal controllers discussed in this dissertation are dual rings, unless otherwise specified.

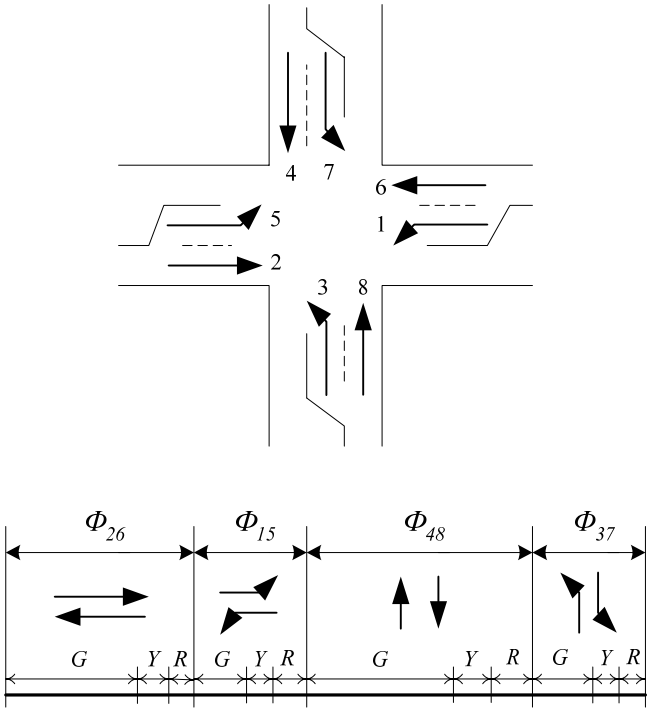


Figure 1.1 Geometric layout and phase diagram of a typical 4-leg intersection

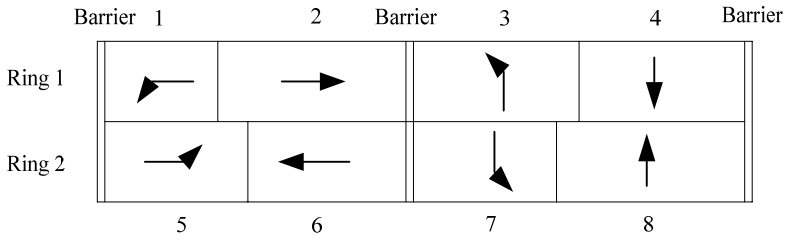


Figure 1.2 A standard NEMA dual-ring, eight phase signal control

1.3 Classification of Traffic Signal Control Systems

Traffic signal can be classified as fixed-time, actuated and traffic adaptive control.

1.3.1 Fixed-Time Traffic Signal

The **fixed-time traffic signal** has the longest history. If the signal is timed well using historical data, it can handle light traffic demand and recurrent congestion efficiently. Fixed-time signals are widely used in many urban traffic networks. With this type of control, the cycle length, phase duration, and phase sequence of each movement are preset. The timing is based on historical traffic data, obtained using some demand forecasting methodology.

Time-of–day (TOD) signal timing is widely used in an urban traffic network operating fixed-time traffic signals. Traffic engineers usually divide the day (24 hours) into four periods, for example, an AM-peak plan that operates from 06:00 – 08:30, a Mid-day plan that operates from 08:30 – 15:00, a PM-peak plan that operates from 15:00 – 19:00, and an off-peak plan for the remainder of the day [6]. TOD signal timing is easy to implement and modify, and arterial traffic flow can be synchronized

with fixed-time traffic signals. The shortcoming of this type of control is that it unable to handle random events such as unusual traffic conditions and capacity reductions due to accidents or road maintenance.

1.3.2 Actuated Traffic Signal

Actuated traffic signal can be semi-actuated, full-actuated or volume-density control [6]. For a semi-actuated signal, the intersection has a major street with high traffic demand and a minor street with relatively low demand and fluctuation. Detectors are only placed at the stop bars of the minor approaches. The signal always gives the green phase to the main street if there is not any demand detected on the minor streets. The green phase for minor approaches needs a detector call from either a vehicle or a pedestrian. The full-actuated traffic signal is suitable for isolated intersections with random arrivals. Phase length of each approach is determined by the detector calls. Each phase also has upper and lower bound. Actuated traffic control handles fluctuating traffic demand better than fixed-time control. It may not suitable for intersections with heavy traffic loads, where each phase tends to reach its upper bound and the behavior may look like a fixed-time signal. Full actuated control is rarely used where synchronization is required with other signals since its cycle length is changeable. Volume-density Control can be viewed as an improvement of actuated control, whose parameters can be adjusted to response the real time traffic.

(1) **Semi-Actuated Coordinated Control**: It is usually used in an asymmetric

intersection, which means that the intersection consists of a minor street with relatively low traffic volume and a main street with higher traffic volume. The detector is placed on the minor street. The traffic signal always services the main street until there is a detector call on the minor street, then a phase to serve minor street vehicles is included in the cycle. For coordination, the major green phase is synchronized for arterial traffic progression. If there is a call for service from the side street, it can only be served after a preset yield point (a time point) is reached; also the corresponding phase has to be finished pointing the time to ensure that the next coordinated green phase for the main street begins on time. The discipline of this kind of design is to keep an acceptable progressing for arterial while responding to demand from side streets. If there is no demand for the minor street, signal will keep serving arterial. Thus coordinated arterial through movement always receives a minimum green time plus some green time extension, while the green phase for the minor street is traffic responsive. If there is no demand from the minor street, the corresponding phase will be skipped, and the spare time is given to the coordinated green phase or other phases. Normally there is no exclusive left-turn phase for the minor street because of low traffic demand.

(2) **Full-Actuated Control**: This type of control has detectors on all approaches. The signal serves one direction till the queue is dissipated and

no new vehicle has been detected in a preset gap time or max out, then it might switch to serve the other direction if any vehicle arrival is detected. The controller should have a minimum time duration for any green phase for safety considerations. The green phase can be extended, but as soon as a vehicle detection occurs in the conflicted movement, the extension becomes constrained to a given duration. The signal then switches to serve the conflicted movement at the end of the extension. The green phase extension prevents very long waiting times for conflicted movement. A green phase may be extended before it reaches the minimum green time length, if there is a vehicle detection for that movement. Each new detection will extend the green phase by a passage time (the time required for a vehicle traveling from the detector to the stop bar). If there is no new vehicle arrival and demand exists from the conflicting direction, the signal will switch to serve the other direction, which is called “gap out”. Another situation arises when the green phase hits the upper bound, then the signal switches to serve the other direction. Because the green phases for all directions are traffic responsive, the cycle length is not constant. This type of intersection is normally used at isolated intersections (an intersection is considered as isolated if upstream signalized intersection is more than 1 mile away) in most situations. It is not suitable for signal coordination. Compared with fixed-time signals, actuated control handles large traffic

fluctuations more efficiently.

(3) **Volume-density Control**: This type of control can be viewed as an advanced Full-Actuated Control. It needs detectors to be placed far upstream of the intersection. These detectors are used to estimate the queue length at the beginning of the green phase based on detector calls during the red phase. The estimated initial queue length will be used to determine the initial green phase length, which is the time needed to serve the queue at the beginning of green phase. Normally the initial green phase also has an upper bound. Another difference is the gap timing, which is constant and equal to the passage time in actuated control. Gap timing is used to determine if the signal should be switched. If there is no new vehicle detected during the gap timing (calculated from the last arrival), the signal will switch. In the volume-density control, gap timing is set to the passage time at the beginning. It is reduced when a vehicle is detected in the conflicted approach, which means the signal tends to switch quicker if there are waiting vehicles at the conflicted direction. volume-density control has two advantages: (a) the initial minimum green time reflects the number of vehicles waiting in the queue, which is more efficient than actuated control; and (b) the gap timing reduction follows a preset curve, and reflects the changes of the saturation flow rate and passage time since the beginning of green phase (passage time is always longer at the

beginning of the green phase, because of the “start lost” time).

1.3.3 Traffic Adaptive Control

The word “adaptive” means that the phase timings are set with real-time traffic data. Adaptive traffic control systems use computer technology, allowing the system to adjust the signal timing automatically using designated algorithms. Traffic adaptive control is normally used to control a traffic network rather than an isolated intersection, where information such as detector data and signal timing of peer intersections can be obtained through the controller using a communication network. In some cases, a central system is in charge the whole network. The most well-known traffic adaptive control systems are SCOOT (Split, Cycle and Offset Optimizer Technique), SCATS (Sydney Coordinated Adaptive Traffic System) and RHODES (Real-time Hierarchical Optimized Distributed Effective System).

The SCOOT system has been developed by three companies in the United Kingdom, Ferranti, GEC, and Siemens, using algorithms developed by the Transportation Road and Research Laboratory (TRRL). SCOOT, which is employed extensively in Great Britain, is able to optimize the signal parameters (cycle length, offsets and splits) for a traffic network by using real-time detector data. The build-in traffic model uses real-time data, such as the current signal timing of each intersection signal controller and the vehicle presence data from the induction loop detectors. It also uses some static data, such as the detector locations, intersection lane settings, signal phase sequences

and some other parameters [7].

The SCATS traffic adaptive signal control system was developed by the Roads and Traffic Authority (RTA) of New South Wales, Australia. SCATS adjusts the signals' cycle length, split and offset to optimize performance measure. The traffic network intersections under control are grouped into several subsystems. Each subsystem has a critical intersection whose signal timing settings are adjusted directly by a controller that is in charge of the subsystem. Other intersections in the subsystem are always coordinated with the critical intersection, by using same cycle length and phase split. And they are coordinated to the critical intersection. The critical intersection is chosen based on the historical traffic condition of the area, such as total traffic demand, congestion or traffic intensity. Each intersection controller of the subsystem can work as an independent actuated signal if the traffic demand fluctuated dramatically. In this situation it still keeps the common cycle length of the subsystem for coordination purposes. The basic real-time traffic data used by SCATS is the "degree of saturation" (also known as the volume to capacity ratio), which is the ratio of the rate of flow rate to the capacity of the intersection. The controller adjusts the common cycle time of a subsystem by maintaining a high degree of saturation for the critical intersection's critical lane. This will reduce the amount of wasted green phase and decrease the delay. The phase splits of the critical intersection are adjusted by maintaining equal degree of saturation on competing approaches. The offsets among the intersections in a subsystem are selected by maximizing the green bound of the direction of dominant

traffic flow. SCATS does not optimize any specific objective value, such as average vehicle delay or queue lengths [8][9].

Since 1991, the College of Engineering and Mines of University of Arizona and the ATLAS research center has been conducting research in the area of Intelligent Traffic Control Systems, funded through grants and contracts from the FHWA, the ADOT, the City of Tucson and some other agencies. The research has concentrated on traffic management and logistics. A particular highlight of the research is the development of RHODES, an innovative computer-based traffic control system. It has proven to be a very effective traffic control system dealing with congestion and widely varying traffic conditions in simulation testing. Various state and federal agencies have recognized the potential of RHODES and have scheduled field operational testing in the Tucson and Tempe, both in Arizona, Clearwater (Fl) and Seattle (WA). Currently, new attributes are being added to RHODES and as its module, such as transit priority and emergency vehicle preemption. “The RHODES architecture is based on decomposing the control-estimation problem into three hierarchical levels: (1) intersection control; (2) network control; and (3) network loading” [10].

At the lowest level, intersection control, built-in models use real-time traffic information, typically from detectors, along with operational constraints established by the traffic engineer to predict traffic flows and optimize the signal phase and duration. By using a dynamic programming algorithm, these decisions are updated on

a second-by-second basis. At the middle level, the network control level, real-time information of peer intersections are used to predict platoon flows and establish coordination in the network. These decisions are made less frequently, say, every 200-300 seconds. At the highest level, the network loading model estimates the general traffic demand over a much longer duration, typically one hour. These demands will be used to calculate future platoon sizes at or near the control boundaries. In addition to this hierarchical decomposition of the control/estimation problem, the control system is distributed over the traffic network spatially. The functions performed at the intersection control level (signal timing and link flow prediction) are distributed to each intersection controller. This kind of structure reduces the computational requirements for the central control function as well as the intersection-to-central communication requirements. On the other hand, intersection-to-intersection communication is required [10].

In summary, adaptive signal control systems have the following common points,

(1) More detailed information is needed, including real-time and static information. Real-time information usually includes traffic information, such as detector data, current timing setting of local controller and peer controllers. Static information often includes the intersection geometrics, detector location and phase constrains, such as required phase sequence, if any, and upper and lower bounds of the phase lengths.

(2) Traffic controller is “intelligent”. Normally the controller has a build-in computer system which runs developed optimization algorithms rather than fixed-time and actuated signal settings.

(3) Some adaptive traffic control systems have build-in traffic dynamics models, such as RHODES, which simulate and predict the real traffic situation for the near future, based on the detector data, intersection geometrical layout and vehicle speeds.

(4) The adaptive control systems usually controls a traffic network rather than an isolated intersection. Neighboring intersection controllers may share information using an appropriative communication network. Coordination can be achieved efficiently in such a kind of system. The intersection controller are interconnected, and can be coordinated by a central controller. The intersection controller can also control the intersection independently offline in case the network is down. For example, the local controller of RHODES has a backup strategy, which is capable of running a TOD plan or be in an actuated control mode if the network communication system fails.

(5) The architecture of the adaptive control systems may have multiple layers. Lower layer focuses on the intersection signal optimization while higher layer often focus on the network-wide issues, such as coordination of platoon movement and optimization of traffic assignment.

1.4 Left Turn Treatment

Left turn treatment is very important in traffic signal control because it is related to traffic safety and intersection efficiency. Based on the traffic demand and conflicting traffic volume and speed, it might be permitted, protected or combined.

(1) Permitted **Left Turns**: left turns are permitted during the green phase of the through movement. Vehicles can make a left turn when there is an acceptable gap in the opposing through traffic.

(2) Protected **Left Turns**: There is a designated green phase for the left turn movement only. It may come along with the green phase for the through movement of the same approach or the left turn movement of the opposing approach.

(3) Combined **Left Turns** (Protected + Permitted or Permitted + Protected): As the name implied, the left turn treatment is composed of two types, either protected left turns followed by permitted left turns or permitted left turns followed by protected left turns.

If considering the phase sequence settings, left turn phase can be leading or lagging. Leading left turn means left turn movements are made ahead of the through movement, otherwise it is lagging. More details on left turn treatment will be discussed in the signal timing design.

1.5 Basic Signal Timing Design

1.5.1 Two-Phase Signal with Permitted Turns

Two-phase signal is the simplest traffic signal which is widely used at four-leg intersections with relatively low traffic volumes (see Figure 1.3 [6]). The approach receiving green lights has through movements in protected type, while left turn and right turn movements are made in permitted type. Different movements may share one lane or have their own exclusive lanes. The phase diagram shows all E-W movements are made in phase A, while N-S movements are made in phase B. The ring diagram shows that the northbound and eastbound movements belong to ring 1, other two movements belong to ring 2.

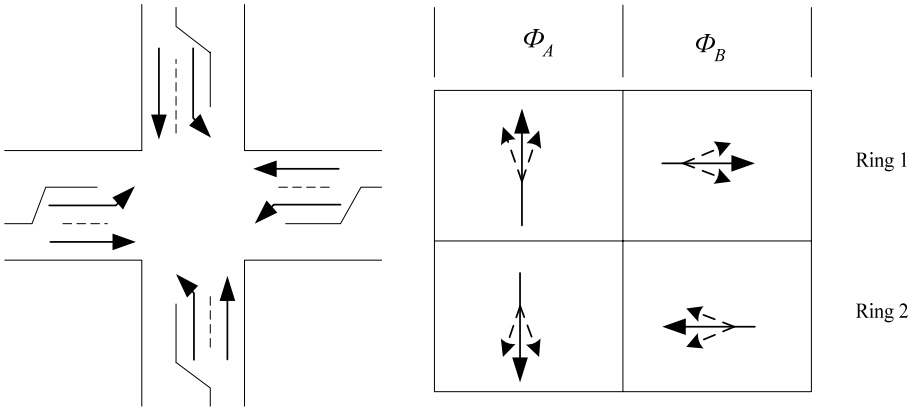


Figure 1.3 Two-phase signal with permitted left-turn of a 4-leg intersection

1.5.2 Left–Turn Phase

If the left turn traffic demand is fairly large, the left turn movements usually have protected phases (see N-S direction in Figure 1.4a). The following three general guidelines should be followed to determine if dedicated left turn phases should be used [6].

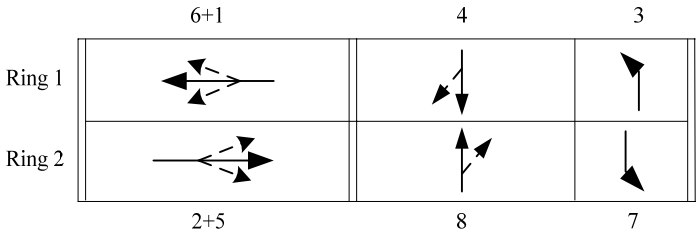
(1) Protected left-turn phase is rarely used if the demand of left turn traffic is low, say, less than 100 vehicle/hour.

(2) On the other hand, if the demand of left turn traffic is over 250-300 vehicle/hour, the protected left-turn phase is usually used.

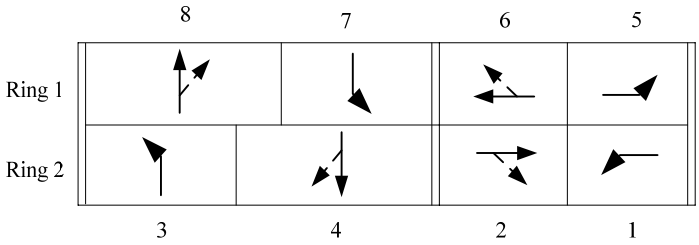
(3) If the demand is between those two thresholds above, the traffic engineer should consider other factors to determine if the left-turn phase should be used. Such the opposing traffic volume, number of lanes for through movements, and the field view and grade of the intersection.

If one movement of the intersection warrants the condition of protected left-turn phase while the opposite direction doesn't, a three-phase signal might be used (see Figure 1.4a). Most traffic signal of four-leg major intersection has protected left-turn phase for both directions. There are several possible configurations for protected left turn phasing; Figure 1.4b and 1.4c give two examples. Left turn movement can have a permitted phase plus a protected one. The permitted left turn might be prohibited because of safety consideration, such as high speed through traffic and/or poor view.

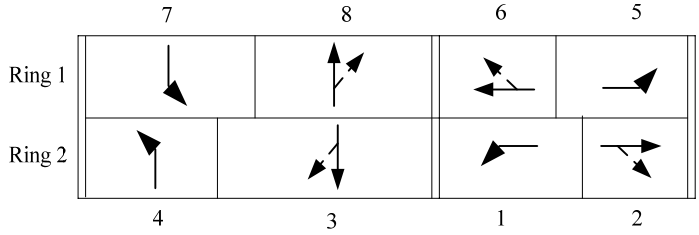
A red arrow is often used in this situation.



a) A three-phase signal with permitted left turn for one direction and protected left turn for the other direction



b) Example if a protected left turn treatment



c) A second example for protected left turn treatment

Figure 1.4 Example treatments for left-turns

Exclusive left turn for opposing approaches might be made ahead of the through movement of the same approach, or behind it. The former case is called leading left turn; the latter is called lagging left turn. Figure 1.2 shows a signal timing with leading left turn. Some researches report [11][12] that the leading left-turn can reduce

the delay for both the turning movements and through movements simultaneously, when comparing with lagging left turn, and that there is no significant difference on progression between those two types of left-turns. This can be partially explained by the following: if the left turn bay doesn't have enough capacity to handle the left-turn traffic. The left-turn queue may spill back to the through lane when the left-turn phase is lagged. The researches also report the permitted left–turn along with the protected left-turn always reduces the delays compared with protected left-turn only.

If the left-turn traffic volume fluctuates dramatically, the exclusive left-turn phase should be actuated.

Figure 1.2 is also applicable to a typical eight phase dual ring timing plan with actuated left-turn phase. There are two barriers in the diagram, which are presented by double line crossing through two rings. The barrier means the phase switch in two rings has to be synchronized (say switch from phase 2 to phase 3, and phase 6 to phase 7), otherwise conflicting movements will enter the intersection simultaneously. The single lines belong to different rings that do not have to be synchronized.

1.5.3 NEMA Eight-Phase Actuated Control

For a four-leg intersection with left turn phase, several different phase sequence plans may be applied. A generalized example showing the possible combinations is the standard National Electrical Manufacturers Association (NEMA) eight-phase actuated

control (see Figure 1.5 [6]). This control strategy is designed for four-leg intersection with exclusive left turn lanes, as showed in Figure 1.1. The eight-phase actuated control has adjustable phase lengths and sequences. It responses to the real time traffic demand information by using detectors.

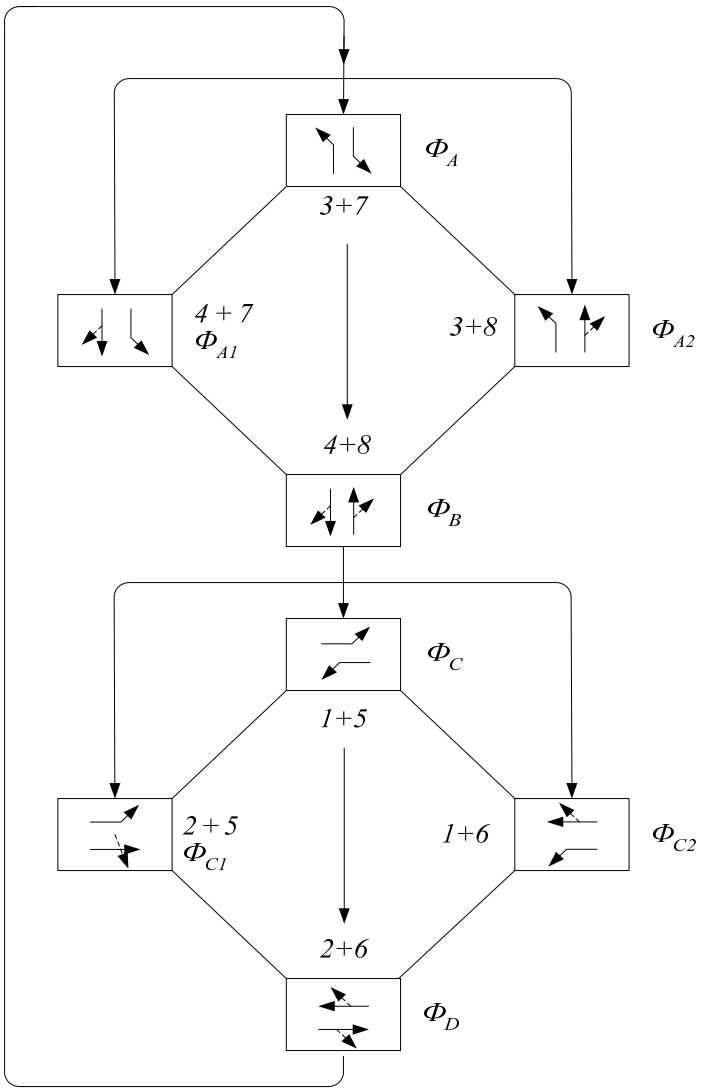


Figure 1.5 A standard NEMA eight-phase actuated phase plan

1.6 Methodologies of Analyzing Signalized Intersection

Much research has been focused on signalized intersection. A major goal of such research has been on developing new methodology to solve the following two problems: (a) how to evaluate the performance of the signalized traffic system, and (b) how to optimize traffic signal timing. The most predominant performance measures of signalized intersection are average delay, average queue length, and maximum queue length. Most researchers have used mathematical analysis and traffic simulation software to analyze delays and queues.

In this dissertation, a new stochastic queuing model is proposed. The developed model treats the traffic system as a queuing system. By assuming the vehicle arrivals follow Poisson distribution, one can analyze stochastic properties of several signalized intersection performance measures, such as queue length, green phase length and average delay. Different types of signal control systems are analyzed using this queuing model. The results are compared with each other and with results from a simulation model.

1.7 Modeling and Analysis of Ramp Metering Control System

1.7.1 History of Ramp Metering History

Freeway congestion is caused by the traffic demand exceeding system capacity. It can be classified into two basic types, recurrent or non-recurrent. Recurrent congestion is

normally due to weekday commuting, which can be predicted fairly easily. Non-recurrent congestion is hard to predict and caused by, for example, traffic accident and road works. For a freeway system, if the congestion exceeds a threshold, the system throughput drops dramatically. If too many vehicles keep entering the freeway without any control at the on-ramp, excess demand will keep increasing the congestion and reducing the system throughput. In order to relive congestion for freeway traffic, one possible solution is to use a ramp metering system to regulate the entering traffic to freeway, based on the current traffic condition. A ramp meter is a special type of traffic control signal, with two phases (red and green phase only, there is no yellow phase), located at the on-ramp of the freeway.

The first ramp metering system was installed on the Eisenhower Expressway (Interstate 290) in Chicago, Illinois in 1963. This first manually operated ramp metering system required a police officer who would stop traffic and release merging traffics on an entrance ramp with a predetermined rate, so that the vehicles could merge onto the freeway safer and smoother. Since then modern ramp-meters (which can be set with time-of-day rates without manual operator, either police officer, or by traffic engineer) have been deployed in many urban areas in the USA [13].

1.7.2 Effectiveness of a Ramp Metering System

It is claimed that ramp meters can reduce congestion (increase speed and volume) on freeways by reducing traffic demand and by making the merging traffic flow

smoother and at an appreciate rate. Two variations of demand reduction are commonly mentioned; one is access rate, the other is diversion [13].

In 2000, the Minnesota Department of Transportation (Mn/DOT) used ramp metering systems to control freeway access on 210 miles freeways in the Minneapolis-St. Paul metropolitan area (known as the Twin Cities region) [14]. Mn/DOT tested the ramp meter as a possible method of optimizing freeway operations in 1969. Currently, there are approximately 430 ramp meters installed to help merge traffic onto freeways and to manage the flow of traffic through bottlenecks [14]. Although Mn/DOT has a long history of using ramp meters to operate and optimize the freeway system, some members of the public questioned the necessarity of the ramp metering strategy in use. As a response, a bill passed in 2000 session by the Minnesota Legislature, which asked Mn/DOT to analyze the effectiveness of ramp meters in the Twin Cities Region by a comparison study [14]. Mn/DOT formed two committees, an advisory committee and a technical committee, to conduct the study while representing the public and ensuring the credibility and objectivity [14]. The study was conducted in the fall of 2000; and the results were reported to the Legislature and the public in early 2001 [14]. “The goal of the comparison study was to evaluate and report any relevant facts, comparisons, or statistics concerning traffic flow and safety impacts associated with deactivating system ramp meters for a predetermined amount of time” [14].

The research concluded that ramp metering was a cost-effective investment in the

following ways [14]:

(1) **Traffic Volumes and Throughput**: After the meters were switched off, traffic volume on freeways declined by an average of nine percent and there was no significant traffic volume fluctuation on parallel arterials included in this ramp metering study. During the peak hour period, throughput of freeway mainline decreased by an average of 14 percent while the ramp metering system were switched off.

(2) **Travel Time**: Without ramp meters, the additional travel time on freeway facilities are greater than the ramp delay reduction.

(3) **Travel Time Reliability**: Without ramp metering system, variance of freeway travel time is almost twice as that with ramp metering. The study showed ramp metering system reduced approximately 2.6 million hours of unexpected delay annually.

(4) **Safety**: Without the ramp metering system and considering the seasonal variations, peak hour period accidents on metered freeways and ramps increased by 26 percent. This result meant ramp metering reduced 1,041 crashes annually or approximately four crashes per day.

(5) **Emissions**: This study also showed ramp metering saved 1,160 tons of emissions annually.

(6) **Fuel Consumption**: Ramp metering increased fuel consumption by 5.5 million gallons annually. This was the only criteria category which was

deteriorated by ramp metering system.

(7) **Benefit/Cost Analysis**: The study showed ramp metering saved approximately $40 million to the Twin Cities traveling public per year. The benefits of ramp metering outbid the costs by a net benefit of $32 to $37 million for each year. “The benefit/cost ratio indicates that benefits are approximately five times greater than the cost of entire congestion management system and over 15 times greater than the cost of the ramp metering system alone” [14].

The results show that ramp meters is an efficient method of improving the operations of a freeway network in an urban area in several ways, such as increase in system throughput, decreases of delays, improved safety and lower emission [14].

Another study also indicates that ramp meters can efficiently reduce system travel time, while the savings are network dependent [15].

Besides the benefits that can be achieved by using a ramp metering system, it also has some disadvantages and negative effects to the traffic system:

(1) It encourages drivers to select longer routes because of the extra delay at the meter,

(2) Only the through traffic gain the benefits, not the local traffic,

(3) Queue spillbacks might worsen the operations of local traffic network,

(4) Local surface street network will have heavier traffic load because of the diversion trips,

(5) Extra cost is needed to operate and maintain the system.

Even through it has some shortcomings, ramp metering system is widely used in many metropolitan areas of US because of the good cost benefit ratio and better level of service of freeway system.

1.7.3 Overview of Different Types of Ramp Metering Systems

Current ramp metering control systems can be classified in several ways. Based on the system architecture, a ramp metering system can be local or system wide [16],

(1) **Local Ramp Metering**: A local ramp controller operates independently of the other controllers on the freeway, it only manages the traffic condition of a local section of the freeway.

(2) **System-wide Ramp Metering**: A system-wide ramp meters manage several connected freeway sections, sometimes they even control a freeway network. In this kind of system, local ramp meter exchanges real-time traffic condition through a designated communication network. If the system consists of a hierarchical structure, usually a central controller is in charge of the whole system. Individual ramp meters can get some commands from the central controller, adjusts its own signal timing to

optimize the whole traffic network operations under control, rather than only optimize the operation of the local section.

Based on the control strategy, ramp metering systems can be classified into the following three types [16]:

(1) **Operator Controlled Metering**: The ramp meter is manually operated by an operator. Some real-time information might be required, such as loop detector information or surveillance video.

(2) **Fixed-time Ramp Metering**: The signal timing of fixed-time ramp meter is preset, normally based on historical traffic data. Some fixed-time ramp meters use time-of-day signal timing. With time-of-day timing, the meter may only regulate the entering traffic during the peak hour. It increases the metering capacity or even switches off during the off-peak period. Fixed-time ramp metering is reliable and easy to be implemented and maintained. Its behavior is easy to be predicted by the user too. And it is fairly efficient in dealing with recurrent congestions of the freeway. The fixed-time ramp metering system also has some shortcomings. It is an open-loop control, which means it doesn't response to the occasional fluctuation of real-time traffic, such as special events and accidents. Generally, timing settings for this kind of system tends to be more conservative by underutilizing the freeway, which prevents overloading the system during normal traffic

demand fluctuations. Therefore it may decrease the efficiency of the system by wasting some system capacities. If there is some heavy traffic congestion caused by a non-recurrent event, the system can become unstable because the merging traffic through unadjusted signal from the on ramps may exceed the reduced system capacity. Some times it may even cause unnecessary delay at ramp area. If there are high demand fluctuations at a ramp area, spill-back queues may exist and block the urban street intersection.

(3) **Traffic Responsive Ramp Metering**: This type of ramp control can respond to the real-time traffic situation. The system use loop detectors or some other detecting methods to collect the real-time traffic condition of the freeway and ramp area, such as traffic volume, traffic speed on freeway and vehicle queue size on the ramp. Based on the optimization objective and real-time system state, the ramp meter adjusts its signal timing by using some designated optimization method. Since it responds to the real-time traffic condition, it handles the fluctuation of traffic on freeway and merging traffic more efficiently. It is also capable of dealing with non-recurrent traffic fluctuations. The system is more complicated and expensive compared with the fixed-time ramp meter system.

The predominant traffic-responsive ramp controls are demand-capacity control. With demand-capacity control, ramp rates are calculated based on a real-time comparison

of upstream traffic flow and downstream capacity of the freeway. The difference between the upstream flow and the downstream capacity is used to set the maximum entrance ramp flow rate [17]. This ramp flow is expressed as a metering rate to be used during the next control interval. One example of such an algorithm is Asservissement LINeaire d'Entree Autroutiere (ALINEA) [17], developed by engineers at the Technical University of Munich. ALINEA is a local-feedback control algorithm that manages the metering rate to keep the downstream occupancy of the freeway at a certain level, which is called the "occupancy set point" [18].

Modern ramp meter systems even consider the interdependency of ramp operations of a freeway network rather than one section. The primary objective of integrated ramp control is to optimize operations and reduce congestion on the freeway network. Therefore, the optimization objective of each ramp considers the demand-capacity control of the whole network rather than the demand-capacity control of the section it is located in [18]. A good example is MILOS [19], which uses real-time traffic data and optimization algorithms to set ramp metering systems to optimize a given performance measure.

There is very limited research focusing on the analytical modeling and analysis of traffic at a ramp meter system. In this dissertation, an analytical queuing model is proposed. It can be used to analyze fixed-time and simplified adaptive ramp meter control systems. The model can solve the steady state probabilities analytically and

estimate the vehicle arrival rate by comparing the results with an upstream ramp detector's occupancy.

1.8 Outline of the Dissertation

This chapter introduces traffic signal control, including its history, and a classification of different types of control systems. In the last part of this chapter, ramp metering for freeways is introduced briefly. Chapter 2 reviews research on signalized intersection analysis. Such efforts focus on the intersection delay analysis and on optimizing fixed-time signal setting based on estimated traffic volumes.

Chapter 3 presents a methodology for modeling queues in a simplified adaptive signal control system. In the model, intersection has two one-way traffics. Vehicle arrivals follow Poisson process. With a simplified adaptive signal control the half cycle lengths are random variables. Their distribution can be calculated sequentially until a steady state is reached. The results are compared with those given by simulation based model.

Chapter 4 develops a queuing model for fixed-time signal control of the same intersection as that analyzed in Chapter 3. With a fixed-time traffic signal and Poisson arrivals, a residual queue distribution at the end of each cycle can be calculated sequentially until a steady state distribution is reached. Several performance measures can be calculated using this queuing model. The results are compared with those

given by adaptive control in Chapter 3 and with the results obtained by Webster's formula [20].

In Chapter 5, the exact distribution of actuated green phase is calculated and applied to eight-phase dual-ring actuated signal control. Because of the random arrivals, the actuated left turn phase length becomes a random variable. The method developed iteratively calculates the distribution of the left turn phase length and corresponding residual queue length until a steady state is reached. The intersection performance measures are compared with the simulation based results.

Chapter 6 proposes a queuing modeling for a ramp metering system. For both fixed metering rate setting and adaptive rate setting, the model can analytically solve for the steady state queue distribution with a specific arrival rate. By comparing the detector occupancy data with the model's results, one can estimate the arrival rate of merging traffic.

Chapter 7 summarizes the dissertation, points out its contributions, and discusses some areas of future research.

CHAPTER 2 LITERATURE REVIEW

Several performance criteria can be used to evaluate signalized intersection, such as average number of stops per vehicle, vehicle throughput, average queue length and average stop delay. Among these indicators, average delay is perhaps the most predominant since it is easily understood, related to travel costs and vehicle emission. For example, the Highway Capacity Manual (1985) uses it as the only measure to determine the level of service of signalized intersections [21].

2.1 Deterministic Intersection Delay Model

Several researchers have established relationships between cycle lengths and intersection delays over many years. The earliest one can be traced back to Clayton [22], who assumed both the arrival and the departure processes were constant and deterministic. If the traffic demand is light, this model gives a fairly good estimate on vehicle delay. Given the uniform arrival and departures, this model assumes queue is dissipated at the end of each green phase. That means the capacity utilization is always 100% and a residual queue never occurs. If traffic loads are fairly high, this model tends to underestimate the delay.

Matson, Smith, and Hurd gave a delay formula based on 3.7 seconds of lost time per phase and 2.1 seconds of average headways. They also gave the optimal cycle length for a two-phase operation [23],

$$\frac{T}{n} = \frac{R + 4.75}{2} - \frac{Q(A - 2.1)}{R + 4.75}, \tag{2.1}$$

$$c_o = \frac{34200}{3600 - (V_1 + V_2) * h}, \tag{2.2}$$

where

T/n is the average delay,

R is the interval of stop signal aspect in seconds,

Q is adjustment factor,

V_1 and V_2 are the approaching critical lane volumes in vehicles per lane per hour,

h = headway (assumed as 2.1 seconds per vehicle),

c_o = cycle length which gives minimum average delay.

A general equation based upon saturation flows is as follows:

$$c_o = \frac{nk}{1 - \sum_{i=1}^{n} \frac{V_i}{s}}, \tag{2.3}$$

where

s = saturation flow, vehicle per hour per lane,

V_i = critical lane volume per hour per lane for approach i,

k = 4.75,

s= lost time per phase plus 1.05s,

n = number of phases.

Thus, nk = 9.5s, 13.75s, and 18.0s for two, three, and four-phase intersections, respectively.

Very similar to Clayton's model, this formula doesn't consider the randomness of the arrival process. It underestimates the average vehicle delay when traffic load is heavy.

2.2 Webster's Formula

The most well known intersection delay model is that due to Webster, who came up with analytical expressions for delays and queues under the assumption of Poisson arrivals. In 1958, Webster published the first widely used delay equation for fixed-time signals [20]. Delays were computed from a function of green times, cycle lengths, and traffic flows, assuming that the saturation flow remains constant and the intersection is under-saturated over the period of time considered. The derivation of the delay equation is based on the theoretical and numerical simulation approaches,

$$d = \frac{c(1-\lambda)^2}{2(1-\lambda x)} + \frac{x^2}{2q(1-x)} - 0.65\left(\frac{c}{q^2}\right)^{1/3} x^{(2+5\lambda)}, \tag{2.4}$$

where

d = average delay per vehicle on the particular approach of the intersection,

c = cycle time,

λ = proportion of the cycle which is effectively green for the phase under consideration (i.e. g/c, g is the green phase length),

q = traffic flow rate,

s = saturation flow rate,

x = the degree of saturation. This is the ratio of the actual flow to the maximum flow that can move through the intersection from this approach, and is given by $x = q/(\lambda \cdot s)$.

As mentioned before, Expression (2.4) is not a pure theoretical derivation. The first term of the equation represents the delay caused by uniform arrivals. The second term represents the extra delay caused by the randomness of arrival process. It is because residual queue may exist and cause extra delay if number of excess arrivals is larger than the capacity of this approach. If the average delay is represented by only these two terms, then average delay would be given by Equation (2.5),

$$d = \frac{c(1-\lambda)^2}{2(1-\lambda x)} + \frac{x^2}{2q(1-x)}. \tag{2.5}$$

The average delay given by (2.5) has fair agreement with simulation results. The third term of Equation (2.4) is an empirical correction term, used to further improve the agreement with the simulation results.

Another very important contribution of Webster's research is the settings of optimal cycle length. Here the term "optimal" means the signal with minimum vehicle delay. By taking the first derivative of the delay formula and equate it to zero we obtain the optimal cycle length with minimum delay, given by,

$$c_o = \frac{1.5nL + 5}{1 - Y}, \tag{2.6}$$

where,

$$Y = \sum_{i=1}^{n} y_i .$$

y_i is the critical movement's volume to saturation flow rate for phase i,

L is the lost time for each cycle.

Based on Webster's finding, the lower bound of cycle length is approximately half of c_o. If the cycle length is below half of c_o, queues will accumulate forever, which means the intersection traffic is not stable. If the cycle length falls in the range of 0.75 c_o to 1.5 c_o, the corresponding delay is within 120% of the minimum delay.

The Webster's formula gives a fairly accurate estimation of intersection delay for

fixed-time signals and random arrival. It is still widely used by traffic engineers.

The drawback of Webster's formula is the limitation of applicable situation. It can only analyze intersection with low to medium traffic load. Otherwise the value of *Y* will be too close or greater than 1, which will result either in an excessive long or a negative cycle length (see Equation 2.6).

2.3 Stochastic Intersection Delay Model

Miller proposed a stochastic model to analyze the traffic delay and optimal settings of fixed-time traffic signals [24]. In this model, the average delay can be approximately estimated using the following equation,

$$\bar{d} = \frac{(c-g)}{2c(s-q)}\left\{\frac{sI(2x-1)}{q(1-x)} + s(c-g) + I - 1 + q/s\right\}. \tag{2.7}$$

The notation in this equation is the same as that of Equation (2.4), with *I* as the ratio of the variance to the mean of the number of arrivals.

In this model, the arrival process is assumed to be random, which means the intersection is isolated. The inter arrival times can follow any identical and independent distribution; only the ratio of the variance to the mean is required. If a steady state of the system exists, the estimation of residual queue and average delay can be derived analytically.

Miller also tried to minimize average delay through optimizing signal settings,

$$c = \frac{L + \sqrt{I_i L(1 - q_i / s_i)/[s_i \pi_i (1 - \pi_i)^2]}}{1 - (q_i / s_i \pi_i)}, \tag{2.8}$$

where

$$\pi_i = \frac{\sqrt{(y_i I_i)} + 1.2\sqrt{(y_i y_j)}[\sqrt{(y_i I_j)} - \sqrt{(y_j I_i)}]}{\sqrt{(y_i I_i)} + \sqrt{(y_j I_j)}}.$$

Another modeling approach was introduced by Newell who treated traffic as continuous fluid, with either deterministic or stochastic properties, when he analyzed traffic delays for fixed-time signals [25]. In his method, approximations were made by the law of large numbers or the central limit theorem, which are not very sensitive to the detailed stochastic structure of the arrival or departure processes [25]. His model also requires *I*, the ratio of the variance to the mean of the number of arrivals. The average delay derived by this method is,

$$E\{w\} = \frac{R^2}{2(R+G)(1-q/s)} + \frac{IH(\mu)}{2s}[\frac{G}{R+G} - \frac{q}{s}]^{-1} + \frac{RI}{2s(R+G)(1-q/s)^2}. \tag{2.9}$$

In this formula, most notation is same as that used for Webster and Miller's formula; $H(\mu)$ is a function related to the stochastic properties of queue lengths.

Newell pointed out that Webster's, Miller's and Newell's formulae were probably equally useful for delay estimation in typical applications. The main virtue of the

Newell's formula is that it is based on central limit theorem and the formula tends to be more accurate for sufficiently large values of *sG* [25]. Another common point and also a limitation of these formulae is that these are steady state results, which assumes the arrival patterns remain the same throughout the period under study.

Newell and Osuna also developed stochastic intersection delay model for vehicle actuated signal for both one-way and two-way streets [26][27]. Using continuous fluid approximation, they derived the following delay formulae for the signalized one-way street,

$$E\{W_1\} \cong \frac{q_1}{2(1-q_1/s_1)} \cdot [(E\{R\}+Y)^2 + \operatorname{var} R + \frac{I_1(E\{R\}+Y)}{s_1(1-q_1/s_1)} + \frac{V_1}{s_1 q_1}], \tag{2.10a}$$

$$E\{W_2\} \cong \frac{q_2}{2(1-q_2/s_2)} \cdot [(E\{G\}+Y)^2 + \operatorname{var} G + \frac{I_2(E\{G\}+Y)}{s_2(1-q_2/s_2)} + \frac{V_2}{s_2 q_2}]. \tag{2.10b}$$

The notation in the above equations are same as those of Equation (2.4) except,

$E\{W_1\}$ is the delay for movement 1 and $E\{W_2\}$ is the delay for movement 2,

R and *G* are the red and green phase of movement 1, respectively,

V_i is the start lost time for movement *i*, and

I is a specific ratio which represents the relationship between the variance of the departures and the mean of corresponding green phase length.

For a two-way intersection symmetric between E-W (movement 1 and 2) and N-S (movement 3 and 4) direction, Newell derived the following delay formulae

$$E\{W_1\} = E\{W_2\} \cong \frac{q_1}{2(1-q_1/s_1)} \cdot [(E\{R\}+Y)^2 + \operatorname{var} R], \tag{2.11a}$$

$$E\{W_3\} = E\{W_4\} \cong \frac{q_3}{2(1-q_3/s_3)} \cdot [(E\{R\}+Y)^2 + \operatorname{var} R], \tag{2.11b}$$

using same notation, and

$$G = Y(q_1/s_1)/(1-q_1/s_1-q_2/s_2),$$
$$R = Y(q_2/s_2)/(1-q_1/s_1-q_2/s_2),$$

2.4 Time-Dependent Delay Model

Hurdle discussed the limitations of the steady-state analyses for the stochastic and the deterministic models [28]. He pointed out that the steady-state of the stochastic model can estimate the average delay well when the traffic load is low to medium. Also, he states that oversaturated intersection is easy to analyze, since the system throughput would be close to deterministic and equal to the capacity. The hardest part lies between these two situations, where overflow queues might exist. The much used formula to determine the delay caused by overflow queues has developed by Akcelik [29][30],

$$OD = \begin{Bmatrix} (T/4)\{(x-1)+[(x-1)^2+12(x-x_o)/cT]^{1/2}\} & if \quad x > x_o \\ 0 & otherwise \end{Bmatrix}, \tag{2.12}$$

where T is the time during which the arrival flow exceeds the capacity,

$x=v/c$,

and $x_o=0.67+sG/600$.

The generalized time-dependent models which are used in the Highway Capacity Manual [21], and the Australian [29][30] and Canadian [31] counterparts were analyzed and compared by Akcelik [32]. The overflow delay in these models is more empirical than theoretical. He pointed out that the differences between the three models are on setting the calibration parameters. The delay formula used in these three models can be generalized as the follows,

$$d = \frac{0.5c(1-\mu)^2}{1-\mu x} + 900Tx^n[(x-1)+\sqrt{(x-1)^2+m(x-x_o)/QT}], \quad (2.13)$$

where:

d = average overall delay (including stop-start delays) in seconds per vehicle,

c = signal cycle time in seconds,

μ = g/c (ratio of effective green time to cycle time),

x = degree of saturation (ratio of arrival flow to capacity),

T = analysis period in hours,

Q = Capacity in vehicle per hour,

m, n = calibration parameters, and

x_o = the degree of saturation below which the second term of the delay formula is zero, which can be expressed as

$$x_o = a + bsg$$

where,

sg = capacity of each cycle (s = saturation flow rate in vehicles per second and

g = effective green time in seconds), and

a, b = calibration parameters.

Akcelik pointed out that Equation (2.11) can be used to derive HCM [21], Australian [29][30] and Canadian [31] formulae if proper parameters are selected [32]. He also compared the average delay given by those three formulae. The result shows HCM formula tends to give higher average delay than the other two formulae in oversaturated condition, and the differences increase as the degree of saturation increases. He also gave an alternative of HCM formula by adjusting the m, n, a and b values in (2.11), which then gave lower average delay in the oversaturated condition.

2.5 Queuing Model

Olszewski [33][34] proposed a queuing delay model to study the probability distribution of overflow queue with fixed-time traffic signals. In his model, sequential calculation of queue length distribution is performed based on the known distribution of arrival process (either stationary or non-stationary). This model also considered the random variations of cyclic capacities. If arrival process is Poisson, the results agree with Akcelik's model well. The model is also capable of analyzing the variance of average delay. The most obvious advantage of Olszewski's model is that it can be used to analyze non-stationary arrivals, which cannot be performed by other available analytical models. Besides average delay, the variance is another very important signalized intersection performance measure, which can be evaluated by his model.

Many traffic models are proposed to estimate queue-length, but the procedures of these models tends to give different results because of using different queue definitions and different computational approaches. Kenneth [35] developed a ration framework to compare their behavior with that of the proposed Highway Capacity Manual (HCM) 2000 queue model. Various queue estimation models' outputs are converted to their HCM 2000 equivalent. The results showed that HCM 2000 queue estimates are generally higher than those of most other models. The reason is HCM model considered control parameters such as controller type and progression quality as well as for the random and overflow effects associated with traffic flow.

2.6 Focus of this Research

This dissertation will re-visit queuing theory and queuing models in traffic control, and will develop analytical models to evaluate traffic control strategies. Iterative algorithms will be developed to solve each of these models. In particular, the methodology will be applied to

(1) two-phase adaptive signal control at an isolated intersection,

(2) two-phase fixed-time signal control for an isolated intersection,

(3) eight-phase actuated signal control at an intersection, and

(4) ramp-metering control strategies.

The analytical results will be compared with results from simulation models.

The main methodological contribution of the dissertation is the approach developed to numerically solve for the steady-state behavior of the underlying queuing models. This general methodology has not been reported in literature, although there are some limited applications to fixed-time signal control [33][34].

The iteration is based on developing scenario trees. The new methodology proposed in this dissertation can be widely used to analyze different types of traffic control systems, and in other queuing applications where scenario trees can be efficiently developed.

For signalized intersection, probability of each scenario is calculated cycle by cycle with the assumption of Poisson arrivals. Algorithms perform iterative computations until steady-state is reached. The steady-state of ramp metering systems can be calculated analytically by developing state-transition matrices and corresponding steady-state equations. Knowing the exact steady-state probability of each scenario, the stochastic properties of traffic control systems can be analyzed.

The queuing modeling methodology developed in the dissertation can be extended to model and analyze some new scenarios, for example but not limited to, adaptive signal control for two-way traffic, four-phase full-actuated signal control, and signalized intersections with different arrival streams.

CHAPTER 3 QUEUING MODELING OF ADAPTIVE SIGNAL CONTROL SYSTEM

This chapter develops an analytical approach to evaluate traffic-adaptive signal control systems based on queuing models. In particularly, a queuing model is developed for a simplified adaptive control strategy where a signal serves two movements alternatively. In this strategy, the first movement is served until the queue dissipates, then the second movement is served until the queue dissipates, then the signal goes back to serve the first movement and this cyclic process repeats. The strategy is an approximation of the RHODES strategy [10], which usually switches to next phase when most of the served queue clears; but where delay minimization instead of queue clearance criterion is used. A numerical algorithm is developed to compute steady-state performance measures such as average delays and expected queue lengths. These results are compared with simulation-based results and, indeed, the analytically method predicts accurately the simulation results.

3.1 Introduction

In this chapter, a stochastic queuing model is developed for a two-phase simplified adaptive control strategy where the signal is switched from the current phase to the other phase when the served queues just vanish; it switches back to this phase after the served queue in next phase are depleted. We note that such an adaptive control strategy was first presented and studied by Newell [36].

In section 3.2 of this chapter, the expectation of green phase duration, or the conditional "busy period" of this adaptive control strategy, given multiple vehicles waiting at the beginning of green phase, is calculated using transform methods. The result shows that when k vehicles are queued up at the beginning of the green phase, then the expected time to clear the queue is k times the busy period of the corresponding queuing system with the same parameters. This result is used in section 3.3 to build the queuing model for the adaptive control analyzed in this chapter. In section 3.4, numerical results are given and compared with simulation-based results. Several loading patterns are studied, including symmetric intersection traffic with equal arrival rates, and asymmetric intersection traffic with unequal arrival rates. The results show that the analytical model predicts accurately the simulation results.

3.2 Busy period analysis

In a general first-in first-out single-server queuing system, a busy period begins when a customer arrives and finds no queue and the server is free to serve him immediately. A busy period ends when the server completes the service of a customer and finds there are no customers in queue waiting for service (i.e., it is a zero queue again). The interval between the end of a busy period and the beginning of the next constitutes an idle period. The intersection can be viewed as an $M/G/1$ system, where M means the arrival process is Poisson; G means the service time follows general distribution, and 1 means there is single server. Since arrivals are Poisson in an $M/G/1$ system, the idle

period, which ends with the next arrival, has a negative exponential distribution. The distribution of the busy period of *M/G/***1** system is available in many queuing theory textbooks [37].

In this chapter we study different types of busy periods that occur due to the competition of the green phases for the various vehicle streams approaching an intersection.

Assuming vehicles arrive in a Poisson manner with rate λ. Let the service rate (or the saturation flow rate) of the intersection be μ. Let v denote the time to serve one vehicle, its distribution as $b(v)$ with expectation $\overline{v} = 1/\mu$, $B^e(s)$ is its s-transform. Let q be the busy period when there is only one vehicle waiting at the beginning of the green phase. Define $f(q)$ as the distribution function of q, $F^e(s)$ is its s-transform, which is defined as

$$F^e(s) = \int_0^\infty e^{-sq} f(q) dq \,. \tag{3.1}$$

The expectation of q is [37].

$$E(q) = \frac{1/\mu}{1-\lambda/\mu} \,. \tag{3.2}$$

We will now compute the expectation of busy period, say, q' when k vehicles are waiting at the beginning of green phase. Define $f(q')$ as the distribution function of q',

$F'^{e}(s)$ is its s-transform. The length of the busy period, q', consists of the length of service, kv, for the first k vehicles, plus the length of the busy period for the subsequently arriving customers. If n vehicles arrive during the period kv, then these n customers may be visualized as standing aside for service. The first of the n customers gets serviced immediately after the termination of service of the initial k vehicles; if there are any new arrivals during this service time, then, without loss of generality in computing total service time, they may be considered to be served before another customer of the (n-1) remaining in queue is admitted for service. The same modified priority rule may be applied when the next customer goes into service. Therefore, the total busy period is

$$q' = kv + q_1 + q_2 + \cdots + q_n . \tag{3.3}$$

where the q_i are independent (since the arrival process is Poisson) and have the same distribution as q. Consequently, the conditional Laplace-transform of q' given there are n arrivals in time kv is

$$F'^{e}(s/n, kv) = E[e^{-sq'}] = e^{-skv}[F^{e}(s)]^{n} . \tag{3.4}$$

To find the unconditional s-transform, we note that the unconditional distribution may be derived by

$$f(q') = \int_0^{\infty} \sum_{n=0}^{\infty} f(q'/n, kv) P[n/(kv)] b(v) dv , \tag{3.5}$$

where

$$P[n/(kv)] = (\lambda kv)^n e^{-\lambda kv} / n!, \quad n = 0,1,2,\cdots, \tag{3.6}$$

and $b(v)$ is the service time distribution. The s-transform of (3.5) gives

$$F'^{e}(s) = \int_0^\infty \sum_{n=0}^{\infty} F'^{e}(s/n, kv) P[n/(kv)] b(v) dv1. \tag{3.7}$$

Using (3.4) and (3.6) and summing appropriate elements, we get

$$F'^{e}(s) = \int_0^\infty e^{-[ks+k\lambda-k\lambda F^e(s)]v} b(v) dv. \tag{3.8}$$

By definition of s-transform, the above relation becomes

$$F'^{e}(s) = B^e[ks + k\lambda - k\lambda F^e 1(s)]. \tag{3.9}$$

By differentiation, the expected length of a busy period caused by the k vehicles in queue is

$$\begin{aligned} E(q') &= -\frac{d}{ds} F'^{e}(s)\,|_{s=0} \\ &= -\frac{d}{ds} B^e(0) \frac{d}{ds}[ks + k\lambda - k\lambda F^e(s)]\,|_{s=0}. \end{aligned} \tag{3.10}$$

Substituting $\bar{v} = -\frac{d}{ds} B^e(0)$ in (3.10) we have

$$E(q') = \bar{v}[k + k\lambda E(q)]. \tag{3.11}$$

From (3.2) and (3.11) we get

$$E(q') = \frac{k/\mu}{1-\lambda/\mu} = kE(q). \tag{3.12}$$

This result means that the expectation of busy period caused by k vehicles in queue is k times the expectation of the busy period caused by one vehicle in queue. Intuitively, if one treats the queuing discipline as a last-in-first-out system (note that the length of busy period remains unchanged, because the order in which the vehicles are served is irrelevant as far as the busy period is concerned), then all k vehicles can be served one by one at times when the busy period of the previous vehicle is just terminated. Hence, each of these k vehicles constitutes an independent busy period q. The total busy period caused by these k vehicles is thus k times the individual expected busy periods.

3.3 Analysis of adaptive signal control

Consider an isolated two-phase intersection with one-way traffic streams without turning movements. Let phase Φ_1 serve the W-E flow and Φ_2 the N-S flow. As illustrated in Figure 3.1, let the arrival rate and intersection service rate of W-E (N-S) direction be λ_1 (λ_2) veh/s and μ_1 (μ_2) veh/s, respectively. We will assume that there is sufficient intersection capacity so that flow ratio $\rho_i=\lambda_i/\mu_i$ is less than one. Furthermore, for the existence of a steady state for the signal control, we will assume that total effective service rate for both arrival streams, which includes start-up loss during queue discharge, is greater than total arrival rate for both arrivals.

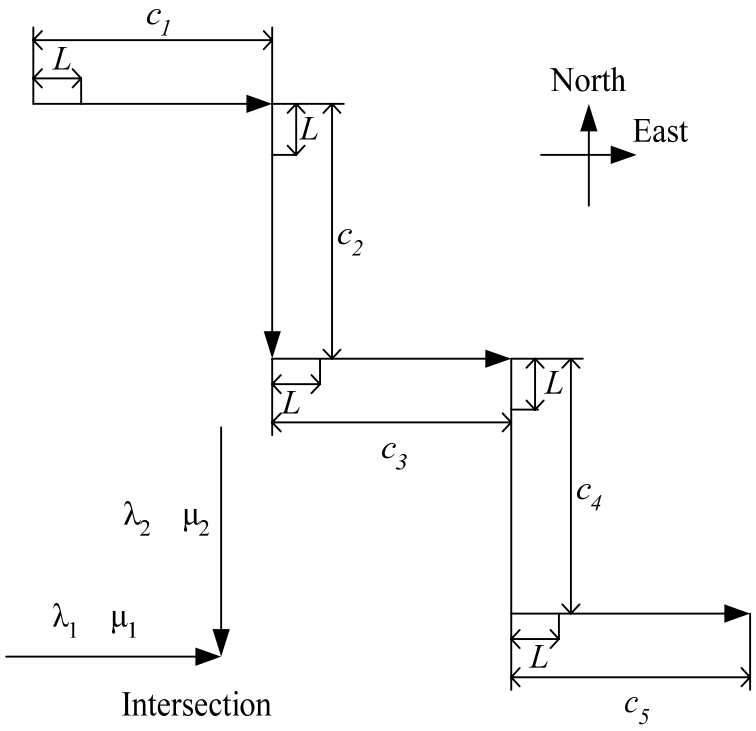


Figure 3.1 Adaptive traffic phases at an intersection

In Figure 3.1, the time axis is divided into half cycles for the convenience of analysis. Each half cycle, c_i, is composed of a constant queue discharge lost time L plus a green phase. The odd (even) subscript on c means it serves W-E (N-S) direction (see Figure 3.1).

In the adaptive control scheme analyzed in this chapter, phase Φ_1 serves the W-E queue and new arrivals until there are no vehicles in queue. Then Φ_2 serves N-S queue and arrivals until there is no queue. Then it goes back to Φ_1 and so on.

If vehicles arrive at deterministically constant rates, then the optimal signal times (with minimum average delay) for adaptive control can be easily computed as follows. During phase Φ_2, exact $(c_1+c_2)\cdot\lambda_2$ vehicles, which arrived in the last half cycle and the current one need to be served, while total service capacity during this half cycle is $(c_2-L)\cdot\mu_2$. Analysis for the next phase Φ_1, for time c_3 is similar. That is

$$\begin{aligned}\lambda_2(c_1+c_2) &= \mu_2(c_2-L)\\ \lambda_1(c_2+c_3) &= \mu_1(c_3-L).\end{aligned} \tag{3.13}$$

At steady state, $c_1=c_3=c_5\ldots$, and $c_2=c_4=c_6\ldots$ Then solving for c_1 and c_2, the optimal half cycle lengths of two directions are

$$c_1 = \frac{1+\rho_1-\rho_2}{1-\rho_1-\rho_2}L, \tag{3.14a}$$

$$c_2 = \frac{1-\rho_1+\rho_2}{1-\rho_1-\rho_2}L, \tag{3.14b}$$

where flow ratio $\rho_i = \lambda_i / \mu_i$.

Now consider Poisson arrivals. For a reasonable initialization of the process needed for our numerical analysis, let us assume we start with half cycle c_1 and c_2 given by (3.14). Starting at the beginning of c_2, W-E vehicle arrival process become Poisson with rate λ_1. Beginning of c_3, N-S vehicles arrive in Poisson manner with rate λ_2. This initialization does not affect the steady state.

Since the adaptive signal serves one direction till queue dissipates, then it switches to the other direction, and the switching continues, the c_3, $c_4 \ldots c_n$ are random variables, whose distributions depend on the number in queue (random) and time to serve them.

Let X_{2i} denote the number of vehicles that arrive in the W-E direction during (c_{2i}+L) and X_{2i+1} the N-S vehicles during (c_{2i+1}+L). Then the probability mass function (PMF) of X_2 is

$$P_{k_2} = P(X_2 = k_2) = \frac{[\lambda_1 (c_2 + L)]^{k_2}}{k_2!} e^{-\lambda_1 (c_2 + L)}. \tag{3.15}$$

From (3.11) we can get the conditional (given k_2) expected green phase length

$$c_3 - L = k_2 \frac{1/\mu_1}{1 - \rho_1}. \tag{3.16}$$

The traffic signal switches as soon as the queue becomes zero. Assuming vehicles depart at saturation flow rate during the green phase, the conditional expected number of vehicles served during (c_3-L) is

$$n_3 = (c_3 - L)\mu_1 = \frac{k_2}{1 - \rho_1}. \tag{3.17}$$

To calculate the total delay of W-E vehicles in cycle (c_2+c_3), we need the distribution of the vehicle arrival times in this cycle. Suppose n vehicles arrived in (0, t). Given the Poisson arrival process, the arrival times of these n vehicles, s_1, $s_2 \ldots s_n$, are

independently and uniformly distributed in this interval [38]. Therefore the conditional expected total delay, D_3, can be viewed as the area of the triangle in Figure 3.2,

$$\begin{aligned} D_3 &= \frac{1}{2} k_2 (c_2 + c_3) \\ &= \frac{1}{2} k_2 (c_2 + L + \frac{k_2 / \mu_1}{1 - \rho_1}). \end{aligned} \tag{3.18}$$

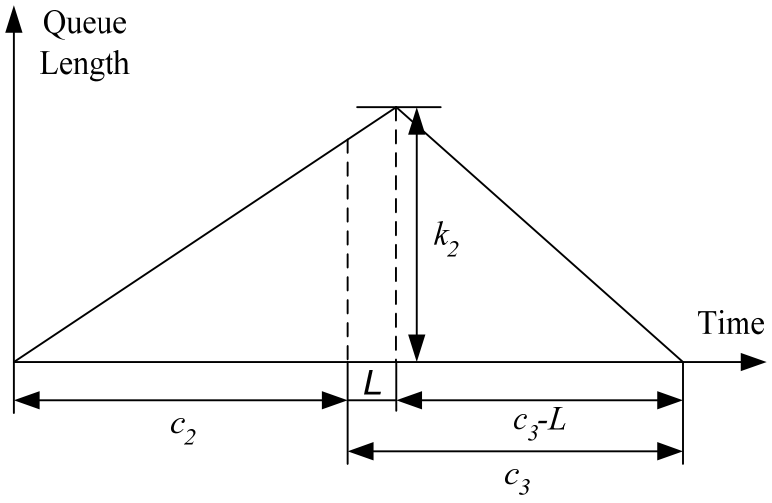


Figure 3.2 Queue length plot in one cycle

Hence, the conditional (given k_2) average delay, d_3, of W-E vehicles during (c_2+c_3) is

$$\begin{aligned} d_3 = D_3 / n_3 &= \frac{1}{2} k_2 (c_2 + L + \frac{k_2 / \mu_1}{1 - \rho_1}) \Big/ (\frac{k_2}{1 - \rho_1}) \\ &= \frac{1}{2} [(c_2 + L)(1 - \rho_1) + k_2 / \mu_1]. \end{aligned} \tag{3.19}$$

The (unconditional) expectation of d_3 is

$$E(d_3) = \sum_{k_2=1}^{\infty} P_{k_2} D_3 / n_3$$
$$= \frac{1}{2}[1 - P(k_2 = 0)](c_2 + L)(1 - \rho_1) + \frac{1}{2}\sum_{k_2=1}^{\infty} P_{k_2} \frac{k_2}{\mu_1}. \quad (3.20)$$

Next step is the calculation of the distribution of X_3, the number of N-S vehicles arriving in the time interval (c_3+L). Since c_3 is a discrete random variable, whose PMF is given by (3.15) and (3.16), the calculation of PMF of X_3 is different from that of X_2. For any realization of (c_3+L), we can get the probability of exactly k_3 N-S vehicle arrivals. The summation of these probabilities over all possible (c_3+L) gives the probability of exactly k_3 vehicle arrivals during (c_3+L), that is the following PMF of X_3

$$P_{k_3} = P(X_3 = k_3)$$
$$= \sum_{all(c_3+L)} P[X_3 = k_3 \mid (c_3 + L)]P[(c_3 + L)], \quad (3.21a)$$

when summing over realization of k_2, is equal to,

$$\sum_{k_2=0}^{\infty} P_{k_2} P\{X_3 = k_3 \mid [(c_3 + L) = 2L + \frac{k_2 / \mu_1}{1 - \rho_1}]\}$$
$$= \sum_{k_2=0}^{\infty} P_{k_2} \frac{[\lambda_2 (c_3 + L)]^{k_3}}{k_3!} e^{-\lambda_2 * (c_3 + L)} \quad (3.21b)$$
$$= \sum_{k_2=0}^{\infty} P_{k_2} \frac{[\lambda_2 (2L + \frac{k_2 / \mu_1}{1 - \rho_1})]^{k_3}}{k_3!} e^{-\lambda_2 (2L + \frac{k_2 / \mu_1}{1 - \rho_1})}.$$

From (3.12), the conditional (given k_3) green phase (c_4-L), which serves N-S direction, is

$$c_4 - L = k_3 \frac{1/\mu_2}{1-\rho_2}. \tag{3.22}$$

The conditional expected number of vehicles served during this green phase is

$$n_4 = (c_4 - L)\mu_2 = k_3 \frac{1/\mu_2}{1-\rho_2} u_2 = \frac{k_3}{1-\rho_2}. \tag{3.23}$$

Similarly, the conditional expected total delay of N-S vehicle during (c_3+c_4) is

$$\begin{aligned} D_4 &= \frac{1}{2} k_3 (c_3 + c_4) = \frac{1}{2} k_3 (c_3 + L + \frac{k_3/\mu_2}{1-\rho_2}) \\ &= \frac{1}{2} k_3 (\frac{k_2/\mu_1}{1-\rho_1} + \frac{k_3/\mu_2}{1-\rho_2} + 2L). \end{aligned} \tag{3.24}$$

Hence, the conditional average delay d_4 during (c_3+c_4) is

$$\begin{aligned} d_4 &= D_4 / n_4 = \frac{1}{2} k_3 (k_2 \frac{1/\mu_1}{1-\rho_1} + 2L + \frac{k_3/\mu_2}{1-\rho_2}) \Big/ (\frac{k_3}{1-\rho_2}) \\ &= \frac{1}{2}(1-\rho_2)(k_2 \frac{1/\mu_1}{1-\rho_1} + 2L) + \frac{1}{2}(\frac{k_3}{\mu_2}). \end{aligned} \tag{3.25}$$

The (unconditional) expectation of d_4 is

$$\begin{aligned} &E(d_4) = \frac{1}{2}(1-\rho_2) \cdot \\ &\left\{ 2L[1 - P(k_3 = 0)] + \sum_{k_2=1}^{\infty} P_{k_2} \frac{1/\mu_1}{1-\rho_1} k_2 \right\} + \frac{1}{2} \sum_{k_3=1}^{\infty} P_{k_3} \frac{k_3}{\mu_2}. \end{aligned} \tag{3.26}$$

Using such expressions on succeeding half cycles until steady state is reached, one can get the steady-state distribution of (1) the number of vehicles been served in a

cycle, (2) green phase lengths, and (3) vehicular delays, including average vehicular delays in each direction, and in each cycle.

In our initialization of half cycle lengths, uniform arrivals were assumed. Other initial half cycles could have been used. The reason we selected these half cycles was because they are close enough to steady-state so that steady-state can be calculated faster.

The final issue in this section concerns the numerical computation of the steady state probabilities and associated performance measures. Since there are infinite possible realizations of X_2, X_3, $\cdots$ and the number of possible scenarios goes up exponentially it may appear that the computational load is unmanageable. But if we truncate the state space by neglecting extremely low probability scenarios, the computational load is very manageable. At any stage, we fathom a node in the scenario tree when its probability is below a given threshold, say, 10^{-6}. We then can get a relationship of the various measures of interest from one half-cycle to another.

Suppose at step 1 of the computational process, a finite PMF for X_2 is calculated (by ignoring very low possibility scenarios). Then one can easily compute the distribution of (c_3+L) since it is linearly related to the PMF of X_2. This allows the computation of the conditional probability of k_3 arrivals for each realization of (c_3+L), which gives us the PMF of X_3 in step 2 of the process. This scenario tree is schematically represented in Figure 3.3; note that now the number of scenarios do not blow-up exponentially.

The formulae to compute various measures from step to step are derived below. In most cases, less than 30 steps were needed for the expressions of steady-state probabilities to converge within a threshold of 10^{-6}.

When the distribution of X_{i-1}, number of vehicle arrivals in $(c_{i-1}+L)$, is known, we can get the distribution of the expected length of (c_i+L), since it is linearly related to k_{i-1} as $c_i + L = 2L + \frac{k_{i-1}/\mu_j}{1-\rho_j}$ where $j = 1$ when i is odd and $j = 2$ when i is even.

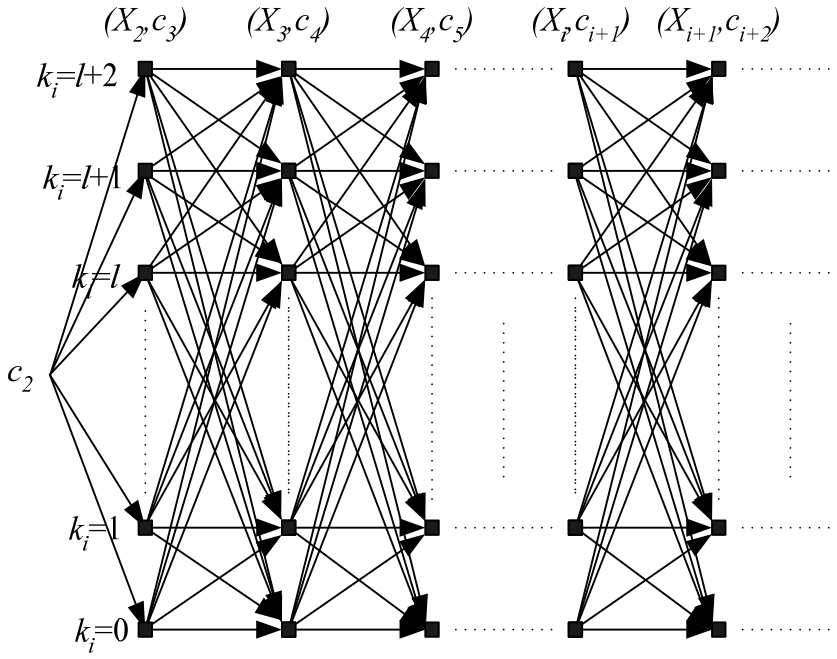


Figure 3.3 The scenario tree of the system

Then the distribution of X_i is given by

$$
\begin{aligned}
P_{k_i} &= P(X_i = k_i) \\
&= \sum_{all(c_i+L)} P[X_i = k_i \mid (c_i + L)]P(c_i + L) \qquad (3.27a) \\
&= \begin{cases} \sum_{k_{i-1}=0}^{\infty} P_{k_{i-1}} P\{X_i = k_i \mid [(c_i + L) = 2L + \frac{k_{i-1}/\mu_1}{1-\rho_1}]\} & i \quad is \quad odd \\ \sum_{k_{i-1}=0}^{\infty} P_{k_{i-1}} P\{X_i = k_i \mid [(c_i + L) = 2L + \frac{k_{i-1}/\mu_2}{1-\rho_2}]\} & i \quad is \quad even. \end{cases}
\end{aligned}
$$

Since X_i arrivals are Poisson, (3.27a) can be simplified to

$$
P_{k_i} = \begin{cases} \sum_{k_{i-1}=0}^{\infty} P_{k_{i-1}} \dfrac{[\lambda_2(2L + \frac{k_{i-1}/\mu_1}{1-\rho_1})]^{k_i}}{k_i!} e^{-\lambda_2(2L + \frac{k_{i-1}/\mu_1}{1-\rho_1})} & i \quad is \quad odd \\ \sum_{k_{i-1}=0}^{\infty} P_{k_{i-1}} \dfrac{[\lambda_1(2L + \frac{k_{i-1}/\mu_2}{1-\rho_2})]^{k_i}}{k_i!} e^{-\lambda_1(2L + \frac{k_{i-1}/\mu_2}{1-\rho_2})} & i \quad is \quad even. \end{cases} \qquad (3.27b)
$$

Since the relationship between the conditional green phase length (c_{i+1}-L) and k_i is

$$
c_{i+1} - L = \begin{cases} \dfrac{k_i/\mu_2}{1-\rho_2} & i \quad is \quad odd \\ \dfrac{k_i/\mu_1}{1-\rho_1} & i \quad is \quad even, \end{cases} \qquad (3.28)
$$

the conditional (given k_i) expected number of vehicles served during this green phase, given by the product of the green phase length and the service rate, is

$$
n_{i+1} = \begin{cases} \mu_2(c_{i+1} - L) = \mu_2 \dfrac{k_i/\mu_2}{1-\rho_2} = \dfrac{k_i}{1-\rho_2} & i \quad is \quad odd \\ \mu_1(c_{i+1} - L) = \mu_1 \dfrac{k_i/\mu_1}{1-\rho_1} = \dfrac{k_i}{1-\rho_1} & i \quad is \quad odd. \end{cases} \qquad (3.29)
$$

Using queue length plots such as shown in Figure 3.2, the conditional (given k_i)

expected total delay during the two half cycles ($c_i + c_{i+1}$) is

$$
\begin{aligned}
&D_{i+1} = \frac{1}{2} k_i (c_i + c_{i+1}) \\
&= \begin{cases} \frac{1}{2} k_i (c_i + L + \frac{k_i / \mu_2}{1-\rho_2}) & i \quad is \quad odd \\ \frac{1}{2} k_i (c_i + L + \frac{k_i / \mu_1}{1-\rho_1}) & i \quad is \quad even \end{cases} \\
&= \begin{cases} \frac{1}{2} k_i (\frac{k_{i-1} / \mu_1}{1-\rho_1} + \frac{k_i / \mu_2}{1-\rho_2} + 2L) & i \quad is \quad odd \\ \frac{1}{2} k_i (\frac{k_{i-1} / \mu_2}{1-\rho_2} + \frac{k_i / \mu_1}{1-\rho_1} + 2L) & i \quad is \quad even. \end{cases}
\end{aligned}
\tag{3.30}
$$

The conditional average delay d_{i+1} during (c_i+c_{i+1}) is

$$
\begin{aligned}
&d_{i+1} = \frac{D_{i+1}}{n_{i+1}} = \begin{cases} \frac{1}{2} k_i (\frac{k_{i-1} / \mu_1}{1-\rho_1} + 2L + \frac{k_i / \mu_2}{1-\rho_2}) \Big/ \frac{k_i}{1-\rho_2} \\ \frac{1}{2} k_i (\frac{k_{i-1} / \mu_2}{1-\rho_2} + 2L + \frac{k_i / \mu_1}{1-\rho_1}) \Big/ \frac{k_i}{1-\rho_1} \end{cases} \\
&= \begin{cases} \frac{1}{2}(1-\rho_2)(\frac{k_{i-1} / \mu_1}{1-\rho_1} + 2L) + \frac{1}{2}(\frac{k_i}{\mu_2}) & i \quad is \quad odd \\ \frac{1}{2}(1-\rho_1)(\frac{k_{i-1} / \mu_2}{1-\rho_2} + 2L) + \frac{1}{2}(\frac{k_i}{\mu_1}) & i \quad is \quad even. \end{cases}
\end{aligned}
\tag{3.31}
$$

Finally, the (unconditional) expectation of d_{i+1} is

$$
\begin{aligned}
&E(d_{i+1}) = \\
&\begin{cases} L(1-\rho_2)[1-P(k_i=0)] + \\ \frac{(1-\rho_2)}{2(1-\rho_1)} \sum_{k_{i-1}=0}^{\infty} P_{k_{i-1}} \frac{k_{i-1}}{\mu_1} + \frac{1}{2} \sum_{k_i=1}^{\infty} P_{k_i} (\frac{k_i}{\mu_2}) & i \quad is \quad odd \\ L(1-\rho_1)[1-P(k_i=0)] + \\ \frac{(1-\rho_1)}{2(1-\rho_2)} \sum_{k_{i-1}=0}^{\infty} P_{k_{i-1}} \frac{k_{i-1}}{\mu_2} + \frac{1}{2} \sum_{k_i=1}^{\infty} P_{k_i} (\frac{k_i}{\mu_1}) & i \quad is \quad even. \end{cases}
\end{aligned}
\tag{3.32}
$$

Equation (3.32) was used to propagate expected delay from half cycle to half cycle

until it converged.

3.4 Model evaluation

In this section, the analytical queuing model is evaluated by comparing its numerical solution with simulation-based results. The simulation model of the adaptive control strategy was built in MATLAB. The random number generator of MATLAB was to generate two groups of exponentially distributed random variables for vehicle inter-arrival times in the two directions. Service rates and lost time were the same as those of the analytical models. For each type of loading pattern, the simulation model was run 10 times, with length of 500,000 seconds each. The first 10,000 seconds of each run were considered warm-up periods and truncated.

Several loading patterns were simulated. In each case we examined (i) the average number of served vehicles in one cycle, (ii) the average half cycle lengths given by the numerical method and simulation, (iii) the average delays and (iv) the number of iterations for the numerical method to converge.

Case 1: Symmetric intersection with balanced flow ratios

In this case, the saturation flow rates of the two directions are 0.5 veh/s, the arrival rates of two traffic streams are equal, and range from 0.05 veh/s to 0.2 veh/s. Queue discharge lost time was set to 4 seconds for all the cases.

Flow ratio of N-S	Number of vehicles served		Half cycle length	
	Analytical	Simulation	Analytical	Simulation
0.10	0.50	0.50	5.00	5.00
0.12	0.63	0.63	5.26	5.26
0.14	0.78	0.78	5.56	5.55
0.16	0.94	0.94	5.88	5.88
0.18	0.13	1.12	6.25	6.25
0.20	1.33	1.33	6.67	6.66
0.22	1.57	1.57	7.14	7.14
0.24	1.85	1.84	7.69	7.68
0.26	2.17	2.17	8.33	8.34
0.28	2.55	2.54	9.09	9.08
0.30	3.00	3.00	10.00	9.99
0.32	3.56	3.53	11.11	11.05
0.34	4.25	4.24	12.5	12.49
0.36	5.14	5.12	14.29	14.24
0.38	6.33	6.32	16.67	16.64
0.40	8.00	8.07	20.00	20.13

Table 3.1 Comparison of average number of vehicles served in one cycle and average half cycle lengths (case 1)

Since the system is symmetric, performance measures should be the same for either direction and, therefore, only measures for one direction are listed in Table 3.1, which gives the average number of vehicles served in a cycle and average half cycle lengths. The table shows that the means of these measures computed by the analytical model match well with those given by the simulation model. Another interesting finding is that the means of these two measures for adaptive signal control are equal to those of

a system with deterministically constant arrivals (with same arrival and service rate) and corresponding optimal half cycles.

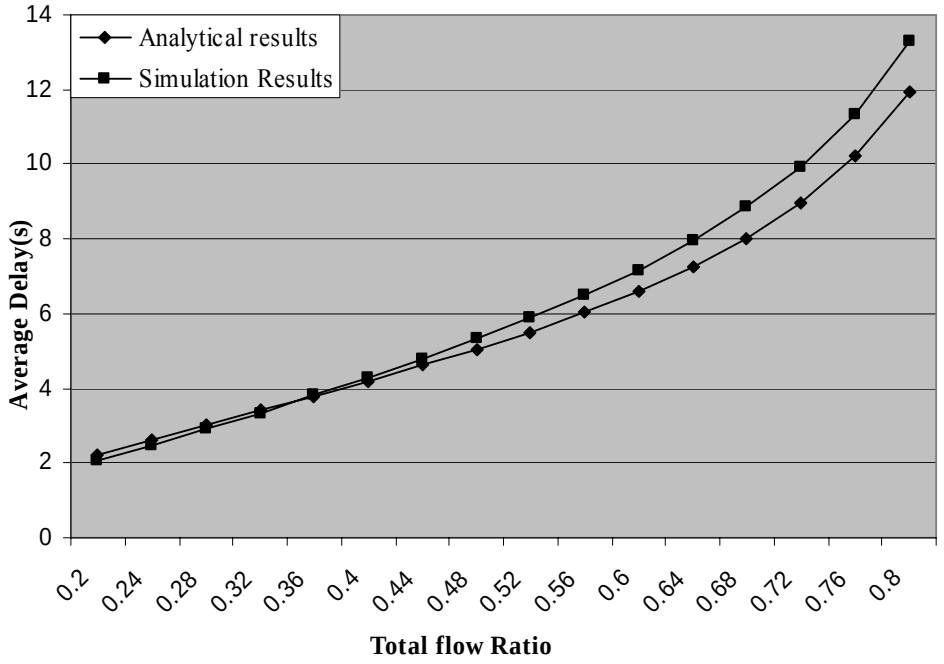


Figure 3.4 Average intersection delay (Case 1)

Figure 3.4 shows the average intersection delays calculated by the queuing model and obtained by simulation. The delay plots show that the average delays given by the two methods match well when the total flow ratio is less than 0.6. As the total flow ratio increases beyond this point, the differences increase but are still less than 10%. The analytical model always tends to underestimate the average delays compared to the simulation model. The differences are due to the method of calculating the green phase length; the analytical model truncates low probability states and uses expected green phase length while the simulation model uses actual phase length (i.e. the

distribution of the green phase lengths). Since the truncation threshold is very small, say 10^{-6}, decreasing the threshold won't reduce the differences very much. The difference can be reduced by using the distribution of the phase length.

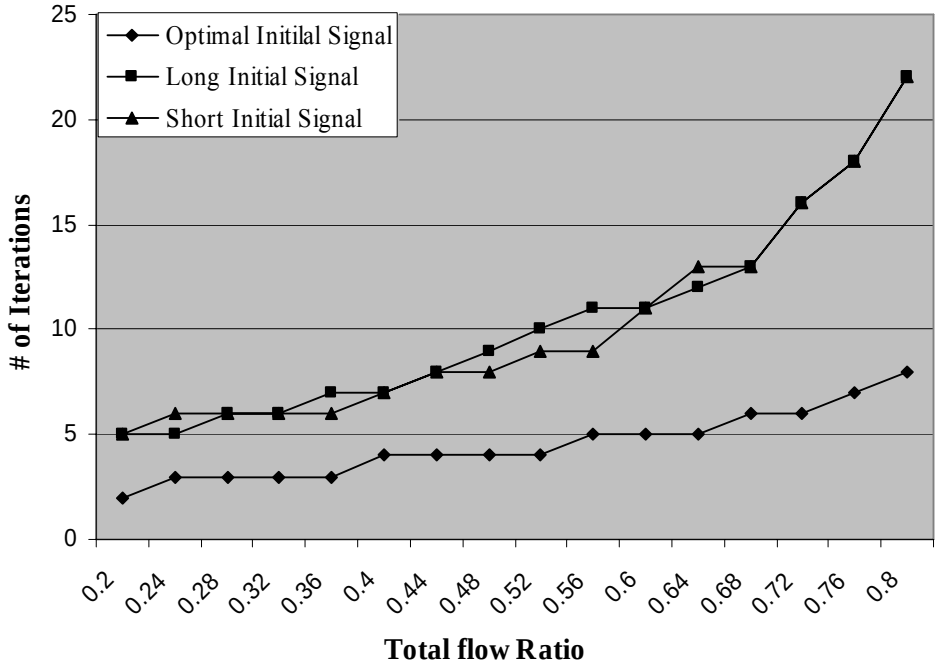


Figure 3.5 Number of iterations to converge (Case 1)

For a numerical method, its convergence, in terms of iterations needed, is a major issue. Figure 3.5 shows the numbers of iterations needed to converge, defined as when the performance measure does not change by more 10^{-6} from iteration to iteration. Of course the time to converge depends on the initial cycle lengths. "Optimal (deterministic) initial cycle" in the figure indicates that the initial cycle length determined by (3.14) based on deterministic arrivals. For comparison, we also used a

long initial cycle (where the optimal cycle length is doubled) and a short initial cycle (where the length is halved). While the initial signal settings do not affect the final states at convergence, the system converges much faster with optimal initial cycle where, in most cases, less than 10 iterations are needed to reach steady state. The figure also shows that, as the total flow ratio increases, the number of iterations needed to converge also increases.

Case 2: Symmetric intersection with unbalanced flow ratios

The departure flow rates for both directions were set to 0.5 veh/s again. The arrival rates of N-S vehicles ranged from 0.03 veh/s to 0.13 veh/s, while the arrival rates of E-W vehicles were twice those of N-S vehicles.

Table 3.2 gives the average numbers of vehicles served in one cycle and average half cycle lengths in the N-S direction, given by the analytical model and the simulation model. The results match well.

Flow ratio of N-S	Number of vehicles served		Half cycle length	
	Analytical	Simulation	Analytical	Simulation
0.06	0.29	0.29	4.59	4.59
0.08	0.42	0.42	4.84	4.84
0.10	0.57	0.57	5.14	5.14
012	0.75	0.75	5.50	5.49
0.14	0.97	0.97	5.93	5.94
0.16	1.23	1.23	6.46	6.45
0.18	1.57	1.56	7.13	7.13
0.20	2.00	2.00	8.00	8.00
0.22	2.59	2.57	9.18	9.14
0.24	3.43	3.44	10.86	10.88
0.26	4.72	4.75	13.45	13.50

Table 3.2 Comparison of average numbers of vehicles served in one cycle and average half cycle lengths in N-S direction (Case 2)

Figure 3.6 and 3.7 show the average delays in the two directions given by the two methods. The delay plots show that the average delays given by the two methods are very close when the flow ratio is relatively low, say less than 0.3 for each direction. As the intersection flow ratio increases, the differences increase. As before, the numerical queuing model tends to underestimate the average delay as compared with the simulation model. The differences in the E-W direction (which has a higher flow ratio) are larger. Figure 3.8 shows the average intersection delays.

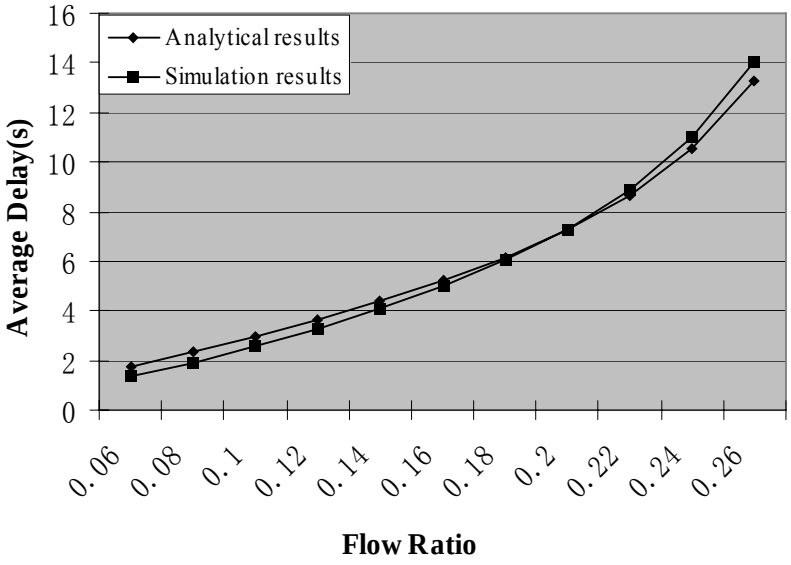


Figure 3.6 Average delay of N-S direction (Case 2)

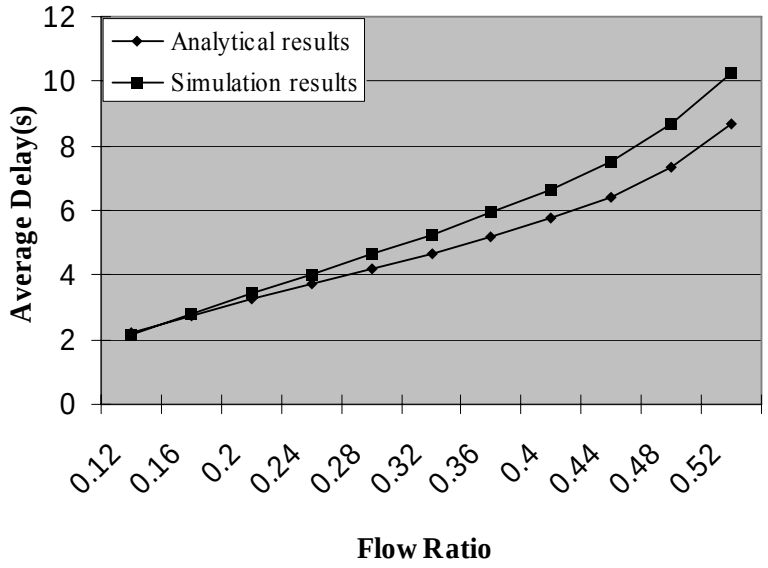


Figure 3.7 Average delay of E-W direction (Case 2)

Figure 3.9 plots the numbers of iterations needed to converge at various flow ratios.

Again, when optimal initial cycle lengths are used, the numerical method converges much faster; with "bad" initial cycle it still converges to the same steady state, but much slower. As before, more iterations are needed to converge when the flow ratio increases.

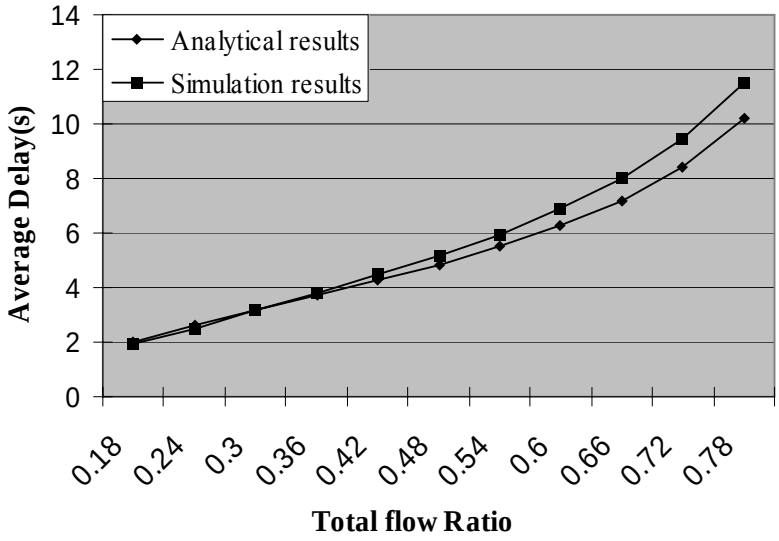


Figure 3.8 Average intersection delay (Case 2)

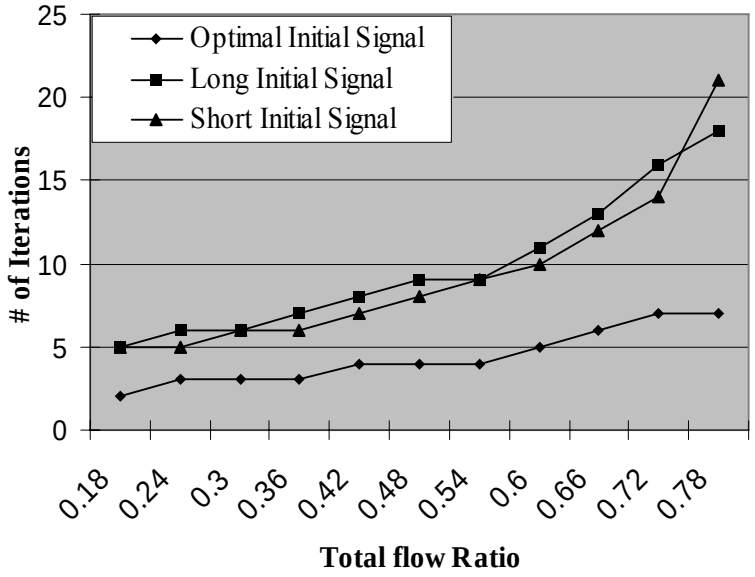


Figure 3.9 Number of iterations to converge (Case 2)

Case 3: Asymmetric intersection with balanced flow ratios

In this third case, the departure rate of N-S vehicles is 0.5 veh/s, and of E-W vehicles is 1.0 veh/s. The arrival rates of N-S direction ranges from 0.05 veh/s to 0.20 veh/s and the arrival rates of E-W direction are twice of those of N-S vehicles.

Table 3.3 gives the average numbers of vehicles served in one cycle and average half cycle lengths in E-W direction given by the analytical and the simulation models. Again the results of two models match well.

Flow ratio of E-W	Number of vehicles served		Half cycle length	
	Analytical	Simulation	Analytical	Simulation
0.10	1.00	1.00	5.00	5.00
0.12	1.26	1.26	5.26	5.26
0.14	1.56	1.57	5.56	5.57
0.16	1.88	1.88	5.88	5.88
0.18	2.25	2.25	6.25	6.25
0.20	2.66	2.66	6.67	6.66
0.22	3.14	3.14	7.14	7.14
0.24	3.69	3.72	7.69	7.72
0.26	4.33	4.34	8.33	8.34
0.28	5.09	5.10	9.09	9.1
0.30	6.00	6.02	10.00	10.02
0.32	7.11	7.11	11.11	11.11
0.34	8.50	8.49	12.50	12.49
0.36	10.28	10.29	14.29	14.29
0.38	12.67	12.7	16.67	16.7
0.40	16.00	16.10	20.00	20.09

Table 3.3 Comparison of average numbers of vehicles served in one cycle and average half cycle lengths of E-W direction (Case 3)

The average delays in the two directions and the combined intersection delays are given in Figures 3.10, 3.11 and 3.12. The numbers of iterations needed to converge for different flow ratios and initial signal cycles are given in Figure 3.13. In summary, the numerical solution of the queuing model approximates well the simulation results and converges quickly.

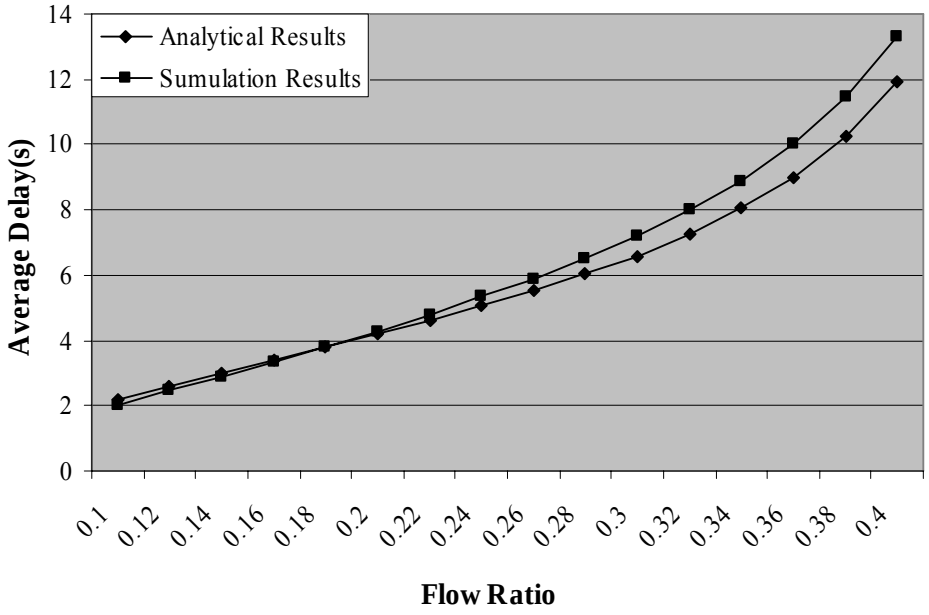


Figure 3.10 Average delay of N-S direction (Case 3)

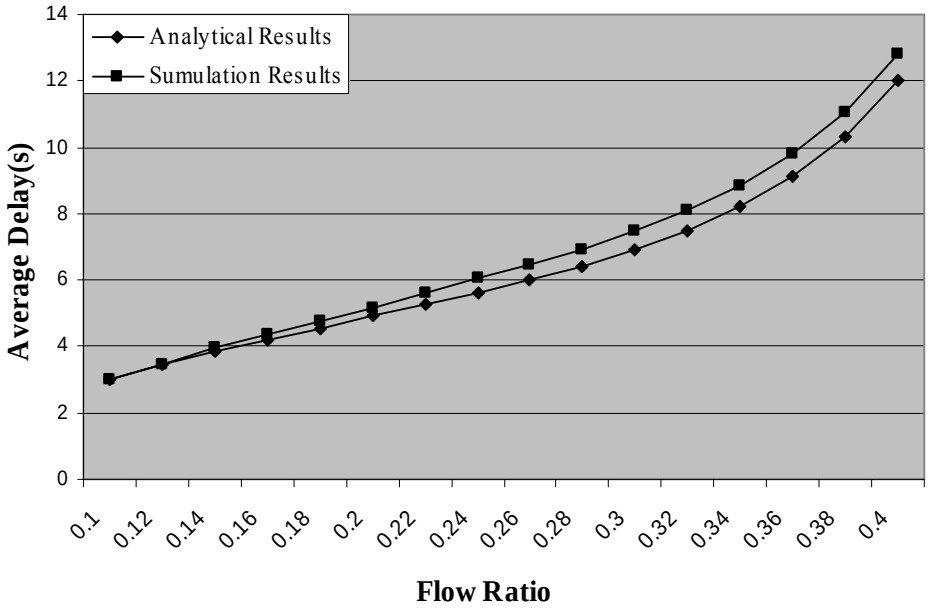


Figure 3.11 Average delay of E-W direction (Case 3)

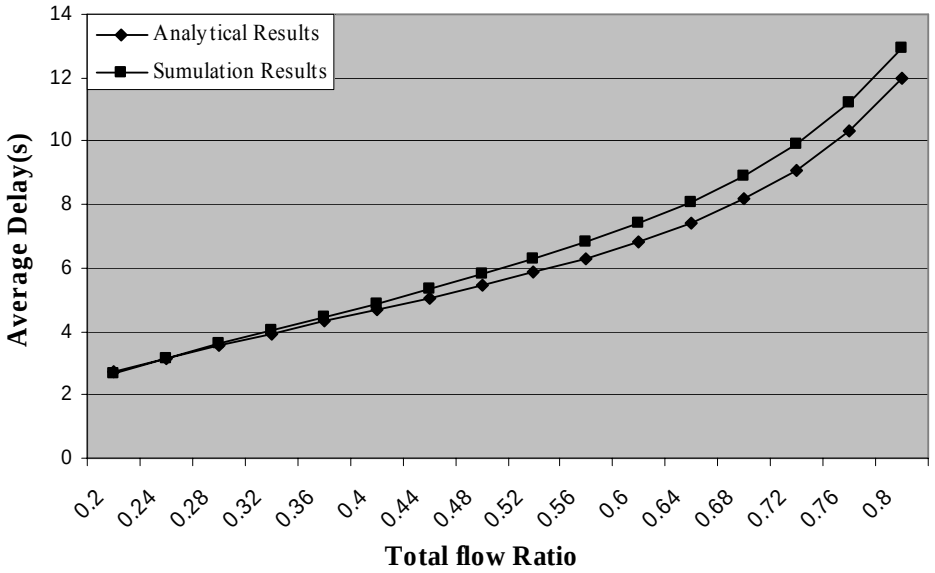


Figure 3.12 Average intersection delay (Case 3)

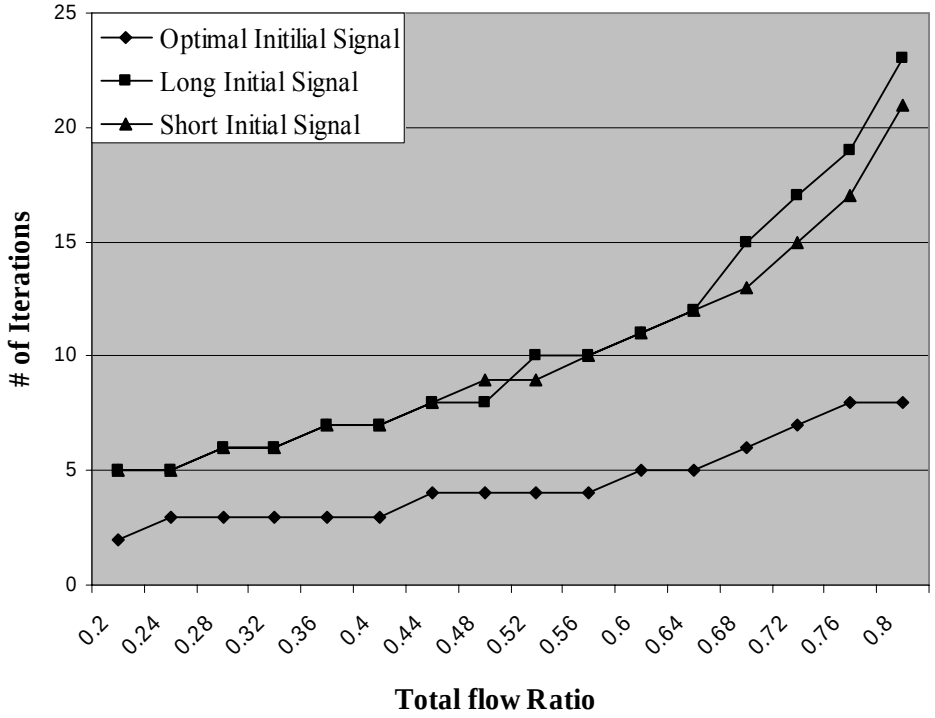


Figure 3.13 Number of iterations to converge (Case 3)

Case 4: Asymmetric intersection with equal arrival rates

In this last case, the departure rates in the two directions are same as those of the case 3. The arrival rates in the two directions are equal and range from 0.06 to 0.28.

Table 3.4 gives the average numbers of vehicles served during one cycle and average half cycle lengths in N-S direction given by analytical and simulation models.

Flow ratio of N-S	Number of vehicles served		Half cycle length	
	Analytical	Simulation	Analytical	Simulation
0.12	0.59	0.59	5.17	5.18
0.16	0.84	0.84	5.68	5.68
0.2	1.14	1.15	6.29	6.29
0.24	1.50	1.50	7.00	7.00
0.28	1.93	1.93	7.86	7.87
0.32	2.46	2.46	8.92	8.91
0.36	3.13	3.14	10.26	10.29
0.4	4.00	3.97	12.00	11.95
0.44	5.17	5.16	14.35	14.32
0.48	6.86	6.86	17.71	17.72
0.52	9.45	9.47	22.91	22.95
0.56	14.00	13.99	32.00	31.97

Table 3.4 Comparison of average numbers of vehicles served in one cycle and average half cycle lengths of N-S direction (Case 4)

The average delays in the two directions and for the whole intersection are given in Figures 3.14, 3.15 and 3.16. The numbers of iterations needed for convergence for

difference flow ratios and initial signal cycles are given in Figure 3.17. Again, the analytical results match well with the simulation results and the numerical solution of the analytical model converges quickly.

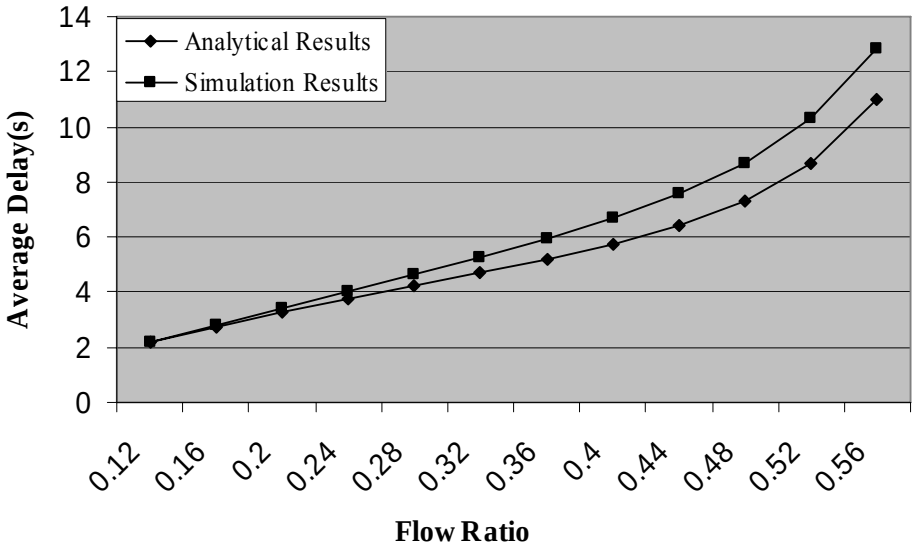


Figure 3.14 Average delay of N-S direction (Case 4)

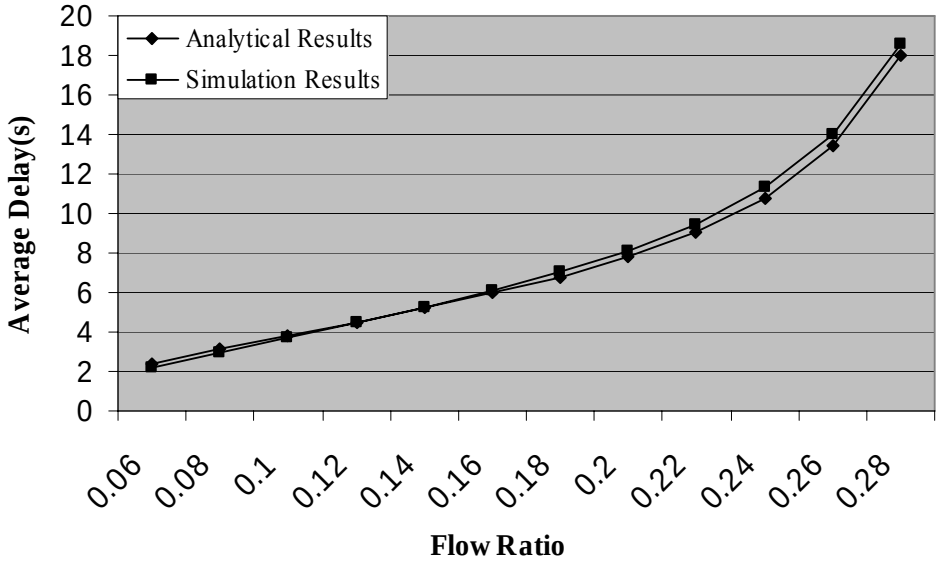


Figure 3.15 Average delay of E-W direction (Case 4)

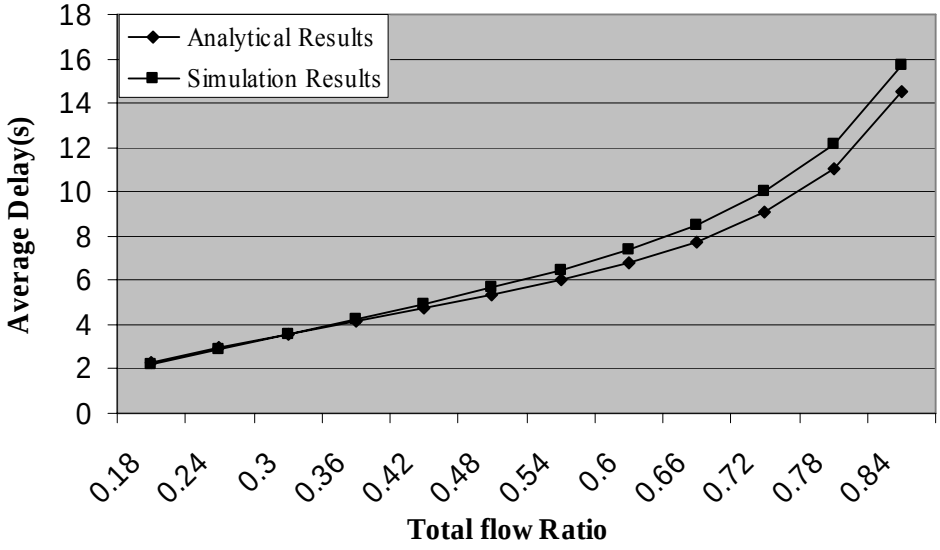


Figure 3.16 Average intersection delay (Case 4)

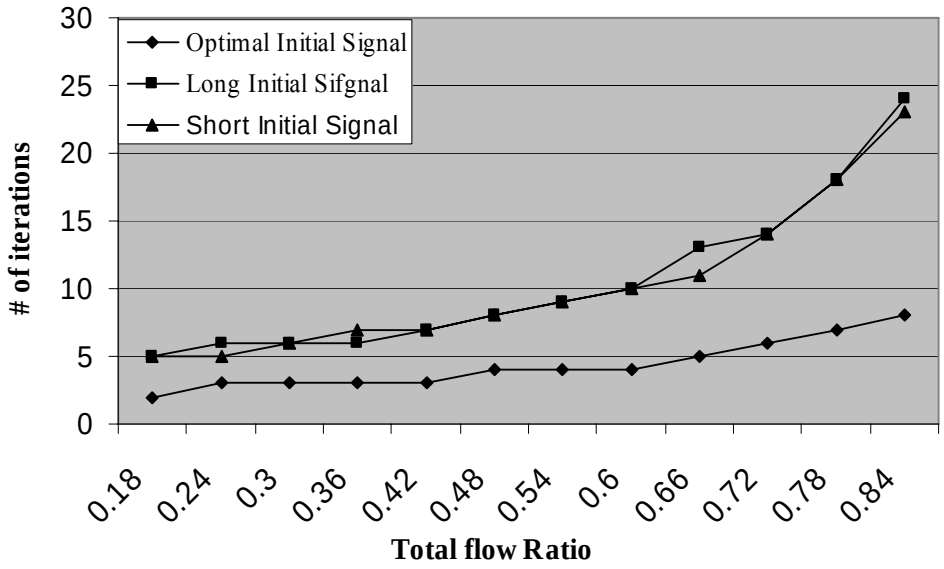


Figure 3.17 Number of iterations to converge (Case 4)

Summarizing the evaluation of the queuing model of the four different cases of intersection loading, we conclude that:

(1) The numerical solution of the queuing model estimates very well the average number of vehicles served in each cycle and average half cycle length of the strategy.

(2) When intersection total flow ratio is lower than 0.6, the estimations of average delays given by the analytical model are very close to the simulation-based results. The queuing model underestimates the average delay slightly when the intersection flow ratio is higher than 0.60.

(3) When the initial cycle lengths are set at the optimal ones for deterministic arrivals, the numerical method converges very quickly, in less than 10 iterations. With other initial cycle lengths, say, twice or half these optimal

lengths, the model still converges, but at a slower rate. Also, in general, number of more iteration is needed when the flow ratio is higher.

3.5 Summary

This chapter develops an analytical queuing model to analyze the queues and delays at an isolated 2-phase intersection control that uses a simple adaptive signal strategy originally proposed by Newell [36]. In this strategy, the first movement is served until the queue dissipates, then the second movement is served until the queue dissipates, then the signal goes back to serve the first movement and this cyclic process repeats. A queuing model and a corresponding numerical solution procedure was developed to compute steady-state performance measures such as average delays and expected queue lengths.

The numerical solution of the model was compared with results from a simulation model. The comparison shows that the analytical model approximates very well the simulation results when the intersection total flow ratio is less than 0.60. The model slightly underestimates the average delays when flow ratio increases.

In Chapter 4, a stochastic queuing model for an intersection operating under fixed signal timing is developed. The results are compared with the results for adaptive control case presented in Chapter 3. The major complication in modeling fixed signal timings is the consideration of residual queues, when they occur, from one cycle to

another (In our adaptive strategy there are no residual queues). We will also use this modeling methodology to model an adaptive control strategy for a four-phase intersection control, where the signal switches to the next phase after queues being served in the current phase are depleted.

CHAPTER 4 QUEUING MODELING OF FIXED-TIME SIGNAL CONTROL

4.1 Two-way traffic with fixed-time signal control

In Chapter 4 we consider the same intersection of Chapter 3 but now with fixed-time traffic signal. Since the arrival and departure process of each direction is independent, and signal timing is constant, we only analyze one direction, say N-S, to derive expressions for queue-based performance measures. The analysis of E-W direction is similar. The average delays given by the corresponding numerical method are compared with results obtained in Chapter 3. The results show that when the arrivals are Poisson, adaptive signal always reduces the average vehicular delay compared to the fixed-time signal. And the delay is even lower than fixed-time control with uniform traffic streams.

Figure 4.1 illustrates the phasing for fixed-time control, where phase Φ_1, serves W-E direction vehicles and half cycle is denoted by a fixed c_1, and phase Φ_2 serves N-S vehicles and half cycle is denoted by a fixed c_2.

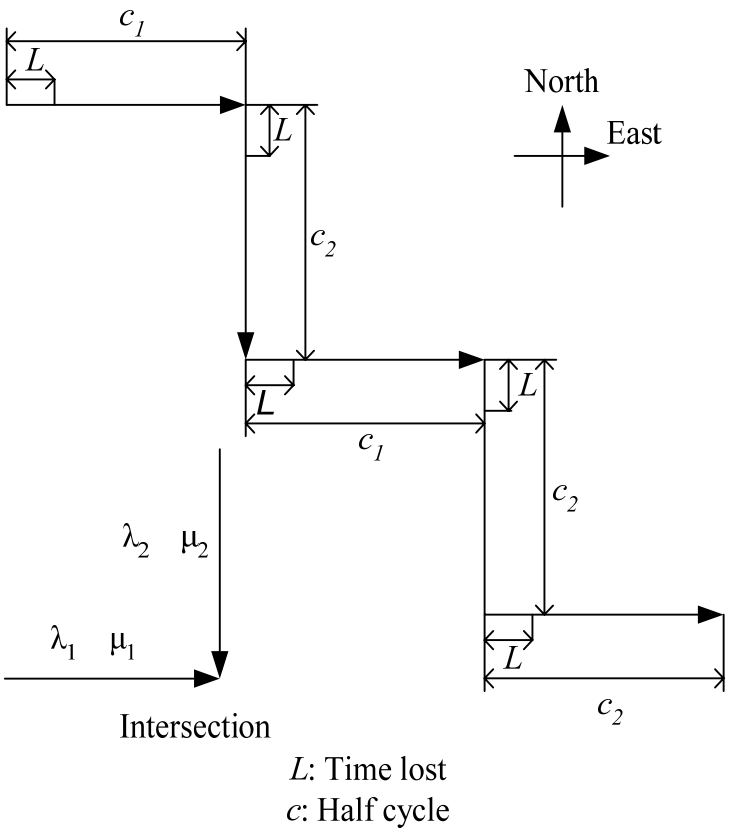


Figure 4.1 Traffic signal of the intersection

4.2 Queuing model for fixed-time control

The optimal half cycle length for deterministic constant arrival rates, derived in Chapter 3 (see (3.14)), cannot be used directly for Poisson arrivals, because when the cyclic capacity is equal to the average traffic demand (i.e., ρ = 1), queues due to random arrivals tend to infinity. Therefore to compute optimal half cycles, we simply introduce a multiplier $\beta > 1$ for both c_1 and c_2, which is obtained by (3.14a)- (3.14b).

$$c_1 = \beta \frac{1+\rho_1-\rho_2}{1-\rho_1-\rho_2} L \quad \beta > 1.0\,, \tag{4.1a}$$

$$c_2 = \beta \frac{1-\rho_1+\rho_2}{1-\rho_1-\rho_2} L \quad \beta > 1.0 . \tag{4.1b}$$

Appropriate multiplier values make the queue length finite and system stable. The new signal setting is given in (4.1). If the β is large, say greater than 3, the average delay begins to increase dramatically because very long green phases waste the effective system capacity. If β is too small queue delays become excessive. To find the optimal β, we simply search a β value between 1.01 to 3.00 to minimize the average delay; where the average delay for any c_1, c_2 is determined by the numerical approach described below.

We assume vehicles arrive at the intersection in a Poisson manner and there is no residual queue on N-S direction at the beginning of the first half cycle c_1, which also serves N-S direction. Let X_i denote the number of N-S arrivals in the i^{th} cycle, and X_i' denote the total number of vehicles requesting N-S green time in the i^{th} cycle (obviously $X_i = X_i'$ if there is no residual queue from the $(i\text{-}1)^{th}$ vehicle). Poisson arrival pattern implies

$$P_{k_i} = P(X_i = k_i) = \frac{N^{k_i}}{k_i!} e^{-N} \; for \quad i = 0,1,2\cdots, \tag{4.2}$$

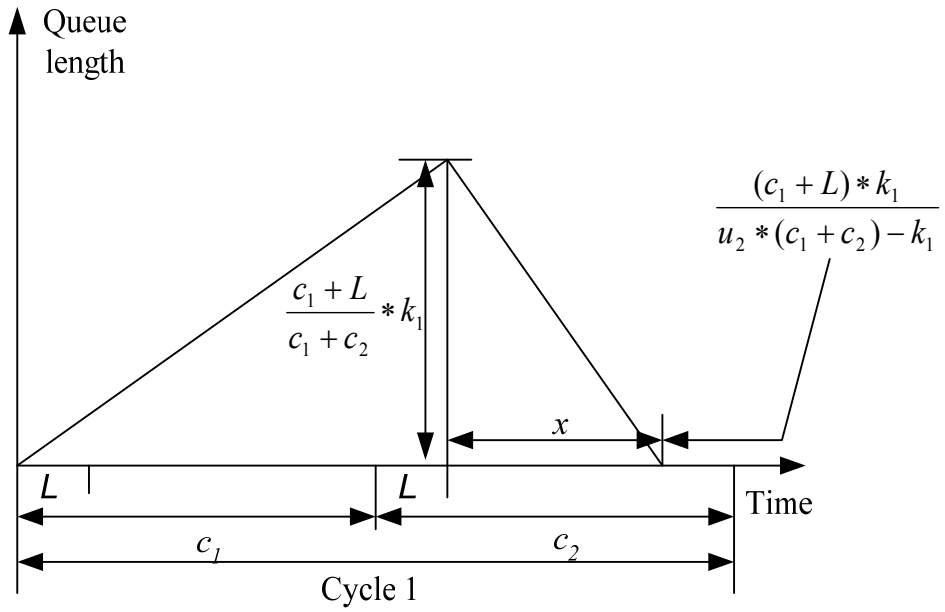
where $N = \lambda_2 (c_1 + c_2)$, which is the expected number of N-S arrivals in one cycle.

The cyclic capacity, or the maximum number of arrivals that can be served by the intersection during one cycle, in the N-S direction, is

$$N' = u_2(c_2 - L)\,. \tag{4.3}$$

Obviously we must have $N'> N$ if the system is to be stable.

There are two possible cases on what may happen at the end of cycle 1.



L: Time lost
c: Half cycle

Figure 4.2 Queue length plot of the first cycle in case A

Case A: There is no residual queue at the end of the first cycle when $X_1 \le N'$ (see Figure 4.2);

Case B: There is a residual queue (see Figure 4.3).

The maximal queue occurs at the beginning of the green phase (see, e.g, Figure 4.2). Given k_1, N-S arrivals in half cycle c_1, the expected maximal queue in this cycle is

$$Q_{\max}^{(1)} = k_1 \frac{c_1 + L}{c_1 + c_2}. \tag{4.4}$$

Case A:

Assume it takes x seconds to clear the queue in the first green phase, that is, the intersection serves N-S vehicles with rate u_2 in the interval $[(c_1 + L), (c_1 + L + x)]$. Given k_1 arrivals in cycle 1 (i.e., during c_1+c_2), the arrival in the above time interval, $k_1(c_1 + L + x)/(c_1 + c_2)$, are served at rate μ_2 for duration x until queue clears. That is,

$$k_1 \frac{(c_1 + L) + x}{c_1 + c_2} = u_2 x,$$

Solving for x, we get

$$x = \frac{k_1(c_1 + L)}{u_2(c_1 + c_2) - k_1}. \tag{4.5}$$

In this case, the conditional expected total delay, D_1^A, given k_1, can be viewed as the area of the triangle in Figure 4.2

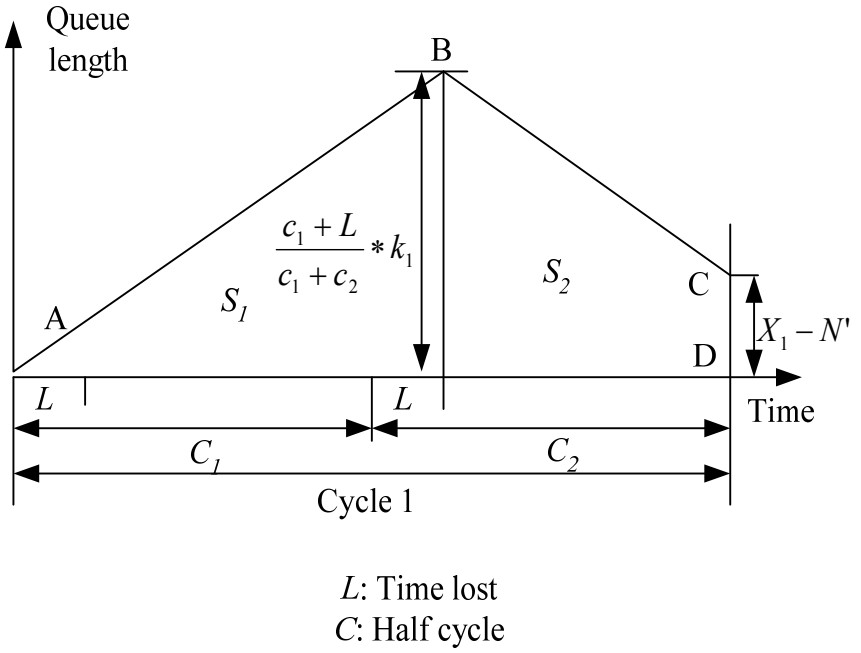
$$\begin{aligned} D_1^A &= \frac{k_1}{2}[c_1 + L + \frac{k_1(c_1 + L)}{u_2(c_1 + c_2) - k_1}]\frac{c_1 + L}{c_1 + c_2} \\ &= \frac{k_1(c_1 + L)^2}{2(c_1 + c_2)}[1 + \frac{k_1}{u_2(c_1 + c_2) - k_1}]. \end{aligned} \tag{4.6}$$

Hence, the conditional average delay, d_1^A, is

$$d_1^A = D_1^A / k_1 = \frac{(c_1 + L)^2}{2(c_1 + c_2)}[1 + \frac{k_1}{u_2(c_1 + c_2) - k_1}]. \quad (4.7)$$

Case B:

When a residual queue exists at the end of the cycle with size $(X_1 - N')$ (see Figure 4.3), the conditional expected total delay, D_1^B, of the served N-S vehicles can be viewed as the area of the polygon ABCD in Figure 4.3.



L: Time lost
C: Half cycle

Figure 4.3 Queue length curve of the first cycle in case B

The area of S_1 and S_2 can be calculated directly

$$S_1 = \frac{k_1}{2}(c_1 + L)\frac{c_1 + L}{c_1 + c_2}, \quad (4.8a)$$

$$S_2 = \frac{1}{2}(c_2 - L)[(k_1 - N') + \frac{k_1(c_1 + L)}{c_1 + c_2}]. \tag{4.8b}$$

The total delay of the served vehicles is

$$\begin{aligned} D_1^B &= S_1 + S_2 \\ &= \frac{k_1}{2}(c_1 + L) + \frac{1}{2}(c_2 - L)(k_1 - N'). \end{aligned} \tag{4.8c}$$

Then given k_1, the conditional average delay, d_1^B, of the served N-S vehicles is

$$d_1^B = D_1^B / k_1 = \frac{1}{2}(c_1 + L) + \frac{1}{2k_1}(c_2 - L)(k_1 - N') \tag{4.9}$$

Average delay for first cycle

Let P_1^A be the probability for case A, and P_1^B be the probability for case B, then

$$P_1^A = P(X_1 \le N') = \sum_{k_1=N'+1}^{\infty} \frac{N^{k_1}}{k_1!} e^{-N}, \tag{4.10a}$$

$$P_1^B = P(X_1 > N') = \sum_{k_1=N'+1}^{\infty} \frac{N^{k_1}}{k_1!} e^{-N}. \tag{4.10b}$$

The residual queue, R_1, has the following distribution

$$P(R_1 = i) = \frac{N^{N'+i}}{(N'+i)!} e^{-N} \quad for \quad i = 1,2,3\cdots. \tag{4.11}$$

Therefore, the expected average delay of the served vehicles, $E(d_1)$, during the first cycle is

$$E(d_1) = \sum_{k_1=1}^{N'} P_{k_1} d_1^A + \sum_{k_1=N'+1}^{\infty} P_{k_1} d_1^B$$
$$= \sum_{k_1=1}^{N'} P_{k_1} \frac{(c_1+L)^2}{2(c_1+c_2)}[1+\frac{k_1}{\mu_2(c_1+c_2)-k_1}]+ \qquad (4.12)$$
$$\sum_{k_1=N'+1}^{\infty} P_{k_1}[\frac{1}{2}(c_1+L)+\frac{1}{2k_1}(c_2-L)(k_1-N')].$$

Average delay for second cycle

If residual queue R_1= 0, the process is renewed, and cycle 2 simply repeats cycle 1. Thus we now need to consider only the case B situation R_1>0. The distribution of X_2', which is equal to R_1+X_2, can be calculated by using the following convolution formula:

$$P_{k_2'} = P(X_2' = k_2') = \sum_{i=1}^{k_2'}[P(R_1 = i)P(X_2 = k_2' - i)]$$
$$= \sum_{i=1}^{k_2'} \frac{N^{N'+i}}{(N'+i)!} e^{-N} \frac{N^{(k_2'-i)}}{(k_2'-i)!} e^{-N}. \qquad (4.13)$$

At this situation, when there is residual queue from the last cycle (say cycle 1), the probability that no residual queue exists at the end of current cycle (say cycle 2), $X_2' \le N'$, is

$$P(X_2' \le N') = \sum_{k_2'=1}^{N'} \sum_{i=1}^{k_2'}[P(R_1 = i)P(X_2 = k_2' - i)]. \qquad (4.14)$$

There are two events that lead to no residual queue at the end of cycle 2: (1) no residual queue from cycle 1, and the total arrivals in this cycle is less or equal to N'; and 2) there is a residual queue from cycle 1 and $X_2' \le N'$ happens, whose probability is

given in (4.14). The total probability of no residual queue at the end of cycle 2, P_2^A, is

$$P_2^A = P(R_2 = 0) = P_1^A(\sum_{k_2=0}^{N'} \frac{N^{k_2}}{k_2!} e^N) + P(X_2' \le N'). \tag{4.15}$$

When there is a residual queue of size i at the end of cycle 2, there are two events that could have lead to it: (1) cycle 1 has no residual queue, exact $N'+i$ vehicles arrived in this cycle, N' vehicles were served and i vehicles are left; and (2) a residual queue with size j, j in the range from 1 to ($N'+i$), was present at the end of cycle 1, and there were ($N'+i-j$) vehicles arrived in cycle 2. Total number of vehicles served in this cycle is N'. The probability of this happening is

$$\begin{aligned} &P(R_2 = i) = P_1^A P(X_2 = N' + i) + \\ &\sum_{j=1}^{N'+i} [P(R_1 = j) P(X_2 = N' + i - j)]. \end{aligned} \tag{4.16}$$

To calculate the delay of vehicles in cycle 2, consider the case where residual queue exists at the end of cycle 1. System is renewed when there is no residual queue and the results of cycle one can be used directly. Let i denote the size of residual queue at the end of cycle 1 and k_2 denote the number of vehicles arrivals in cycle 2. The probability of this to happen is $[P(R_1=i)P(X_2=k_2)]$.

The maximal queue $Q_{\max}^{(2)}$ at the beginning of the green phase is (see Figure 4.4)

$$Q_{\max}^{(2)} = i + \frac{k_2(c_1 + L)}{c_1 + c_2}. \tag{4.17}$$

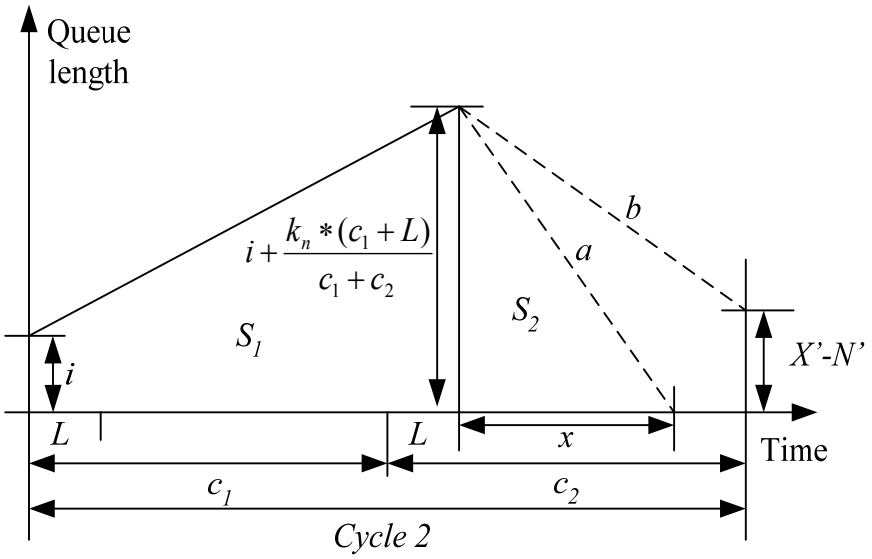


Figure 4.4 Queue length plot of the second cycle in case B

The dashed lines in Figure 4.4, *a* and *b*, represent two sub-cases, residual queue exists (for line *b*) or not (for line *a*) at the end of cycle 2. Figure 4.4 illustrates the total delay in this cycle is composed of two portions, S_1 and S_2. It is easy to get

$$S_1 = \frac{1}{2}(c_1 + L)[2i + \frac{k_2(c_1 + L)}{c_1 + c_2}]. \tag{4.18}$$

To calculate the area of S_2, we consider two sub-cases individually.

Case A_2: No residual queue at the end of cycle 2

If $X_2' \leq N'$, no residual queue at the end of cycle 2 (line *a* in Figure 4.4), S_2 is a triangle. The expected queue dissipating point *x* in this green phase is given by

$$i + \frac{(c_1 + L) + x}{c_1 + c_2} k_2 = \mu_2 x \Rightarrow x = \frac{k_2(c_1 + L) + i(c_1 + c_2)}{\mu_2(c_1 + c_2) - k_2}. \tag{4.19}$$

The area of S_2 in this sub-case, say S_2^A, is

$$S_2^A = \frac{1}{2} x Q_{\max}^{(2)} = \frac{[i(c_1 + c_2) + k_2(c_1 + L)]^2}{2(c_1 + c_2)[\mu_2(c_1 + c_2) - k_2]}. \tag{4.20}$$

The total delay, D_2^A, is

$$\begin{aligned} D_2^A &= S_1 + S_2^A = \frac{1}{2}(c_1 + L)[2i + \frac{k_2(c_1 + L)}{c_1 + c_2}] \\ &+ \frac{[i(c_1 + c_2) + k_2(c_1 + L)]^2}{2(c_1 + c_2)[\mu_2(c_1 + c_2) - k_2]}. \end{aligned} \tag{4.21}$$

The average delay, d_2^A, is

$$d_2^A = D_2^A / (i + k_2) = D_2^A / k_2'. \tag{4.22}$$

Case B_2: Residual queue at the end of cycle 2

If $X_2'>N'$, residual queue exists at the end of cycle 2 (line *b* in Figure 4.4), S_2 is a trapezium. The residual queue size is X_2'-N', the area of trapezium S_2, S_2^B, is

$$\begin{aligned} S_2^B &= \frac{1}{2}(c_2 - L)[Q_{\max}^{(2)} + (i + k_2 - N')] \\ &= \frac{1}{2}(c_2 - L)[2i + k_2(1 + \frac{c_1 + L}{c_1 + c_2}) - N')]. \end{aligned} \tag{4.23}$$

The total delay in this sub-case is

$$D_2^B = S_1 + S_2^B = \frac{1}{2}(c_1 + L)[2i + \frac{k_2(c_1 + L)}{c_1 + c_2}] + \frac{1}{2}(c_2 - L)[2i + k_2(1 + \frac{c_1 + L}{c_1 + c_2}) - N')]. \quad (4.24)$$

The average delay is d_2^B

$$d_2^B = D_2^B / (i + k_2) = D_2^B / k_2'. \quad (4.25)$$

From above we can see the expectation of delay in cycle 2 is composed of three parts: (1) no residual queue at the end of cycle 1, the expected delay is equal to that of cycle 1; (2) there is residual queue at the end of cycle 1, and no residual queue at the end of cycle 2, which is sub case A; and (3) there is residual queue from cycle 1, and a residual queue also at the end of cycle 2, which is sub case B. Combine these, the average delay during the second cycle has the expectation of

$$E(d_2) = P_1^A E(d_1) + \sum_{k_2'=1}^{N'} \sum_{i=1}^{k_2'} [P(R_1 = i) P(X_2 = k_2' - i) d_2^A] + \sum_{k_2'=N'+1}^{\infty} \sum_{i=1}^{k_2'} [P(R_1 = i) P(X_2 = k_2' - i) d_2^B]. \quad (4.26)$$

(4.26) gives the expected average delay during cycle 2. And we also get the distribution of the residual queue at the end of cycle 2. Some other measures of performance can also be calculated, such as the average queue length during cycle 2, the residual queue size at the end of cycle 2 and the number of vehicles being served in this cycle.

With the residual queue distribution at the end of cycle 2, we can analyze cycle 3, then cycle 4 and so on. This iteration continues until a steady state of the system is reached.

Average delay for n^{th} cycle

The following equations give the general formula for the analyses of the n^{th} cycle. Assume we know the residual queue distribution at the end of $(n\text{-}1)^{th}$ cycle; then let:

$P^{A}_{n\text{-}1}$ = probability {no residual queue of cycle $(n\text{-}1)$}, and

$P(R_{n\text{-}1}=i)$ = probability {the residual queue of cycle $(n\text{-}1) = i$}.

If there is no residual queue from $(n\text{-}1)^{th}$ cycle), the n^{th} cycle simply repeats cycle 1. Thus we only discuss the event with queue from $(n\text{-}1)^{th}$ cycle.

With $X_n'=X_n+R_{n\text{-}1}$, the distribution of X_n' is

$$P_{k_n'} = P(X_n' = k_n') = \sum_{i=1}^{k_n'} [P(R_{n-1} = i)P(X_n = k_n' - i)]. \tag{4.27}$$

The probability that there is no residual queue exists at the end of n^{th} cycle, given a residual queue from $(n\text{-}1)^{th}$ cycle, $X_n' \le N'$, is

$$P(X_n' \le N') = \sum_{k_n'=1}^{N'} \sum_{i=1}^{k_n'} [P(R_{n-1} = i)P(X_n = k_n' - i)]. \tag{4.28}$$

Probability that there is no residual queue at the end of n^{th} cycle, given no residual

queue from (n-1)th cycle, is simply $P_{n-1}^{A}(\sum_{k_n=1}^{N'}\frac{N^{k_n}}{k_n!}e^{N})$

The unconditional probability that no residual queue exists at the end of the n^{th} cycle, P_n^{A} is

$$P_n^{A} = P(R_n = 0) = P_{n-1}^{A}(\sum_{k_n=1}^{N'}\frac{N^{k_n}}{k_n!}e^{N}) + P(X_n' \le N') . \tag{4.29a}$$

The probability that residual queue exists with size i at the end of the n^{th} cycle, $P(R_n=i)$ is

$$P(R_n = i) = P_{n-1}^{A}P(X_n = N'+i) + \sum_{j=1}^{N'+i}[P(R_{n-1} = j)P(X_n = N' + i - j)]. \tag{4.29b}$$

Let i denote the residual queue size at the end of the (n-1)th cycle and k_n the number of vehicle arrivals in the n^{th} cycle, which occurs with probability [$P(R_{n-1}=i)$*$P(X_n=k_n)$]. The maximal queue Q^{n}_{max} at the beginning of the green phase in n^{th} cycle is

$$Q_{\max}^{(n)} = i + \frac{k_n(c_1 + L)}{c_1 + c_2}. \tag{4.30}$$

Figure 4.4 illustrates the total delay in this cycle is composed of two portions, S_1 and S_2. S_1 depends on i and k_n. Using same analysis as for cycle 2, we get

$$S_1 = \frac{1}{2}(c_1 + L)[2i + \frac{k_n(c_1 + L)}{c_1 + c_2}] . \tag{4.31}$$

To calculate the area of S_2, we consider again two sub-cases,

Case A_n: No residual queue at the end of cycle n

If $X_n'=R_{n-1}+X_n \leq N'$, there is no residual queue at the end of the n^{th} cycle (line *a* in Figure 4.4), S_2 is a triangle. The time needed to clear the queue in this green phase is

$$x = \frac{k_n(c_1 + L) + i(c_1 + c_2)}{\mu_2(c_1 + c_2) - k_n}. \tag{4.32}$$

The area of S_2 in this sub-case, S_2^A, is

$$S_2^A = \frac{1}{2} x Q_{\max}^n = \frac{[i(c_1 + c_2) + k_n(c_1 + L)]^2}{2(c_1 + c_2)[\mu_2(c_1 + c_2) - k_n]} . \tag{4.33}$$

The total delay of this sub-case, D_n^A, is

$$D_n^A = S_1 + S_2^A = \frac{1}{2}(c_1 + L)[2i + \frac{k_n(c_1 + L)}{c_1 + c_2}] + \frac{[i(c_1 + c_2) + k_n(c_1 + L)]^2}{2(c_1 + c_2)[\mu_2 * (c_1 + c_2) - k_n]}. \tag{4.34}$$

whose expectation is

$$d_n^A = D_n^A / (i + k_n) = D_n^A / k_n' . \tag{4.35}$$

Case B_n: Residual queue at the end of cycle n

If $X_n'=R_{n-1}+X_n > N'$, residual queue exists at the end of the n^{th} cycle (line *b* in Figure 4.4). S_2 is a trapezium. The residual queue size is $X_n'-N'$, the area of S_2 in this sub-case,

S_2^B, is

$$S_2^B = \frac{1}{2}(c_2 - L)[Q_{\max}^n + (i + k_n - N')]$$
$$= \frac{1}{2}(c_2 - L)[2i + k_n(1 + \frac{c_1 + L}{c_1 + c_2}) - N')]. \tag{4.36}$$

The total delay of sub-case B, D_n^B is

$$D_n^B = S_1 + S_2^B = \frac{1}{2}(c_1 + L)[2i + \frac{k_n(c_1 + L)}{c_1 + c_2}]$$
$$+ \frac{1}{2}(c_2 - L)[2i + k_n(1 + \frac{c_1 + L}{c_1 + c_2}) - N')], \tag{4.37}$$

whose expectation is

$$d_n^B = D_n^B / (i + k_n) = D_n^B / k_n'. \tag{4.38}$$

Using the analysis as for cycle 2, the expected average delay, $E(d_n)$, is found to be

$$E(d_n) = P_{n-1}^A E(d_1) + \sum_{k_n'=1}^{N'} \sum_{i=1}^{k_n'} [P(R_{n-1} = i) P(X_n = k_n' - i) d_n^A]$$
$$+ \sum_{k_n'=N'+1}^{\infty} \sum_{i=1}^{k_n'} [P(R_{n-1} = i) P(X_n = k_n' - i) d_n^B]. \tag{4.39}$$

The expectation of average vehicular delay during the n^{th} cycle is given by (4.38). The residual queue distribution at the end of this cycle is also given in (4.28), which will be used in the computation of the $(n+1)^{\text{th}}$ cycle.

4.3 Model evaluation

In this section, the results of the proposed numerical method are presented, and the average vehicular delays for fixed-time and adaptive control schemes are compared. For comparison purposes, the delays with same rates but uniform arrivals (deterministic) are calculated and presented. The delays of Webster's formula are also given. The results show that with Poisson arrivals, the average vehicular delay can always be reduced with adaptive signal timing compared to the fixed-time signal.

In this analysis the numerical values of delay and other performance measures are calculated cycle by cycle until the steady state is reached.

If vehicle arrivals are uniform, the optimal signal setting is given by (3.14). And queue length plot is given by Figure 4.5. Average delay can be calculated analytically. Equations (4.40) to (4.42) give the calculation of N-S direction.

$$k_2 = \lambda_2(c_1 + L), \tag{4.40}$$

$$D_u = \frac{1}{2} k_2(c_1 + c_2), \tag{4.41}$$

$$d_u = D_u / [\lambda_2(c_1 + c_2)]. \tag{4.42}$$

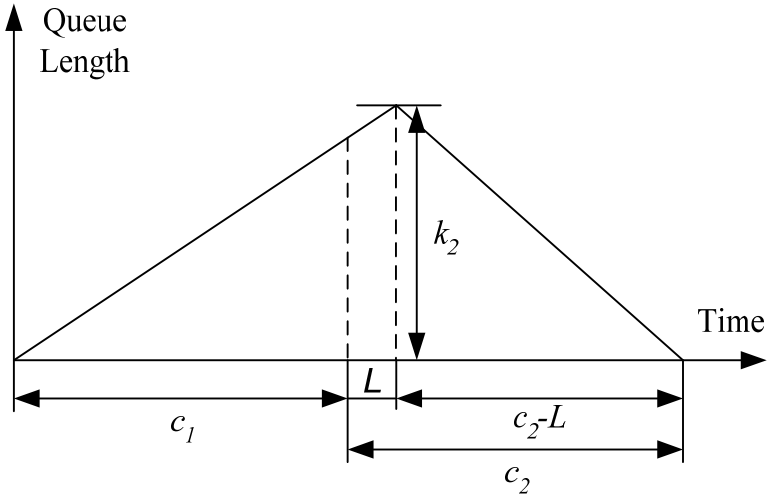


Figure 4.5 Queue length curve of in one cycle

Case 1: Symmetric intersection with equal arrival rates

In this case, the saturation flow rates in the two directions are 0.5veh/sec, the arrival rates of two traffic streams are equal, and range from 0.05veh/sec to 0.2veh/sec. Time lost is 4 seconds for all the following cases.

Figure 4.6 shows the adaptive control always reduces the average vehicular delay compared with fixed-time control. Its performance is even better than the case with uniform arrivals. This is can be explained as following. With Poisson arrival, there is a possibility of no arrival during the red phase, which means delay is zero, this will not happen with uniform arrival since then there are same number of arrivals in each phase. The figure also shows that lower the arrival rate (the possibility of no arrival for adaptive control during the red phase is higher), larger the difference between these two systems. As the arrival rate increases, the delay plots of these two systems

converge eventually. The results imply the overall performance of the adaptive signal control system with Poisson arrivals (very random) is slightly better than the system with uniform arrivals and optimized fixed-time control. The delay obtained by proposed queuing model with fixed-time control matches results of Webster's formula fairly well.

Figure 4.7 gives the percentage of average delay reduction with adaptive signal vs. fixed-time signal for Poisson arrivals. As the degree of saturation of the intersection increases from 0.20 to 0.80, the adaptive signal can reduce the average delay from 35% to 60% comparing with the fixed-time signal.

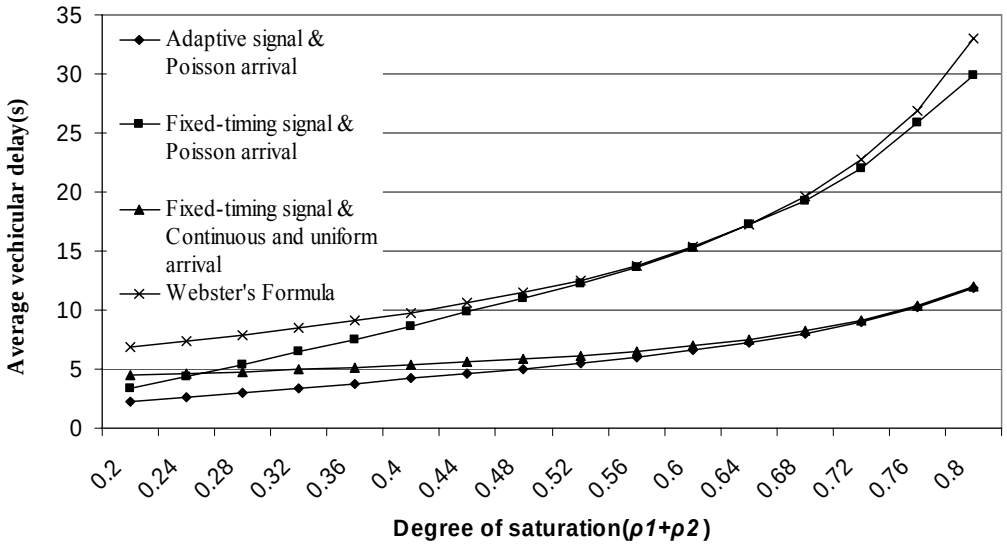


Figure 4.6 Delay of symmetric intersection (Case 1)

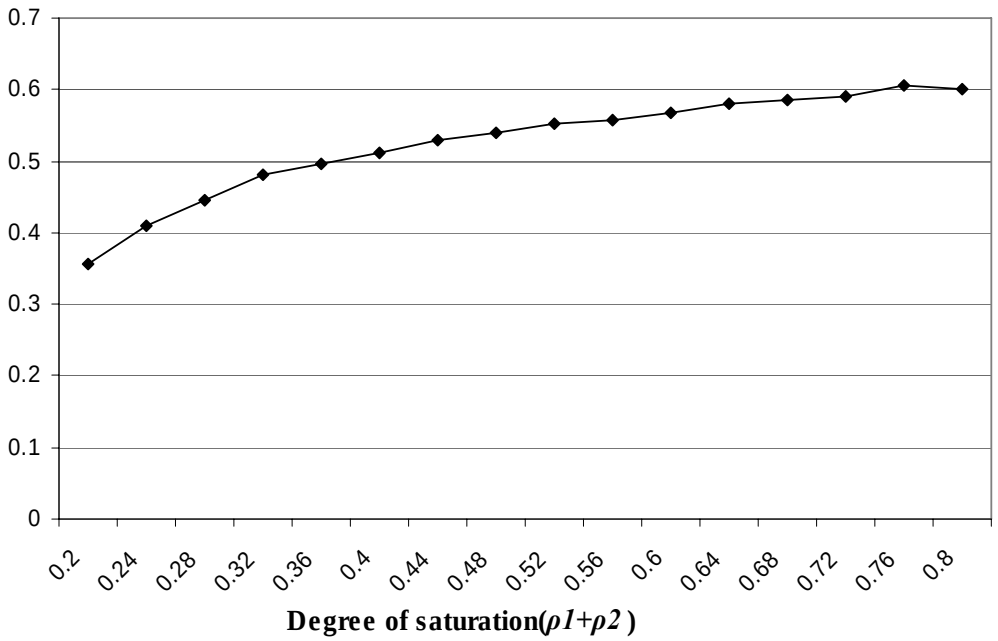


Figure 4.7 Percentage of average delay reduction (Case 1)

Case 2: Symmetric intersection with unequal arrival rates

In case 2, the departure rates of two directions are 0.5veh/sec. The arrival rate of N-S direction ranges from 0.025veh/sec to 0.135veh/sec, while the arrival rate of E-W direction is doubled. Figure 4.8 and 4.9 gives very similar conclusions as for case 1.

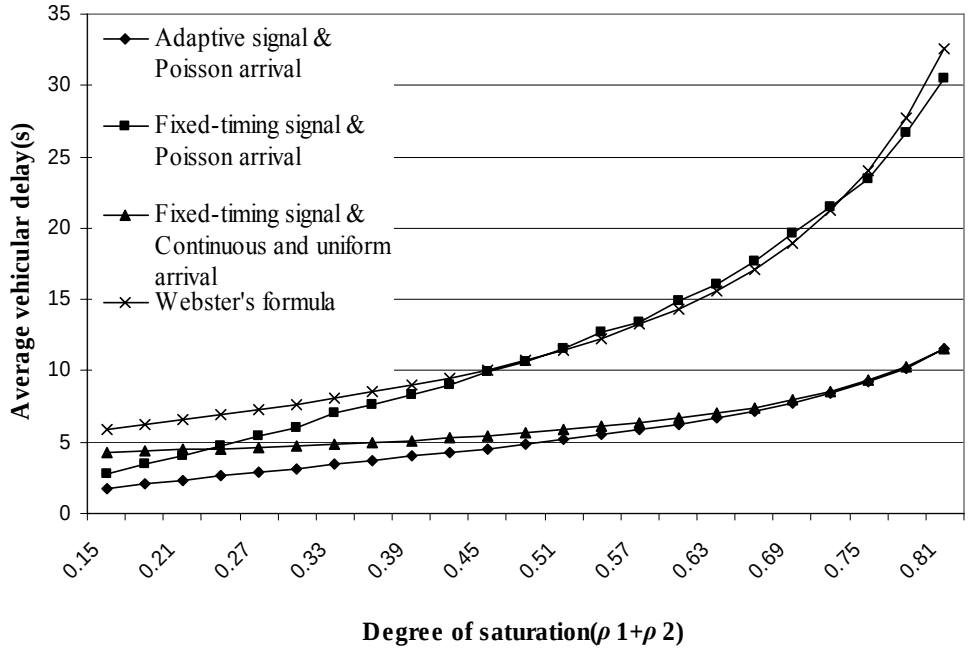


Figure 4.8 Delay of symmetric intersection with unequal arrival (Case 2)

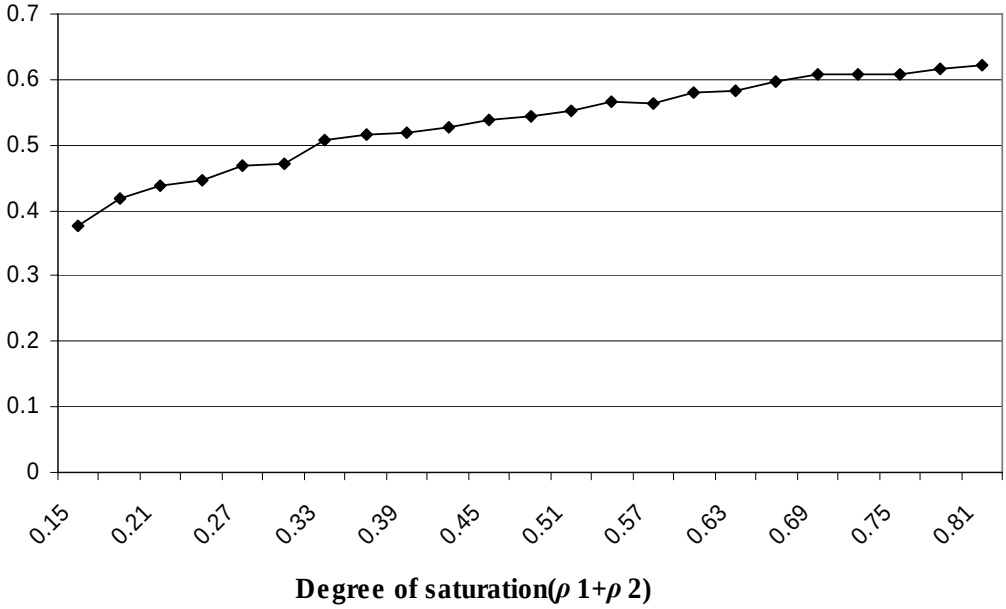


Figure 4.9 Percentage of average delay reduction (Case2)

Case 3: Asymmetric intersection with equal degree of saturation rates

In this case, the departure rate of N-S directions is 0.5veh/sec, the departure rate of E-W directions is 1.0veh/sec. The arrival rate of N-S direction ranges from 0.025veh/sec to 0.20veh/sec. The arrival rate of E-W direction is doubled (see Figure 4.10 and Figure 4.11).

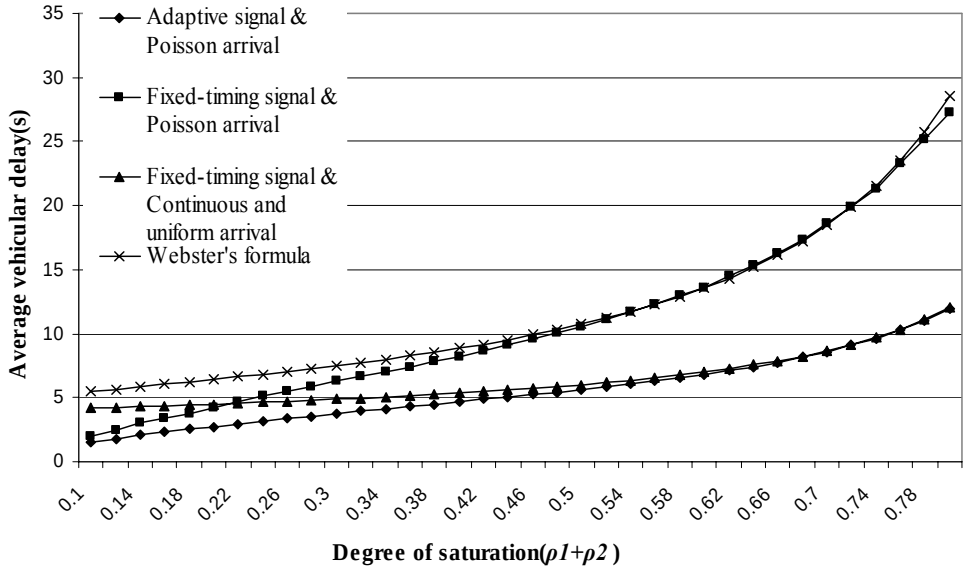


Figure 4.10 Delay of asymmetric intersection with same saturation rate (Case 3)

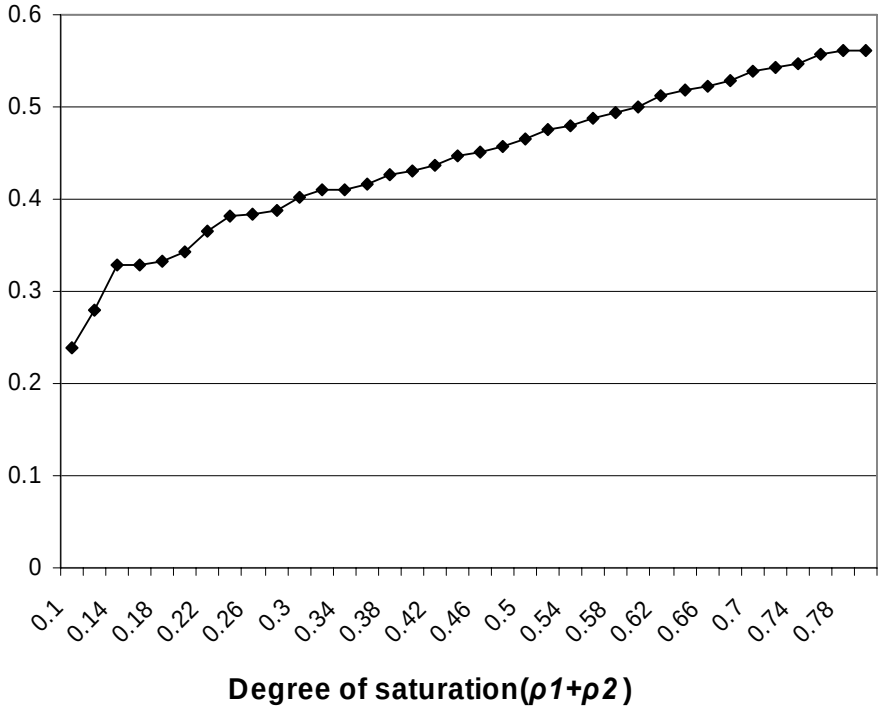


Figure 4.11 Percentage of average delay reduction (Case 3)

Case 4: Asymmetric intersection with equal arrival rates.

In this case, the departure rates of two directions are same as those of the case 3. The arrival rates of two directions are equal and increase from 0.05 to 0.27(see Figure 4.12 and Figure 4.13).

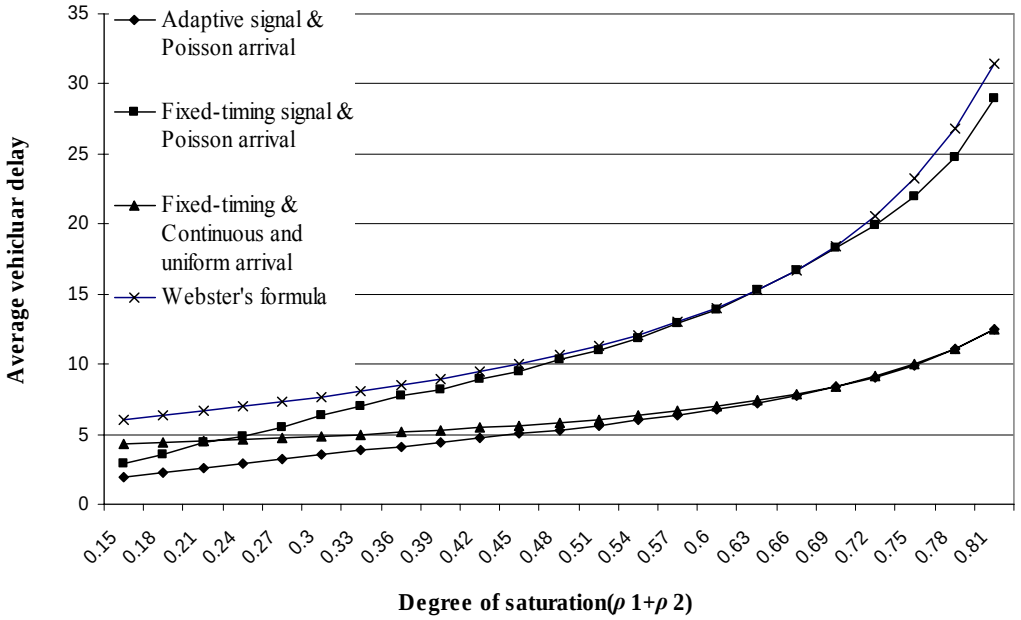


Figure 4.12 Asymmetric intersection with equal arrival (Case 4)

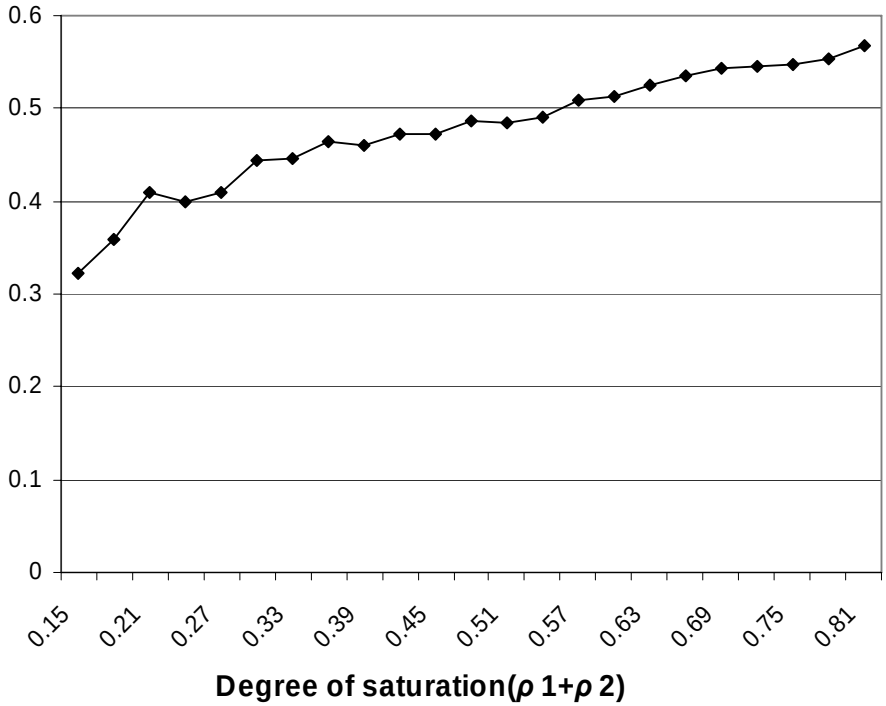


Figure 4.13 Percentage of average delay reduction (Case 4)

From the numerical results for these four cases, we conclude that when arrivals are Poisson, no matter how to set the arrival and departure rates of each approach, the adaptive signal system always reduce the average delay compared to that for the fixed-time signal system the reduction ranging from 30% to 60%, while the total degree of saturation ranges from 0.2 to 0.8.

The results also give some interesting relationship between the adaptive control with Poisson arrival and fixed-time control with uniform arrivals. With same average arrival and departure rate, the expectation of the number of vehicles been served and that of cycle length of adaptive control system is equal to the fixed-time control with uniform arrivals. And the delays of these two systems are close to each other. Actually this difference is caused by the first term of Equation (3.20). If the term (-$P(k_i = 0)$) is eliminated, there is no difference between the average delay of these two systems. This may be explained as following, if the arrival is Poisson, there is a non-zero possibility that no vehicles arrived during a red phase, which creates a zero delay and therefore reduces average delay. The lower arrival rate has a larger possibility of occurrence of zero delay in a cycle.

4.4 Summary

This chapter developed a queuing model to analyze the queues and delays at an isolated intersection operating under fixed-time signal control. Numerical solution of the model was compared with results obtained in Chapter 3 for adaptive control

system. The comparison shows that the adaptive control system always reduces the vehicular delays comparing with fixed-time control. It also show the adaptive control with Poisson arrival has a slightly lower delay comparing with fixed-time control with deterministic constant arrival traffic stream.

A major assumption, and perhaps a weakness, of the model is the assumption of Poisson arrivals. We plan to investigate if such a model can be used for other arrival distributions, in particular arrival pattern that approximate batch (platoon) arrivals

Chapter 5 investigates the use of such a modeling methodology to a more general adaptive control strategy at an intersection operating under a standard eight-phase actuated control.

CHAPTER 5 A QUEUING MODEL FOR ACTUATED SIGNAL CONTROL

5.1 Introduction

In this chapter the stochastic queuing modeling method is further developed to analyze an eight-phase actuated signal control system. First, an algorithm is proposed to calculate the green phase distribution of an actuated signal control, (whereas the methodology used in section 3.2 gives only the expectation of the green phase length in adaptive control) then we develop a stochastic queuing model for actuated signal control system. The algorithm can be used to compute steady–state performance measures such as average delays and expected queue lengths.

The four-leg intersection under consideration has two-way through traffic streams and left turn movements on each leg. For the convenience of delay calculation, we assume the arrival processes are Poisson.

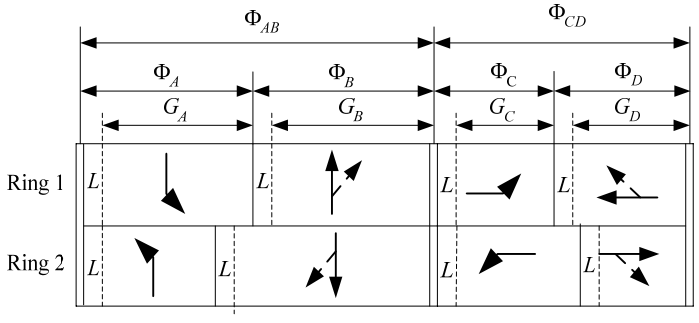


Figure 5.1 An eight-phase dual-ring actuated control

The actuated traffic signal shown in Figure 5.1 has eight phases and two rings. The cycle length is deterministic with two pre-timed barriers (demonstrated by double lines). Each phase consists of a constant time lost L plus an effective green time (for convenience purpose, we use "G" rather than "g" to denote effective green time, for example, $\Phi_A=L+G_A$). The rings and barriers divide eight movements into four independent pairs. For example, movement A and B are one pair. Phase serving each movement only correlates to that of other movement in the same group. The left turn phase, say Φ_A, is actuated and has an upper bound, say $\Phi_{A\max}$. Φ_A may end early before reaching the upper bound if the queue dissipates early; otherwise the traffic signal switches at $\Phi_{A\max}$. Because of the pre–timed barrier, $(\Phi_A + \Phi_B)$ is a constant; similarly $(\Phi_C + \Phi_D)$ is constant. Because of independency between each pair, analysis of each pair of movements is the same. Movements A and B, southbound left turn and northbound through movements are selected to develop the model. By calculating the probability distribution of each possible scenario, the proposed queuing model can analyze the performance measures of the traffic control system at steady state.

5.2 Busy Period Distribution Analysis

In section 3.2, the expectation of green phase duration, or the conditional "busy period" of the adaptive control strategy, given multiple vehicles waiting at the beginning of the green phase, is calculated. The transforms based method in Section 3.2 gave the expectation rather than the distribution of the green phase. This explains

the minor discrepancy between the delays given by the queuing model and those given by the simulation. In this chapter we first propose an algorithm to calculate the distribution of the green phase duration, and then develop the intersection queuing model.

The busy period is the actuated green phase needed to clear the existing queue, same as that in Chapter 3.

Assume vehicle arrival process is Poisson with rate λ, and the intersection service rate (or the saturation flow rate) is a constant μ.

The queue clearing process is a multi–stage stochastic process. Figure 5.2 shows the scenario tree of the process. Assuming there are k_0 vehicles waiting at the beginning of green phase, the first stage is the service time for those k_0 vehicles, whose duration is k_0/μ. We use X_1 to denote the number of arrivals during first stage, which follows the following distribution because the arrival Process is Poisson:

$$P_{k_1} = P(X_1 = k_1) = \frac{(\lambda k_0 / \mu)^{k_1}}{k_1!} e^{-\lambda k_0 / \mu}. \tag{5.1a}$$

Use ρ, flow ratio to denote λ/μ, (5.1a) can be written as

$$P_{k_1} = P(X_1 = k_1) = \frac{(k_0 \rho)^{k_1}}{k_1!} e^{-k_0 \rho}. \tag{5.1b}$$

The second stage serves those X_1 vehicles, whose duration is linearly related to X_1, a

random variable with known distribution.

X_1= 0 means the busy period ends at k_0/μ, whose probability is given by $P(X_1= 0)$. Each realization of X_1 is a child node of the root node in the scenario tree. The scenario with $X_1>0$ also has its own offspring too, analysis of the succeeding stages is similar to the analysis of the first stage.

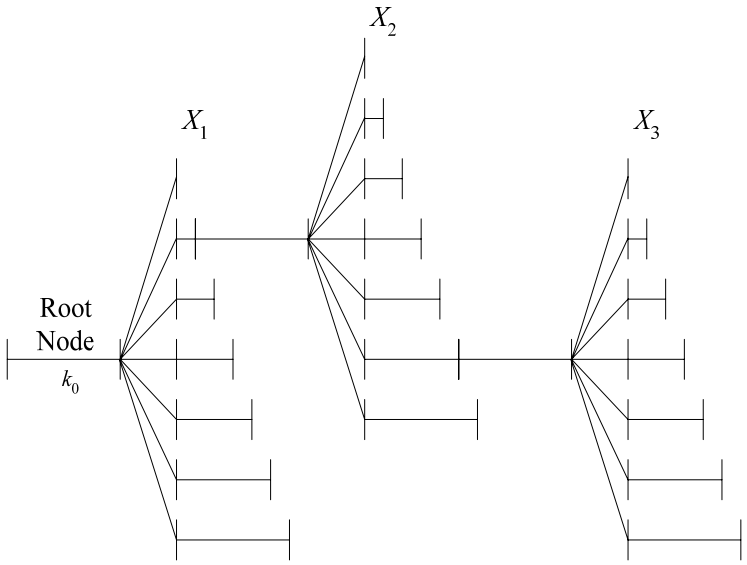


Figure 5.2 Scenario tree of queue clearing process

When we consider the distribution of the busy period, we only care the total number of vehicles being served, not the detailed arrival process described above. For example, if there were n vehicles waiting at the beginning of the green phase, and the busy period was $(n+2)/\mu$, means that two more vehicles arrived and being served during the green phase. These two vehicles may have arrived during stage one, the service time of the first n vehicles. While there were no new arrivals during stage two,

the service time of the first two new arrivals. Another possible scenario is that only one vehicle arrived during stage one, the other vehicle arrived during the stage two and no new arrivals after that. As the total number of served vehicles increases, the total number of possible scenarios increases dramatically with the order of $o(n!)$. Nevertheless, Equation (5.1b) shows that the number of vehicles being served during the busy period is only related to the flow ratio ρ. When the service rate is known, the distribution of the busy period can be calculated from the distribution of the number of vehicles served.

Now we propose an algorithm to calculate the distribution of number of vehicles being served in the actuated green phase given a traffic intensity.

A data structure of the nodes of the scenario tree is defined. Each node represents a feasible stage, which includes the following information:

(1) Number of vehicles being served during current stage;

(2) Total number of arrived vehicles from the beginning to the current stage, which is the total number of arrived vehicles of all nodes on the unique path from the current node to the root node;

(3) Probability of reaching this node from the root node. We assume the current node serves k_i and its parent node serves k_j vehicles, which means there are k_i arrivals during the service time of k_j/μ. Let P_p to denote the known probability of reaching each parent node from the root node (the

parent node is the node located in the upper layer that is directly connected to the current node). The probability of reaching the current node is,

$$P_p \frac{(k_j \rho)^{k_j}}{k_i!} e^{-k_j \rho}. \tag{5.2}$$

(4) Number of child nodes, that is all nodes located in the next layer and directly connected to the current node;

(5) Address of the parent node;

(6) Address of the next little brother node (little brother node shares the same parent node and serves one more vehicle);

(7) Address of the first child node.

To increase the computing speed, a probability matrix is used to store some already computed component data. The first index of each element of the matrix represents the number of vehicles waiting at the beginning of the green phase; the second index is the total number of newly arrived vehicles been served. The element is the probability of this scenario. For example, element m_{22} represents the probability of 2 vehicles waiting at the beginning of the green phase and 4 vehicles in total are served during the green phase. This value can be directly used in the computation of some other scenarios. For example, there are 3 vehicles waiting at the beginning of the green phase and 7 vehicles being served in total, which means 4 new arrivals during the green phase. One possible realization of this scenario is 2 arrivals in stage 1, 2

arrivals in stage two and no new arrival during stage 3. The probability of a partial process mentioned above (two new arrivals in stage 2, which is the service time of the 2 vehicles arrived in stage 1 and no new arrival during stage 3) is equal to the probability m_{22}. Before analyzing this partial process, the algorithm checks the matrix to figure out if the corresponding probability value is available. If it is available, the algorithm uses the result directly, without further calculation, thereby decreasing the computation time.

For each specific traffic intensity value and each number of vehicles (denoted by k) waiting at the beginning of the green phase, the probability distribution of green phase length is determined by the total number of new arrivals during the whole busy period. Considering that n new arrivals are served in the green phase, the possible number of stages in the queue clearing process (busy period) ranges from 2 to (n+1). The first stage is due to the service time of the k vehicles waiting at the beginning. With a fairly large n, numerous stages need to be considered, which can be illustrated as follows. Consider a set of spaces where the first k spaces correspond to the first stage of the process. The next n spaces, with n vehicles and no empty space, are served by stages 2, 3,... The algorithm computes all possible scenarios for stage 2, 3,…, to serve these n vehicles and calculate the probability of each scenario. Summing probabilities of all feasible scenarios gives the probability that exactly (k+n) vehicles are served during the green phase. More details of the algorithm, say Algorithm A, is described below

Algorithm A: To compute probability of (k+n) vehicles served in a busy period

(1) Read the parameters, including arrival and service rate, initial queue size and total number of new arrivals.

(2) Initialize the root node, set total probability to zero.

(3) Create all feasible child nodes of the current node, where the total number of new arrivals (child nodes) is less than or equal to n. Repeat steps 3 to 5 for each child node.

(4) For each child node, update total arrived vehicles and the probability of path from the root node to the current node. Use the following two stopping rules to determine if this branch needs to be fathomed,

   (a) Use the number of newly arrived vehicles, say n_c, at the current node as the first index and (n– n_c) as the second index to check the corresponding element of the probability data matrix. If the data is available for that element, the current branch of the scenario tree is fathomed and the stored probability is used directly.

   (b) Calculate the probability to reach the current node. If it is smaller than a specified threshold, say 10^{-15}, then fathom this branch and go to step 6.

(c) Determine if the total number of new arrivals from the root node to current node is equals to *n*; if yes, then go to step 6, otherwise go to step 3.

(5) Calculate the probability of this scenario based on the number of arrivals in the current node and parent node, and add this probability to the total probability.

(6) When all feasible child nodes are fathomed, retrieve the probability values and store the total probability into the probability data matrix. For each *k*, the algorithm is iterated over increasing values of *n*. As *n* increases, the probability data matrix from (*n*-1) is used directly in the calculations.

5.3 Queuing model for eight–phase actuated signal control

5.3.1 Signal settings for an eight–phase actuated signal control

Assume the arrivals of each movement *i* (*i* = A, B, C, D) are Poisson with a constant rate λ_i, and the corresponding service time is deterministic with rate μ_i, respectively. Let time lost for each phase be *L*. Assume given cycle length is c_o. Recall that the optimal cycle length for a fixed–time signal is given by the Equation (2.6), $c_o = \frac{1.5 \cdot 4L + 5}{1 - I}$, where $4L$ is the total lost time in one cycle. *I* is the total flow ratio of critical movement. For simplification, we may assume *I* is given by parameters of all movements in ring 1,

$$I = \sum_{i=A}^{D} \rho_i = \sum_{i=A}^{D} \frac{\lambda_i}{\mu_i}. \tag{5.3}$$

Based on Webster's formula, the green time allocation should be proportional to corresponding movements' flow ratio,

$$\begin{cases} \Phi_A = L + \dfrac{\rho_A}{I}(c_o - 4L) \\ \Phi_{AB} = 2L + \dfrac{\rho_A + \rho_B}{I}(c_o - 4L) \\ \Phi_{CD} = 2L + \dfrac{\rho_C + \rho_D}{I}(c_o - 4L) \end{cases}. \tag{5.4}$$

We can use this result to set the duration of Φ_{AB} and Φ_{CD}, since the signal timing has a fixed cycle length and two pre–timed barriers. We may also set $\Phi_{A\max} = \Phi_A \beta$, where β were selected from 1.2 to 1.5 to achieve the minimum intersection delay. With the same time lost L, $\Phi_{A\max}$ sets the upper bound of the actuated green phase $G_{A\max}$, which serves southbound left turn traffic.

These are appropriate timing parameters for an 8-phase actuated control. One may use any feasible timing parameter in the algorithm below.

5.3.2 Analysis of movement A

Consider the beginning of the green phase of i^{th} cycle, G_{Ai}. Let the beginning queue size of movement A be X_{Ai}, number of new arrivals during the green phase be Y_{Ai}. The total number of served vehicles during G_A is $X_{Ai}' = X_{Ai} + Y_{Ai}$. Also use XR_{Ai} to denote the residual queue size at the end of G_{Ai}, and let Z_{Ai} denote the number of arrivals

during the red phase following G_{Ai}. If the arrival processes of all movements are deterministic and constant, X_{A1} is given by (5.5). This X_{A1} will be used as the starting point to initialize the stochastic process with Poisson arrivals at the same rate.

$$X_{A1} = \lambda_A (c_o - G_A). \tag{5.5}$$

In section 5.2, an algorithm was developed to calculate the distribution of the actuated green phase for any combination of *v/c* ratio (flow ratio) and number of vehicles waiting at the beginning of the green phase. Let k denote the number of vehicles waiting at the beginning of the green phase (initially $k = X_{A1}$). Let the probability distribution of the actuated green phase be denoted by PA_{kj}, where the subscripts k and j mean the probability (k+j) vehicles are served during the whole busy period, where j is the number of new arrivals during the green phase. Also use C_A to denote the cyclic capacity of movement A for each realization of Φ_A, which is given by,

$$C_A = (\Phi_{A\max} - L)\mu_A = G_{A\max}\mu_A. \tag{5.6}$$

For movement A, the busy period analysis in section 5.2 gives the distribution of the actuated green phase needed to clear the queue, which is denoted by $G_{A1,j}$, where j means the number of new arrivals during the green phase. If the busy period is shorter than the $G_{A\max}$, $G_{A1,j}$ is equal to the busy period. We will call this case α. $G_{A1,j}$ is a discrete random variable in case α. On the other hand, when $G_{A1,j} > G_{A\max}$, = the actuated green phase reaches the upper bound $G_{A\max}$. We call this case β. In this case

$G_{A1} = G_{A\max}$ and a residual queue may exist.

We need to consider two possible cases for cycle 1:

Case α: No residual queue exists at the end of G_{A1}, $G_{A1} < G_{A\max}$, and G_{A1} is a discrete random variable;

Case β: $G_{A1} = G_{A\max}$ residual queue may exist.

Analysis for first cycle

Let P_{A1}^{A} be the probability of case α, and P_{A1}^{B} be the probability of case β, then

$$P_{A1}^{A} = P(X_{A1}' < C_A) = \sum_{j=0}^{C_A-k-1} PA_{k,j}\,, \tag{5.7a}$$

$$P_{A1}^{B} = P(X_{A1}' \geq C_A) = 1 - P_{A1}^{A}\,. \tag{5.7b}$$

Case α:

In case α, G_{A1} is a discrete random variable, ranging from k/μ_A to C_{A1}/μ_A, and it is linearly related to the total number of vehicles being served during the entire green phase, X_{A1}'. The probability of each possible scenario in this case is:

$$P_{A1,j}^{A} = P(Y_{A1} = j) = PA_{k,j} \quad where \quad 0 \leq j \leq (C_A - k)\ . \tag{5.8}$$

For each realization of Y_{A1}, the corresponding phase lengths serving the two movements under consideration are,

$$\begin{cases} G_{A1,j} = (k+j)/\mu_A \\ \Phi_{A1,j} = L + (k+j)/\mu_A = L + G_{A1,j} \\ \Phi_{B1,j} = \Phi_{AB} - \Phi_{A1,j} \end{cases} . \tag{5.9}$$

Figure 5.3 shows the queue length plot of this case.

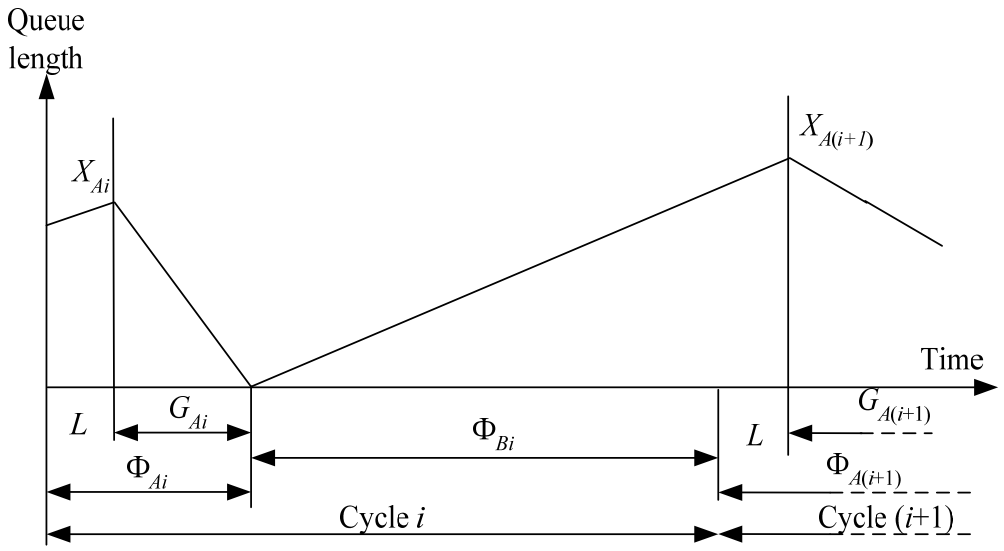


Figure 5.3 Queue length plot of movement A in case α

The corresponding total delay of movement A during $G_{A1,j}$ is,

$$D_{A1,j}{}^{A} = \frac{j}{2} G_{A1,j} . \tag{5.10}$$

For each specific $G_{A1,j}$ the corresponding red phase of movement A, $R_{A1,j}$ is,

$$R_{A1,j} = c_o - G_{A1,j} . \tag{5.11}$$

Since there is no residual queue at the beginning of $R_{A1,j}$, the number of vehicles waiting at the end of $R_{A1,j}$ (also the beginning of G_{A2}), X_{A2}, has the following distribution,

$$P(X_{A2}^{A} = Z_{A1} = n \left| R_{A1,j}) = \frac{(\lambda_A R_{A1,j})^n}{n!} e^{-\lambda_A R_{A1,j}} \right. . \quad (5.12)$$

For all possible $R_{A1,j}$, we have,

$$P(X_{A2}^{A} = Z_{A1} = n \left| R_{A1}) = \sum_{j=0}^{C_A - X_{A1} - 1} \frac{(\lambda_A R_{A1,j})^n}{n!} e^{-\lambda_A R_{A1,j}} \right. . \quad (5.13)$$

The total delay in the first cycle for each possible scenario of case α (k vehicles waiting at the beginning of G_{A1}, j vehicle arrivals during the green phase, G_{A1}<$G_{A\max}$, and n vehicle arrivals during $R_{A1,j}$) is,

$$D_{A1,j,n}{}^{A} = \frac{1}{2}(G_{A1,j}k + R_{A1,j}n) . \quad (5.14a)$$

Therefore the average delay in one cycle is,

$$d_{A1,j,n}{}^{A} = D_{A1,j,n}{}^{A} / (j+n) . \quad (5.14b)$$

The corresponding probability of this scenario is $PA_{X_{A1},j} P(Z_{A1} = n \left| R_{A1,j}) \right.$.

Case β:

For case β, the total number of vehicles waiting the green phase is equal to or greater than the capacity C_A, and the green phase length is equal to $G_{A\max}$. The probability distribution of Y_{A1} is given by (5.15). A residual queue may exist at the end of green phase. Because the arrival process is Poisson, the conditional distribution of the arrival times for a given j new arrivals is uniform. Figure 5.4 shows the queue length

plot for case β.

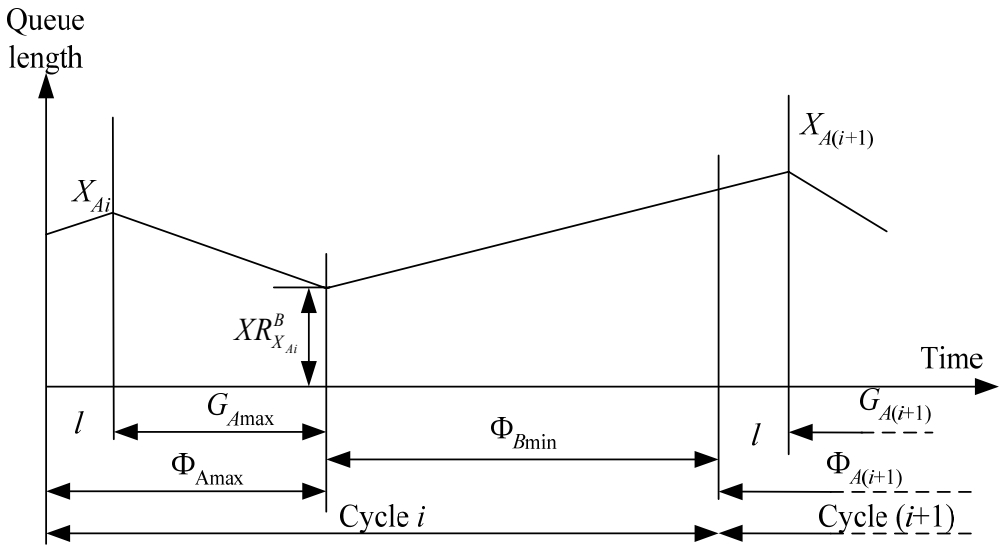


Figure 5.4 Queue length plot of movement A in case β

$$P_{j,G_{A\max}} = P(Y_{A1} = j | G_{A\max}) = \frac{(\lambda_A G_{A\max})^j}{j!} e^{-\lambda_A G_{A\max}} \quad j \geq (C_A - k). \tag{5.15}$$

Then the residual queue size at the end of $G_{A\max}$ is,

$$XR^B_{X_{A1},j} = X_{A1} + Y_{A1} - C_A. \tag{5.16}$$

The conditional (given $X_{A1} = k$) probability of each scenario in case β is,

$$\begin{aligned} P^B_{A1,l} &= P(XR^B_{X_{A1},j} = X_{A1} + Y_{A1} - C_A = k + j - C_A = l) \\ &= P(Y_{A1} = C_A + l - k = j) = P_{j,G_{A\max}} \end{aligned}. \tag{5.17}$$

where $Y_{A1}+X_{A1} \geq C_A$.

The corresponding total delay for movement A during G_{Amax} is,

$$D_{A1,j}{}^{B} = \frac{1}{2} G_{A\max}(X_{A1} + XR^{B}_{X_{A1,j}}). \tag{5.18}$$

In this case, the red phase following $G_{A\max}$, R_{Amin}, is equal to $c_o - G_{A\max}$. Number of arrivals during R_{Amin} has the following distribution,

$$P_{m,R_{A\min}} = P(Z_{A1} = m \mid R_{A\min}) = \frac{(\lambda_A R_{A\min})^m}{m!} e^{-\lambda_A R_{A\min}}. \tag{5.19}$$

The total delay of the served vehicles for this scenario (k vehicles waiting at the beginning of green phase; j vehicle arrival during the green phase $G_{A\max}$; residual queue at the end of $G_{A\max}$ is $l=k+j-C_A$, and there are m vehicles arrived during the red phase, at the beginning of G_{A2}, $n=l+m$ vehicles are waiting for service) is,

$$D_{A1,j,n}{}^{B} = \frac{1}{2}[G_{A\max}(X_{A1} + XR_{X_{A1,J}}{}^{B}) + R_{A\min}(XR_{X_{A1,J}}{}^{B} + X_{A2})], \tag{5.20}$$

where

$X_{A1} = k$,

$Y_{A1} = j$,

$XR^{B}_{X_{A1},j} = X_{A1} + Y_{A1} - C_A = l$,

$Z_{A1} = m$,

$X_{A2} = XR^{B}_{X_{A1},j} + Z_{A1} = n$.

The total number of vehicle arrivals during the whole cycle is ($j+m$), the average delay is,

$$d_{A1,j,n}{}^{B} = D_{A1,j,n}{}^{B}/(j+m). \tag{5.21}$$

The probability of this scenario is

$$P_{j,G_{A\max}} P_{m,R_{A\min}}. \tag{5.22}$$

Queue size at the beginning of the next green phase is determined by the residual queue size at the end of the green phase plus the number of arrivals during the red phase of the current cycle,

$$X_{A2}^{B} = XR_{X_{A1,k}}^{B} + Z_{A1}. \tag{5.23}$$

Given that case β occurred in the first cycle, the probability of exact n vehicles waiting at the beginning of the next green phase is,

$$P(X_{A2}^{B} = n) = \sum_{m=0}^{n} P_{A1,(n-m)}^{B} P_{m,R_{A\min}}. \tag{5.24}$$

Combining (5.13) and (5.23) we get,

$$P(X_{A2} = n) = P(X_{A2}^{A} = n) + P(X_{A2}^{B} = n). \tag{5.25}$$

This result gives the queue distribution at the beginning of the second green phase, which will be used in the analysis of the cycle 2. At the beginning of cycle 2, the number of vehicles waiting at the beginning of the green phase, X_{A2}, is a discrete random variable with known probability given by (5.25).

Analysis for second cycle

Let P_{A2}^{A} denote the probability of case α (green phase length is shorter than $G_{\max}$) in cycle 2, and use P_{A2}^{B} denote the probability of case β (green phase length is equal to $G_{\max}$), then

$$P_{A2}^{A} = \sum_{all\ k<C_A} \left[P(X_{A2} = k) P(X_{A2}' < C_A \middle| X_{A2} = k) \right] = \sum_{all\ k<C_A} \left[P(X_{A2} = k) \sum_{j=0}^{C_A - k - 1} PA_{k,j} \right], \qquad (5.26a)$$

$$P_{A2}^{B} = 1 - P_{A2}^{A}. \qquad (5.26b)$$

Case α_2: No residual queue at the end of green phase in cycle 2

Arbitrarily select a feasible beginning state of cycle 2, say, $X_{A2} = k < C_A$, whose probability, $P(X_{A2} = k)$, is given by (5.25). In case α, G_{A2}, the green phase serving movement A in cycle 2 is a discrete random variable, ranging from k/μ_A to $(C_A - 1)/\mu_A$. It is determined by the total number of vehicles being served during the entire green phase, $X_{A2}' = X_{A2} + Y_{A2}$, where Y_{A2} is the number of new arrivals during the green phase. The conditional probability of each possible scenario is:

$$P_{A2,k,j}^{A} = P(Y_{A2} = j) = PA_{k,j} \quad 0 \le j \le (C_A - k - 1). \qquad (5.27)$$

For each realization of j, the corresponding green phase lengths for the two movements under consideration are,

$$\begin{cases} \Phi_{A2,k,j} = L + (X_{A2} + Y_{A2}) / \mu_A = L + G_{A2,k,j}, \\ \Phi_{B2,k,j} = \Phi_{AB} - \Phi_{A2,k,j}. \end{cases} \tag{5.28}$$

And the probability of this to occur is,

$$PA_{k,j} P(X_{A2} = k). \tag{5.29}$$

The corresponding total delay of movement A occurring during $G_{A2,k,j}$ is,

$$D_{A2,k,j}{}^{A} = \frac{1}{2} k G_{A2,k,j}. \tag{5.30}$$

For each specific $G_{A2,k,j}$ the corresponding red phase, $R_{A2,k,j}$ is,

$$R_{A2,k,j} = c_0 - G_{A2,k,j}. \tag{5.31}$$

Use $G_{A2,v}$ to denote that v, $v < C_A$, which implies it is case α) vehicles are served in total. It is easy to compute $G_{A2,v}$ since it includes all $G_{A2,k,j}$, where $(k+j) = v$.

For each specific $G_{A2,v}$ the corresponding red phase, $R_{A2,v}$ is,

$$R_{A2,v} = c_O - G_{A2,v}.$$

The probability distribution of $G_{A2,v}$ is,

$$P(G_{A2,v}) = P(R_{A2,v}) = \sum_{all \quad k \le v} [P(X_{A2} = k) PA_{k,(v-k)}]. \tag{5.32}$$

It is also straightforward to compute the probability that the green phase length G_{A2} is equal to $G_{A\max}$,

$$P(G_{A2\max}) = 1 - \sum_{t<C_A} P(G_{A2,v}) . \tag{5.33}$$

There is no residual queue at the beginning of each $R_{A2,v}$ because the actuated green phase ends before reaching the upper bound (case α). Thus the number of vehicles waiting at the end of $R_{A2,v}$ (also the beginning of G_{A3}), X_{A3}, has the following distribution,

$$P(X_{A3}^{A} = Z_{A2} = m \left| R_{A2,v}) = \frac{(\lambda_A * R_{A2,v})m}{m!} e^{-\lambda_A R_{A2,v}} . \right. \tag{5.34}$$

The total delay during this cycle for this scenario (k vehicles waiting at the beginning of green phase and $(k+j) = v$ vehicles are served with $G_{A2,v} < C_A$. At the end of cycle 2, there are m vehicles waiting in the queue) is,

$$D_{A2,k,j,m}{}^{A} = \frac{1}{2}(G_{A2,k,j}k + R_{A2,k,j}m) , \tag{5.35}$$

Then the average delay is,

$$d_{A2,k,j,m}{}^{A} = D_{A2,k,j,m}{}^{A} / (j+m) . \tag{5.36}$$

The probability of this scenario (k vehicles waiting at the beginning of green; j vehicles arrived during the green phase; $(k+j)$ vehicles served; m vehicles arrived during the red phase) to occur is,

$$P(X_{A2} = k) PA_{k,j} P(X_{A3}^{A} = m \left| R_{A2,(k+j)}) . \right. \tag{5.37}$$

For all possible $R_{A2,v}$, the probability of exact k (we do not use m here because it is the queue size at the beginning of cycle 3) vehicles waiting at the beginning of green phase in cycle 3, given case α occurred in cycle 2, is,

$$P(X_{A3}^{A} = k) = \sum_{k<C_A} P(R_{A2,v}) \frac{(\lambda_A R_{A2,v})k}{k!} e^{-\lambda_A R_{A2,v}} . \tag{5.38}$$

Case β_2: Residual queue at the end of green phase in cycle 2

When case β occurs in cycle 2, the probability of $X_{A2} = k$, is also given by (5.25). Let j denote the number of new arrivals during the green phase, $G_{A\max}$ (since in case β the green phase is $G_{A\max}$). The probability of each scenario with a specific j is given by (5.15). A residual queue may exist at the end of the green phase. The residual queue size at the end of G_{A1} is,

$$XR_{A2,k,j,l}^{B} = k + j - C_A = l , \tag{5.39}$$

with the conditional probability (given $X_{A2} = k$) of,

$$\begin{aligned} P_{A2,k,j,l}^{B} &= P(XR_{A2,k,j,l}^{B} = l) \\ &= P(j = C_A + k - l) = P_{j,G_{A\max}} , \end{aligned} \tag{5.40}$$

where $j \geq C_A - k$.

The corresponding total delay of movement A during $G_{A\max}$ is,

$$D_{A2,k,j}{}^{B} = \frac{1}{2} G_{A\max}(k + l) . \tag{5.41}$$

In this situation, the red phase following the $G_{A\max}$, R_{Amin}, is equal to $c_o - G_{A\max}$. The probability distribution of queue size at the end of R_{Amin} (also the beginning of cycle 3), n, is,

$$P(X_{A3}^{B} = n \mid X_{A2} = k) = \sum_{all\ \ j} P_{j,G_{A\max}} \frac{(\lambda_A R_{A\min})^{(n-l)}}{(n-l)!} e^{-\lambda_A R_{A\min}}, \tag{5.42}$$

where $l=k+j-C_A$.

The unconditional probability of $P(X_{A3}^{B} = n_3)$ is,

$$P(X_{A3}^{B} = n) = \sum_{all\ \ k} [P(X_{A2} = k) \sum_{all\ \ j} P_{j,G_{A\max}} \frac{(\lambda_A R_{A\min})^{(n-l)}}{(n-l)!} e^{-\lambda_A R_{A\min}}]. \tag{5.43}$$

Combining (5.34) and (5.43) we get,

$$P(X_{A3} = n) = P(X_{A3}^{A} = n) + P(X_{A3}^{B} = n) \tag{5.44}$$

The total delay of movement A during $R_{A\min}$ is,

$$D_{A2,l,n}{}^{B} = \frac{1}{2} R_{A\min}(l+n). \tag{5.45}$$

Therefore the total delay of movement A in cycle 2 is,

$$D_{A2k,j,n}{}^{B} = \frac{1}{2}[G_{A\min}(k+l) + R_{A\min}(l+n)]. \tag{5.46}$$

The corresponding average delay of movement A in cycle 2 is,

$$d_{A2k,j,n}^{\ B} = \frac{1}{2}[G_{A\min}(k+l) + R_{A\min}(l+n)]/(j+m), \tag{5.47}$$

where $m = n - l$.

The probability of this scenario (k vehicles waiting at the beginning of green; j vehicle arrivals during the green phase; green phase equal to $G_{A\max}$; m vehicle arrivals during the red phase $R_{A\min}$) to occur is,

$$P(X_{A2} = k)P_{j,G_{A\max}}P_{m,R_{A\min}}. \tag{5.48}$$

Average delay for second cycle

The expected average delay in cycle 2, considering both case α and case β, is,

$$\begin{aligned} E(d_{A2}) = & \sum_{all} \sum_{k<C_A} \sum_{all} \sum_{j<(C_A-k)} \sum_{all} \sum_{m} P(X_{A2} = k)PA_{k,j}P(X_{A3}^{A} = m \mid R_{A2,(k+j)})d_{A2,k,j,m}^{\ A} \\ & + \sum_{all} \sum_{k} \sum_{j\geq(C_A-k)} \sum_{all} \sum_{m} P(X_{A2} = k)P_{j,G_{A\max}}P_{m,R_{A\min}}d_{A2k,j,n}^{\ B}. \end{aligned} \tag{5.49}$$

The analysis of the second cycle is now completed. For each scenario, the green phase length, number of vehicles being served, queue plot, delay and probability of this scenario to occur are given, which can be used to compute several performance measures of a signalized intersection. The same methodology can be applied to the succeeding cycles until the system reaches steady state.

5.3.3 Analysis of movement B

In this section we analyze queues and delay for movement B. Based on the signal setting, the feasible range of green phase G_B is $(\Phi_{AB} - G_{A\max} - 2L, \Phi_{AB} - 2L - 1/\mu_A) \bigcup (\Phi_{AB} - 2L)$, where the lower bound corresponds to $G_{A\max}$, and the upper bound implies that Φ_A is skipped. For the purpose of convenience, we use G_B's capacity in the analysis, where $C_{B,\min} = (\Phi_{AB} - G_{A\max} - 2L)\mu_A$, $C_{B,\max} = (\Phi_{AB} - L)\mu_B$ and $C_{B,\max}' = (\Phi_{AB} - 2L - 1/\mu_A)\mu_B$. Assume that discrete random variable G_B has n realizations, that is, $C_{B,\max} - C_{B,\min} + 2 = n$. Let $P_{G_B,i}$ denote the probability of each specific green phase $G_{B,i}$ at the steady state, where the subscript i ranges from 1 to n, $i = 1$ denotes the lower bound of the green phase $G_{B,\min}$ and $i = n$ denotes the upper bound $G_{B,\max}$. The corresponding capacity is denoted by $C_{B,i}$, and $C_{B,i} = G_{B,i}\mu_B$. Let $R_{B,i}$ denote the red phase corresponding to $G_{B,i}$. Because of the fixed cycle length, $G_{B,i} + R_{B,i} = c_o$. The phase length distribution of movement B is determined by the movement A, whose steady state distribution is given by the algorithm in chapter 5.3.2. We use P_k to denote the probability that exactly k vehicles are waiting at the beginning of red phase (this time point is viewed as the beginning of each cycle in the analyses). Let $P_{k,i}$ denote the probability for this to happen with each specific $G_{B,i}$, and $R_{B,i}$. Because of the independency of the green phase distribution, the relation between P_k and $P_{k,i}$ is,

$$P_{k,i} = P_k P_{G_B,i}. \tag{5.50}$$

For each whole cycle (beginning with the red phase), the distribution of the number of new arrived vehicles j is given by (5.51)

$$P_{j,c_o} = \frac{(\lambda_B c_o)^j}{j!} e^{-\lambda_B c_o}. \tag{5.51}$$

Since j arrivals are uniformly distributed in the whole cycle for a given j value, then the queue length at the end of red phase is,

$$Q_{k,j,i} = k + j\frac{R_{B,i}}{c_o}. \tag{5.52}$$

If $(k+j) \le C_{B,i}$, all vehicles will be served during $G_{B,i}$, and the number of vehicles waiting at the beginning of the next red phase is zero. Let this be case α. Otherwise, $C_{B,i}$ vehicles are served during $G_{B,i}$, and a residual queue with size of $(k+j- C_{B,i})$ exists at the beginning of the next red phase; let this be case β.

Residual queue, the number of vehicles waiting at the end of the green phase, is given by (5.53),

$$XR_{\mathrm{B}}(k,j,i) = \begin{cases} 0 & if \quad (k+j) \le C_{B,i} \\ k + j - C_{B,i} & otherwise. \end{cases} \tag{5.53}$$

Case α: Residual queue at the end of green phase

For case α, the queue length plot is illustrated in Figure 5.5. Assuming it takes time t_c to clear the queue, the total number of vehicles served from the beginning of the cycle

to the queue clearance point is equal to the number of vehicles requesting the phase. Therefore, the time used to clear the queue, t_c, is given by:

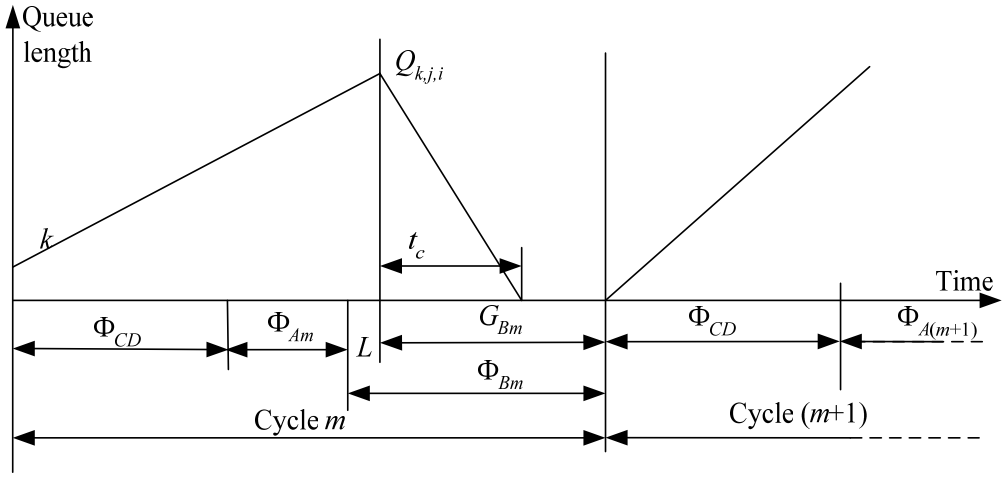


Figure 5.5 Queue length plot of movement B in case α

$$k + j\frac{R_{B,i} + t_c}{c_o} = \mu_B t_c \Rightarrow t_c = \frac{jR_{B,i} + kc_o}{\mu_B c_o - j}. \tag{5.54}$$

Total delay during the whole cycle is,

$$\begin{aligned} D^A_{k,j,i} &= \frac{1}{2}(k + Q_{k,j,i})R_{B,i} + \frac{1}{2}Q_{i,j,k}t_c \\ &= \frac{1}{2}(2k + j\frac{R_{B,i}}{c_o})R_{B,i} + \frac{1}{2}(k + j\frac{R_{B,i}}{c_o})\frac{R_{B,i}j + kc_o}{\mu_B c_o - j}. \end{aligned} \tag{5.55}$$

Average delay during this cycle is,

$$\begin{aligned} d^A_{k,j,i} &= D^A_{k,j,i} / j \\ &= \frac{1}{2j}[(2k + j\frac{R_{B,i}}{c_o})R_{B,i} + (k + j\frac{R_{B,i}}{c_o})\frac{R_{B,i}j + kc_o}{\mu_B c_o - j}]. \end{aligned} \tag{5.56}$$

Case *β*: No residual queue at the end of green phase

For case *β*, the queue length plot for the whole cycle is illustrated in Figure 5.6. Queue size $Q_{k,j,i}$ at the end of red phase is given by (5.52). At the end of the green phase, residual queue size is ($i+j-C_{B,k}$). The total delay and average delay during the whole cycle is,

$$\begin{aligned} D^{B}_{k,j,i} &= \frac{1}{2}(k+Q_{k,j,i})R_{B,i} + \frac{1}{2}(Q_{k,j,i}+k+j-C_{B,i})G_{B,i} \\ &= \frac{1}{2}(2k+j\frac{R_{B,i}}{c_o})R_{B,i} + \frac{1}{2}[2k+j(1+\frac{R_{B,i}}{c_o})-k]G_{B,i}. \end{aligned} \tag{5.58}$$

Average delay during this cycle is,

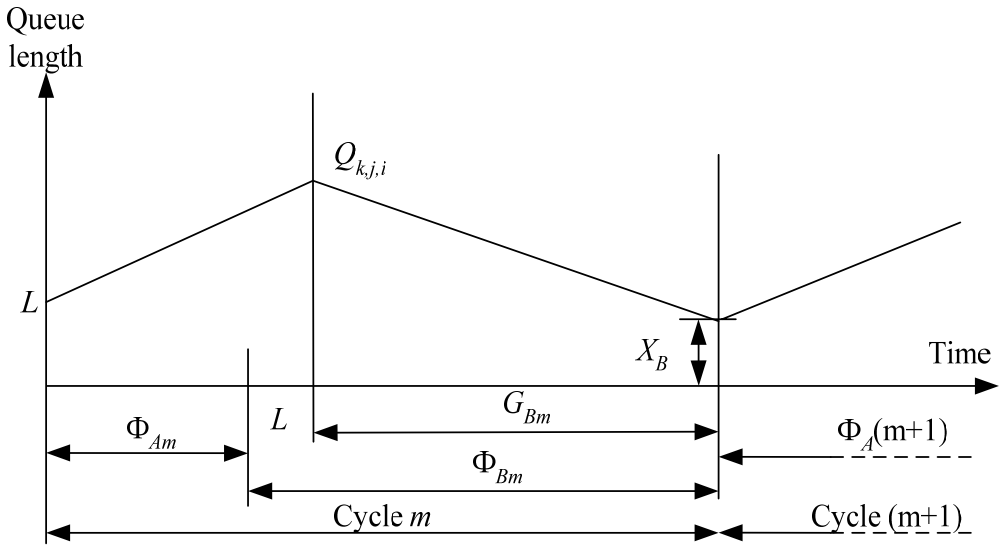


Figure 5.6 Queue length plot of movement B in case *β*

$$\begin{aligned} d^{B}_{k,j,i} &= D^{B}_{k,j,i} / j \\ &= \frac{1}{2j}[(2k+j\frac{R_{B,i}}{c_o})R_{B,i} + [2k+j(1+\frac{R_{B,i}}{c_o})+C_{B,i}]G_{B,i}]. \end{aligned} \tag{5.59}$$

When the system is at the steady state, the queue length at the end of each green phase follows same distribution. The unknown steady-state variables of the system are P_k. We solve the following set of linear equations to obtain P_k.

$$\sum_{all\ \ k,j,i\ \ XR_B(k,j,i)=k} P_{k,i}P_{j,c_o} = \sum_{all\ \ k,j,i\ \ XR_B(k,j,i)=k} P_k P_{G_B,i} P_{j,c_o} = P_k \quad all \quad k. \tag{5.60a}$$

$$\sum_{all\ \ k} P_k = 1. \tag{5.60b}$$

Average delay of movement B

Given the values of P_i, the expectation of average delay of vehicles requesting movement B can be calculated from the conditional probability of each feasible scenario.

$$\begin{aligned} E(d_B) = & \sum_{all\ \ k<C_{B,i}} \sum_{iall\ \ i\ \ all} \sum_{j<(C_B-k)} P_{k,i}P_{j,c_o} d_{k,j,i}^{\ \ A} \\ & + k \sum_{all\ \ iall\ \ iall} \sum_{j\geq(C_B-k)} P_{k,i}P_{j,c_o} d_{k,j,i}^{\ \ B}. \end{aligned} \tag{5.61}$$

5.4 Model evaluation

In this section, the analytical queuing model is evaluated by comparing its numerical solution with simulation-based results. The simulation model of the actuated control was built in MATLAB. Two groups of exponentially distributed random variables were generated by MATLAB for vehicle inter-arrival times of movement A and B. Service rates and lost time were the same as those of the analytical models. For each type of loading pattern, the simulation model was run 20 times, with length of 2000

cycles each. The first 100 cycles of each run were considered warm-up periods and truncated. For each scenario, we assume the flow are symmetric, that is $\rho_A = \rho_B$, $\rho_C = \rho_D$.

Two loading patterns were simulated. In each case we examined (i) the average phase lengths of movement A given by the numerical method and simulation (green phase of movement B is determined by that of movement A), (ii) Number of vehicles of movement A served in one cycle, (iii) the average delays of movements A and B and (iv) the number of iterations for the numerical method to converge.

Case 1: Identical saturation flow rate with balanced flow ratios

In this case, the saturation flow rates of the two directions are 0.5 veh/s, the arrival rates of movement A and B are equal, and range from 0.04 veh/s to 0.09 veh/s. Queue discharge lost time was set to 4 seconds for all the cases.

Table 5.1 gives the average number of vehicles of movement A served in one cycle and average phase lengths of movement A. The table shows that the means of these measures computed by the analytical model match well with those given by the simulation model.

Figure 5.7 and 5.8 shows the average intersection delays calculated by the queuing model and obtained by simulation. The delay plots show that the average delays given by the two methods match fairly well when the intersection flow ratio is less than 0.6.

As the total flow ratio increases beyond this point, the difference of movement B increases. The analytical model always tends to underestimate the average delays compared to the simulation model. The probability distribution of actuated green phase G_A and new arrivals during G_A is calculated by the algorithm presented in section 5.3 if G_A is shorter than $G_{A\max}$. If $G_A = G_{A\max}$, we assume the new arrivals during $G_{A\max}$ follow Poisson distribution. The total probability of new arrivals of these two cases normally close but not equal to 1. This explains the discrepancy between the analytical model and the simulation.

Intersection Flow ratio	Number of vehicles served		Green phase length	
	Ana model	Simulation	Analytical	Simulation
0.32	1.03	1.11	5.53	5.07
0.4	1.55	1.58	6.84	6.34
0.48	2.2	2.16	8.33	7.82
0.56	3.1	2.81	10.01	9.39
0.64	4.34	3.83	12.38	11.56
0.72	6.46	5.76	16.33	15.5

Table 5.1 Comparison of average number of vehicles served in one cycle and average half cycle lengths of movement A (Case 1).

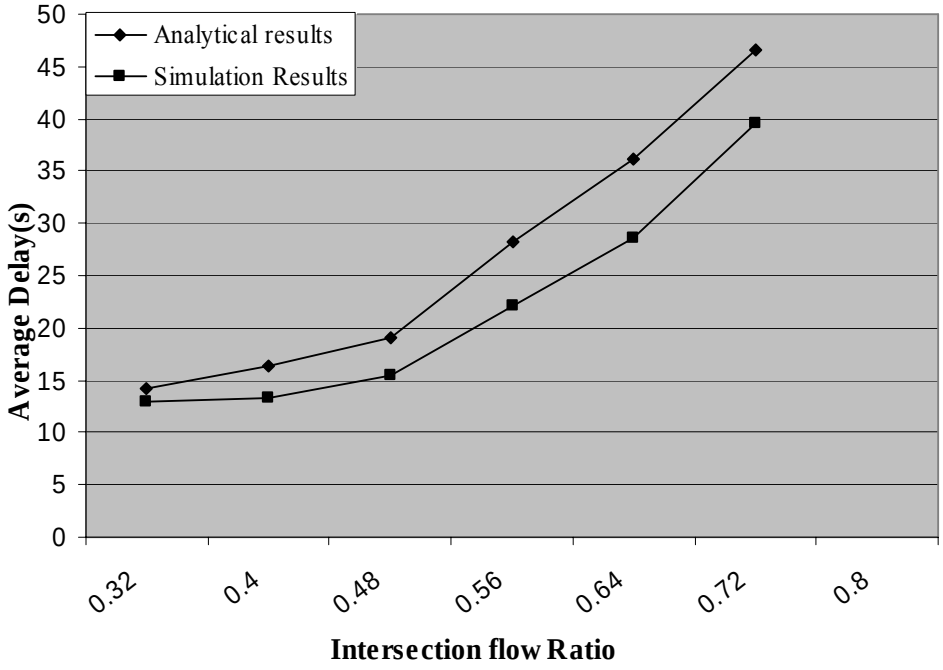


Figure 5.7 Average intersection delay of movement A (Case1)

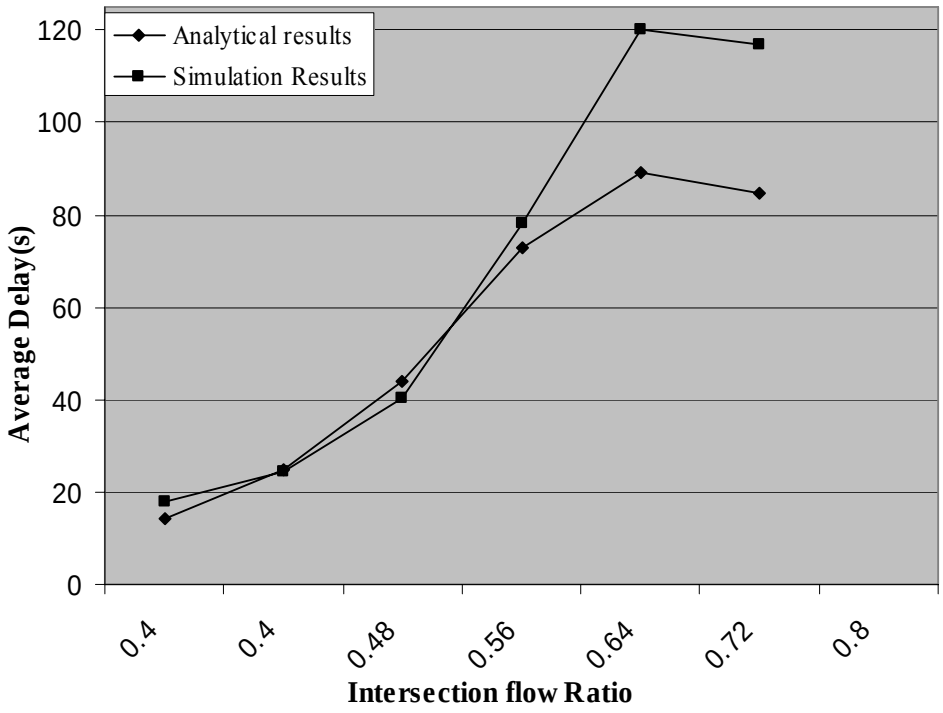


Figure 5.8 Average intersection delay of movement B (Case1)

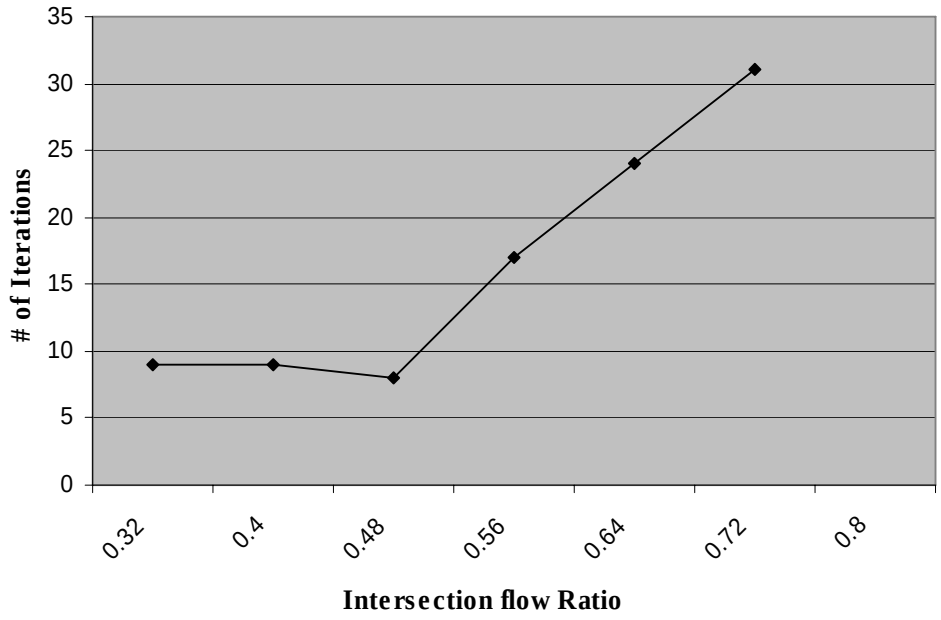


Figure 5.9 Number of iterations to converge (Case 1)

For a numerical method, number of iterations needed to converge is a major issue. Figure 5.9 shows the numbers of iterations needed to converge, defined as when the average delay of movement A does not change by more than 10^{-2} s from iteration to iteration. The figure shows that as the total flow ratio increases, the number of iterations needed to converge tends to increase.

Case 2: Identical saturation flow rate with unbalanced flow ratios.

The departure flow rates for both directions were set to 0.5 veh/s again. The arrival rates of movement A vehicles ranged from 0.03 veh/s to 0.06 veh/s, while the arrival rates of movement B vehicles were twice those of movement A vehicles.

Table 5.2 gives the average numbers of vehicles served in one cycle and average half

cycle lengths of movement A, given by the analytical model and the simulation model. The results match well.

Intersection Flow ratio	Number of vehicles served		Green phase length	
	Ana model	Simulation	Analytical	Simulation
0.18	0.73	0.96	5.16	4.53
0.24	1.18	1.45	6.32	5.95
0.3	2.05	2.2	8.33	7.94
0.36	3.91	3.84	12.14	11.57

Table 5.2 Comparison of average number of vehicles served in one cycle and average half cycle lengths of movement A (Case 2).

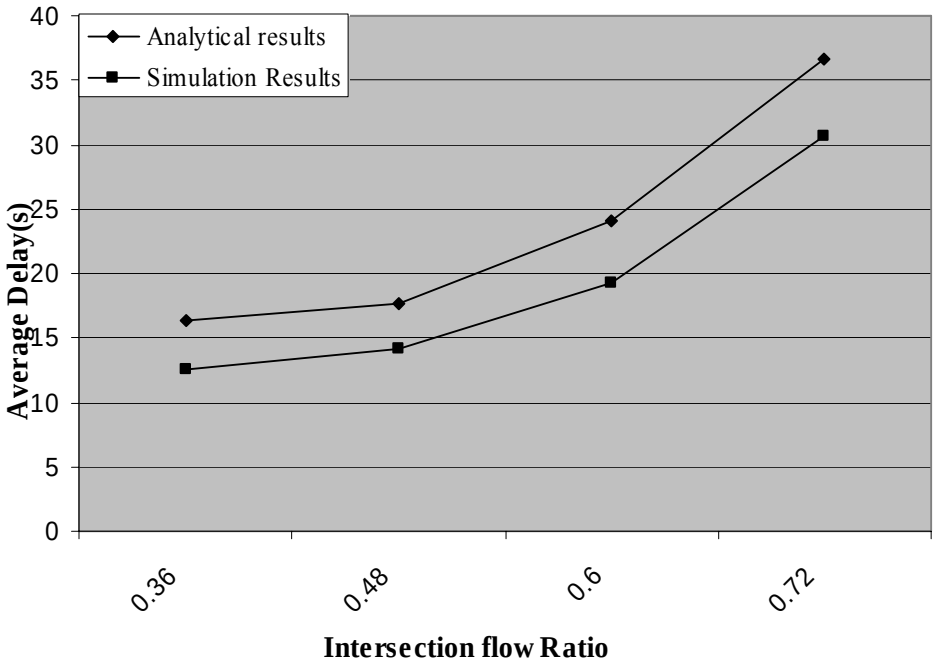


Figure 5.10 Average intersection delay of movement A (Case 2)

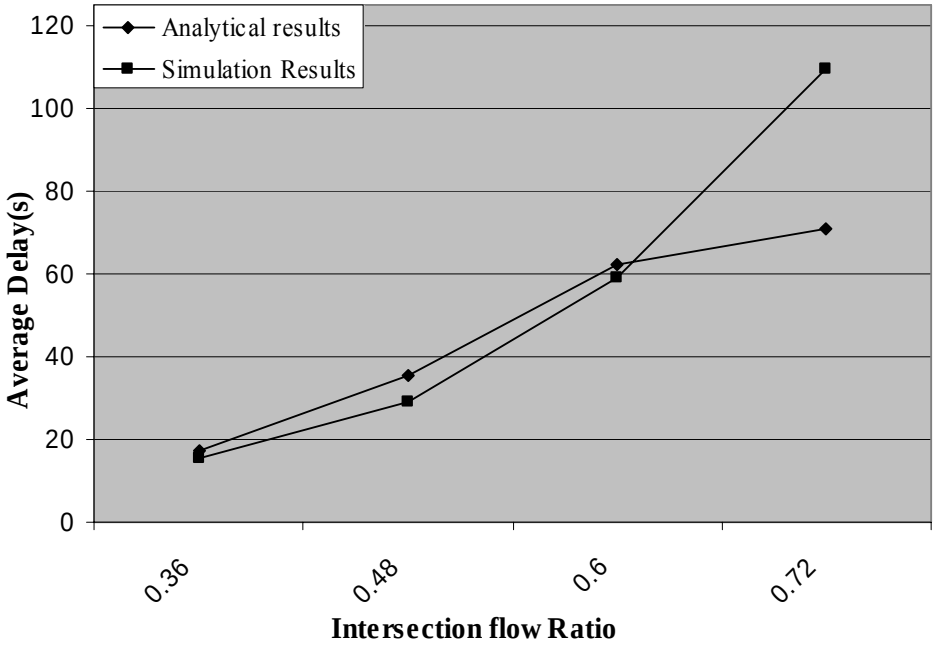


Figure 5.11 Average intersection delay of movement B (Case 2)

Figure 5.10 and 5.11 show the average delays in the two directions given by the two methods. The delay plots show that the average delays given by the two methods are fairly close when the flow ratio is relatively low, say, less than 0.6. As the intersection flow ratio increases, the differences of movement B increase. As before, the numerical queuing model tends to underestimate the average delay as compared with the simulation model.

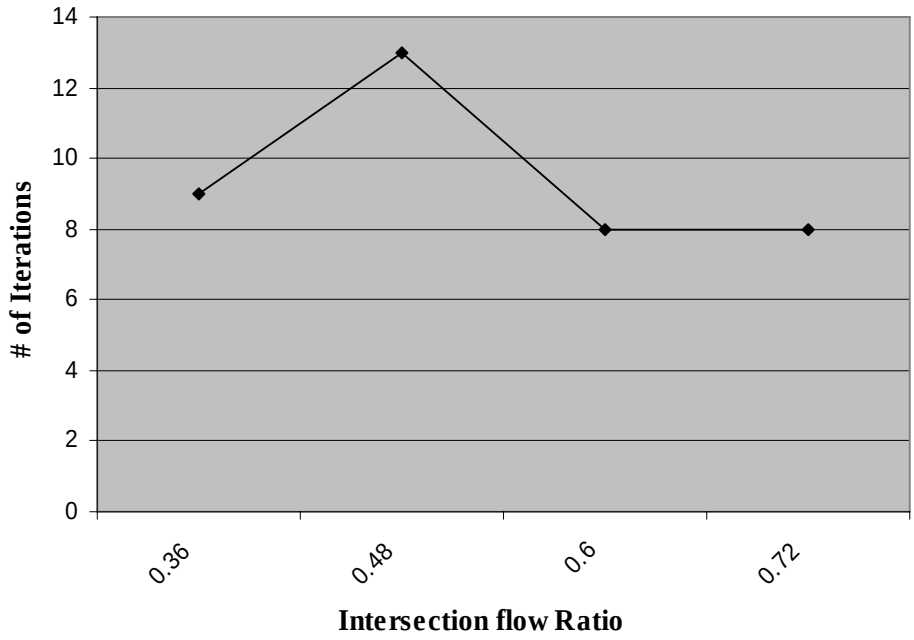


Figure 5.12 Number of iterations to converge (Case 2)

Figure 5.12 plots the numbers of iterations needed to converge at various flow ratios. Number of iterations needed to converge is always less than 15.

5.5 Summary

In this chapter a queuing model is developed to analyze an eight-phase actuated traffic signal. First, the distribution of the actuated green phase duration, or the conditional "busy period" is analyzed. An algorithm is presented to calculate this distribution. Then a simplified eight-phase actuated signal control system is modeled. The leading left turn phase is actuated with an upper bound, followed by a green phase serving the opposing (conflicted) through movement.

The actuated left phase length is a discrete random variable. A residual queue may

exist at the end of the green phase only if the green time reaches the upper limit (i.e. it "max outs").

Since we are interested in steady state performance, we arbitrarily select a starting point to start the iterative computations for calculating the distribution of the left turn phase length and residual queue. For each scenario for the next cycle, we calculate the average delay, left turn phase length and residual queue. Since the probability of each scenario is known, statistics of these performance measures can be easily calculated. The iterative computations are applied from cycle to cycle until statistics converge to a steady state.

The left turn phase length distribution determines the green phase distribution of the opposing through movement. We select the residual queue distribution of through movement at the end of green phase as the system state variable. For each scenario the steady state residual queue distribution at the end of the green phase is calculated. By solving the associated linear equations we can get the unconditional steady-state residual queue distribution. With this distribution and the phase length distribution calculated for the left turn movement, the steady-state statistics of the performance measures for each cycle can be calculated.

The analytical model was evaluated by comparing it with a simulation model. The results show the analytical give fairly good results on the average green phase lengths and number of vehicles served of movement A. The average delay given by the

analytical model tends to be underestimated, but acceptable when the intersection total flow ratio is less than 0.6.

CHAPTER 6 ANALYTICAL MODELING OF RAMP METER CONTROL

6.1 Introduction

Much research has been reported on ramp meter system control strategies and algorithms, but very little of it focuses on the analytical modeling of the ramp meter control itself. In this chapter we propose analytical queuing models for ramp control, with pre-timed and simplified adaptive strategies. With this model we can solve the steady state queue length probabilities analytically and estimate the vehicle arrival rate by comparing the results with the upstream ramp detector's occupancy.

6.2 Pre-timed ramp metering control

As mentioned before, pre-timed control has a deterministic rate setting. One cycle consists a red phase and a green phase. During the green phase, one vehicle can enter the freeway. Figure 6.1 illustrates a possible queue length plot for the system. Assuming vehicle arrivals are Poisson, if we view one cycle as a time interval, we can model the process as a Markovian process where the queue size at the end of each cycle (after departure, if there is any) is the system state.

Let λ denote the vehicle arrival rate. If we observe the state probability distribution at the end of each cycle C, at time points of T, $2T$,…nT, the transition probability matrix is,

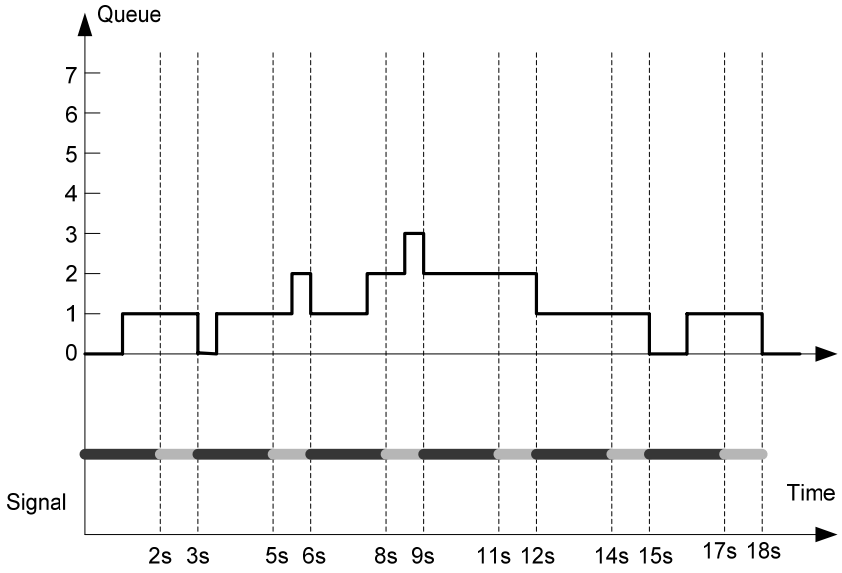


Figure 6.1 Queue length of pre-timed ramp meter system

$$P = \begin{bmatrix} k_0 + k_1 & k_2 & k_3 & k_4 & k_5 & k_6 & \cdots \\ k_0 & k_1 & k_2 & k_3 & k_4 & k_5 & \cdots \\ 0 & k_0 & k_1 & k_2 & k_3 & k_4 & \cdots \\ 0 & 0 & k_0 & k_1 & k_2 & k_3 & \cdots \\ 0 & 0 & 0 & k_0 & k_1 & k_2 & \cdots \\ 0 & 0 & 0 & 0 & k_0 & k_1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \cdots \end{bmatrix} \tag{6.1}$$

where

k_n =Pr{n vehicle arrivals during a cycle c}.

The traffic intensity is defined as

$\rho = \lambda / \mu = \lambda c$,

where μ is the service rate,

$\mu = 1/c$.

Since arrivals are Poisson, we have

$$k_n = \frac{e^{-\lambda C}(\lambda c)^n}{n!} = \frac{e^{-\rho}\rho^n}{n!}. \tag{6.2}$$

Assuming steady state exists, let π_i denote the probability of exactly n vehicles waiting in the queue at the end of each cycle. We have the following equations form the steady-state prospects of the underlying Markov Chain.

$$\pi_0(k_0 + k_1) + \pi_1 k_0 = \pi_0 \tag{6.3.1}$$

$$\pi_0 k_2 + \pi_1 k_1 + \pi_2 k_0 = \pi_1 \tag{6.3.2}$$

$$\pi_0 k_3 + \pi_1 k_2 + \pi_2 k_1 + \pi_3 k_0 = \pi_2 \tag{6.3.3}$$

$$\vdots \qquad\qquad \vdots$$

$$\pi_0 k_i + \pi_1 k_{i-1} + \pi_2 k_{i-2} + \cdots + \pi_i k_0 = \pi_{i-1} \tag{6.3.i}$$

$$\vdots$$

Now define the moment generating functions

$$\Pi(z) = \sum_{i=0}^{\infty} \pi_i z^i \quad (|z| \le 1),$$

and

$$\mathrm{K}(z) = \sum_{i=0}^{\infty} k_i z^i \quad (|z| \le 1).$$

Then multiply Equation (6.3.*i*) by z^{i-1}, sum over *i*, and solve for $\Pi(z)$, which yields

$$\Pi(z) = \frac{\pi_0 k_0 (z-1)}{z - \mathrm{K}(z)} \quad . \tag{6.4}$$

Using the fact that $\Pi(1) = 1$, along with L'Hopital's law, we find that

$$\frac{\pi_0 k_0}{1 - \dot{\mathrm{K}}(z)} |_{z=1} = 1 \quad , \tag{6.5}$$

and realize that

$$\dot{\mathrm{K}}(1) = \lambda / \mu = \lambda c = \rho \; .$$

From(4), we then have

$$\pi_0 = \frac{1-\rho}{k_0} \quad . \tag{6.6}$$

Substituting the value of π_0 into Equations (6.3), we can obtain the values of other π_i, the steady state probabilities at the end of each cycle *c*. These cannot be used directly to estimate the arrival rate. That is because the detector gives the occupancy over the whole time axis, not at the specific points *T*, 2*T*,…*nT*. Below we develop the relationship between P_i and π_i, *i*= 1, 2, …

Normally there is a magnetic detector upstream of the ramp meter. Assume that a presence detector is located at the i^{th} position of a virtual queue. Then a vehicle

stopped over the detector will trigger the detector (in this model, we omit the detector occupancy caused by passing vehicles, but include only vehicle presence). The occupancy of the detector over a long time interval gives the percentage of time during which queue is not less than i. Let p_i be the probability that the queue size is equal to i at an arbitrarily time point. Note that p_i are the steady state probabilities over the time axis, not at the end epochs of each cycle. Obviously there is a relationship between π_i and p_i. To determine this relationship, imagine a general cycle c, whose beginning state and ending state's distribution are given by π_i. In this abstract cycle, the state can change from m to n ($n \geq \max(m\text{-}1,0)$). The p_k, the probability that the system is at any random time is related to following events:

(1) System's beginning state is m ($m \leq k$), and its ending state is n ($n \geq k$). There is one departure at the end of the cycle. There are (n-m+1) arrivals during this cycle, which are uniformly distributed over the cycle (a property of the Poisson process [38]). Then the state is k for a time duration c/(n-m+2).

(2) System's beginning state is k, ending state is k-1, and there is no arrival but one departure; then the state stays in k during the whole time c.

Note, these two scenarios cannot be applied to the special case of k = 0, which will be discussed later.

(3) If the beginning state is 0, ending state is 0, there are two possibilities,

(a) No arrival and no departure; system stays in 0 for the whole time c,

(b) Exactly one arrival and one departure; system stays in state 0 for $c/2$.

(4) If the beginning state is 0, ending state is n, $n \geq 1$, there are $(n+1)$ arrivals and one departure during time c; then system stays in state 0 for $c/(n+2)$.

Based on above observations we can get,

$$p_0 = \pi_0 (k_0 + k_1/2 + \cdots + k_i/(i+1) + \cdots), \tag{6.7.1}$$

$$\begin{aligned} p_1 = {} & \pi_0 \cdot (k_1/2 + k_2/3 + k_3/4 + \cdots + k_i/(i+1) + \cdots) + \\ & \pi_1 \cdot (k_0 + k_1/2 + k_2/3 + \cdots + k_i/(i+1) + \cdots) \end{aligned}, \tag{6.7.2}$$

$$\begin{aligned} p_2 = {} & \pi_0 \cdot (k_2/3 + k_3/4 + k_4/5 + \cdots + k_i/(i+1) + \cdots) + \\ & \pi_1 \cdot (k_1/2 + k_2/3 + k_3/4 + \cdots + k_i/(i+1) + \cdots) + \\ & \pi_2 \cdot (k_0 + k_1/2 + k_2/3 + \cdots + k_i/(i+1) + \cdots) \end{aligned}, \tag{6.7.3}$$

Therefore the general formula for p_i is

$$p_i = \sum_{j=0}^{i} \pi_j \cdot \left(\sum_{l=(i-j)}^{\infty} k_l / (l+1) \right). \tag{6.7.4}$$

For the second summation term in (6.7.4) we can prove,

$$\sum_{l=0}^{\infty} k_l / (l+1) = \sum_{l=0}^{\infty} \frac{e^{-\rho} \rho^l}{l! * (l+1)} = \frac{1 - e^{-\rho}}{\rho}. \tag{6.8}$$

Then Equation (6.7.4) can be rewritten as,

$$p_i = \sum_{j=0}^{i} \pi_j \cdot \left(\frac{1-e^{-\rho}}{\rho} - \sum_{l=0}^{j} k_l/(l+1)\right), \tag{6.9}$$

which greatly simplify the calculations.

Using the analytical equations developed above, we can calculate, for different arrival rate λ, the probability that queue size not less than i, which is equal to (1-p_0-⋯-p_{i-1}). This value is the probability of queue size no less than i over the time axis. Comparing this value with the for arrival rates with the measured detector occupancy, one can find the best match to estimate the average vehicle arrival rate.

6.3 Traffic adaptive ramp metering control

In this section we consider a simplified adaptive ramp meter control. The adaptive control shortens the red phase length when queue size reaches a threshold while the duration of green phase remains unchanged. This means the output at the ramp is always one vehicle per cycle, but the cycle length is adjustable. In the model, if the queue size at the end of last cycle (after the departure) is smaller than the threshold, the cycle length is c, otherwise it is c' ($c' < c$).

To illustrate, suppose we set the queue threshold as 2, when the queue length reaches 2 the cycle length decreases. Figure 6.2 illustrates a possible queue length plot of the traffic adaptive ramp control.

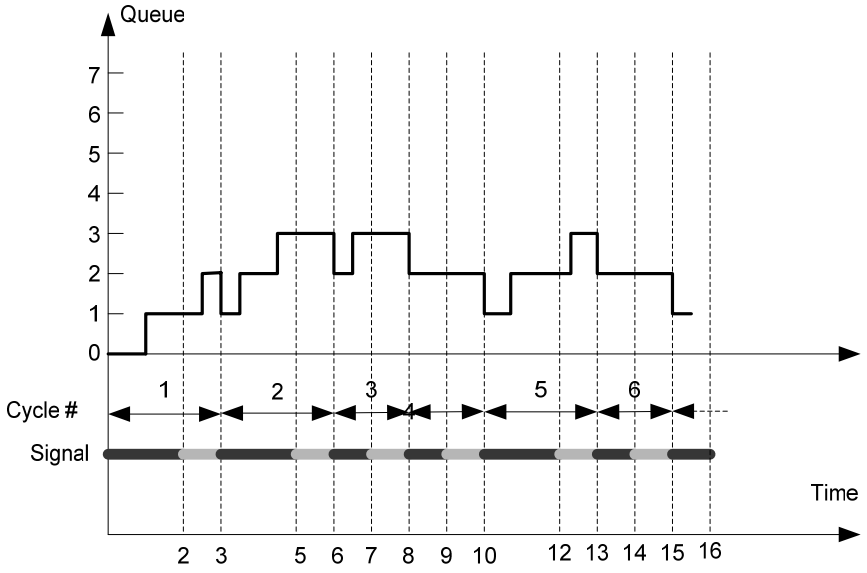


Figure 6.2 Queue length of adaptive ramp meter system

The service rate corresponding to new cycle setting c' is

$$\mu' = 1/c'.$$

Let k_n' =Pr{ n vehicle arrivals during one cycle with length of c'}, then

$$k_n' = \frac{e^{\lambda c'}(\lambda c')^n}{n!}. \tag{6.10}$$

The new traffic intensity is

$$\rho' = \lambda/\mu' = \lambda c'.$$

The generating functions of k' is

$$\mathrm{K}'(z) = \sum_{i=0}^{\infty} k_i' z^i \quad (|z| \le 1).$$

When the queue threshold is 2, the new transition probability matrix is

$$P = \begin{bmatrix} k_0 + k_1 & k_2 & k_3 & k_4 & k_5 & k_6 & \cdots \\ k_0 & k_1 & k_2 & k_3 & k_4 & k_5 & \cdots \\ 0 & k_0' & k_1' & k_2' & k_3' & k_4' & \cdots \\ 0 & 0 & k_0' & k_1' & k_2' & k_3' & \cdots \\ 0 & 0 & 0 & k_0' & k_1' & k_2' & \cdots \\ 0 & 0 & 0 & 0 & k_0' & k_1' & \cdots \\ 0 & 0 & 0 & 0 & 0 & k_0' & \cdots \end{bmatrix}, \tag{6.11}$$

Using similar developments in section 6.2, we have

$$\pi_0 * (k_0 + k_1) + \pi_1 * k_0 = \pi_0, \tag{6.12.1}$$

$$\pi_0 * k_2 + \pi_1 * k_1 + \pi_2 * k_0' = \pi_1, \tag{6.12.2}$$

$$\pi_0 * k_3 + \pi_1 * k_2 + \pi_2 * k_1' + \pi_3 * k_0' = \pi_2, \tag{6.12.3}$$

$$\vdots$$

$$\pi_0 * k_i + \pi_1 * k_{i-1} + \pi_2 * k_{i-2}' + \cdots + \pi_i * k_0' = \pi_{i-1}, \tag{6.12.i}$$

$$\vdots$$

Using z transforms we get

$$\Pi(z) = \frac{[(\pi_0 + z\pi_1) * K(z) + (z-1)k_0] + K'(z) * [\Pi(z) - \pi_0 - z\pi_1]}{z}, \tag{6.13}$$

which can be rewritten as

$$\Pi(z) = \frac{[(\pi_0 + z\pi_1) * K(z) + \pi_0 * k_0 * (z-1)] - K'(z) * (\pi_0 + z\pi_1)}{z - K'(z)}. \quad (6.14)$$

Using the fact that $\Pi(1) = 1$, along with L'Hopital's law, we can get

$$\frac{(\pi_0 + \pi_1) * (\dot{K}(z) - \dot{K}'(z)) + \pi_0 * k_0}{1 - \dot{K}'(z)} |_{z=1} = 1, \quad (6.15)$$

where

$$\dot{K}(1) = \lambda / \mu = \lambda C = \rho,$$

and

$$\dot{K}'(1) = \lambda / \mu' = \lambda C' = \rho'.$$

We then get

$$\pi_0 * (k_0 + \rho - \rho') + \pi_1 * (\rho - \rho') = 1 - \rho'. \quad (6.16)$$

Substitute (6.16) to equation (6.12), we can solve for all values of π_i.

Same as before, let p_i denote the system state distribution over time axis. To solve for p_i we begin with an abstract cycle, whose length can be either c or c′, which is determined by the beginning state of the system (i.e. if the queue length is less than the threshold or not). The expected length of this abstract cycle is,

$$\overline{C} = (\pi_0 + \pi_1) * C + (1 - \pi_0 - \pi_1) * C'. \quad (6.17)$$

Using the method used in section 6.2, we get

$$p_0 * \overline{C} = \pi_0 \cdot (k_0 + k_1/2 + k_2/3 + \cdots + k_i/(i+1) + \cdots) * C, \tag{6.18.1}$$

$$\begin{aligned} p_1 * \overline{C} = {} & \pi_0 \cdot (k_1/2 + k_2/3 + k_3/4 + \cdots + k_i/(i+1) + \cdots) * C + \\ & \pi_1 \cdot (k_0 + k_1/2 + k_2/3 + \cdots + k_i/(i+1) + \cdots) * C \end{aligned}, \tag{6.18.2}$$

$$\begin{aligned} p_2 * \overline{C} = {} & \pi_0 \cdot (k_2/3 + k_3/4 + k_4/5 \cdots + k_i/(i+1) + \cdots) * C + \\ & \pi_1 \cdot (k_1/2 + k_2/3 + k_3/4 + \cdots + k_i/(i+1) + \cdots) * C + \\ & \pi_2 \cdot (k_0' + k_1'/2 + k_2'/3 + \cdots + k_i'/(i+1) + \cdots) * C' \end{aligned}, \tag{6.18.3}$$

$$\begin{aligned} p_i * \overline{C} = {} & \sum_{j=0}^{1} \pi_j \cdot \Big(\sum_{l=(i-j)}^{\infty} k_l/(l+1) \Big) * C \\ & + \sum_{j=2}^{i} \pi_j \cdot \Big(\sum_{l=(i-j)}^{\infty} k_l'/(l+1) \Big) * C' \end{aligned} \quad if \quad i > 2. \tag{6.18.4}$$

With equation (6.8), (6.18.4) can be simplified to

$$\begin{aligned} p_i * \overline{C} = {} & \sum_{j=0}^{1} \pi_j \cdot \Big(\frac{1-e^{-\rho}}{\rho} - \sum_{l=0}^{(i-j)} k_l/(l+1) \Big) * C \\ & + \sum_{j=2}^{i} \pi_j \cdot \Big(\frac{1-e^{-\rho}}{\rho} - \sum_{l=0}^{(i-j)} k_l'/(l+1) \Big) * C' \end{aligned} \quad if \quad i > 2. \tag{6.18.5}$$

As before, now for adaptive ramp meter control, we can calculate the probability queue length is not less than i, which is equal to (1-p_0-$\cdots$-p_{i-1}). Comparing these values for different arrival rate and the measured detector occupancy, one can estimate the underlying vehicle arrival rate.

6.4 Model Evaluation

In this section, the analytical queuing model is compared with simulation-based results. The simulation model of the adaptive control strategy is built in MATLAB. The random number generator of MATLAB was used to generate exponentially distributed random variables for vehicle inter-arrival times. For each scenario, the simulation model was run 10 times, with length of 500,000 seconds each. The first 10,000 seconds of each run were considered warm-up periods and not used in the computation of performance measures.

In the computation experiments, the arrival rate ranges from 0.1 to 0.3 veh/sec for fixed control, with cycle length of 3 seconds. For the adaptive control the arrival rate ranges from 0.15 to 0.35 veh/sec, with queue threshold of 2 vehicles. If queue size is smaller than 2, the cycle length is 3 seconds, otherwise it is 2 seconds.

From the tables we can see the analytical results matches the simulation results very well. The relative difference is less then 1% in most cases.

Queue size	λ= 0.1		λ= 0.2		λ= 0.3	
	Sim	Ana	Sim	Ana	Sim	Ana
=0	0.816	0.816	0.549	0.549	0.164	0.162
≤1	0.973	0.973	0.818	0.818	0.315	0.315
≤2	0.996	0.996	0.929	0.929	0.442	0.443
≤3	0.999	0.999	0.973	0.972	0.550	0.547
≤4	0.999	0.999	0.989	0.989	0.627	0.632
≤5	0.999	0.999	0.996	0.996	0.702	0.701
≤6	N/A	N/A	0.998	0.998	0.754	0.757
≤7	N/A	N/A	0.999	0.999	0.807	0.802
≤8	N/A	N/A	1.000	1.000	0.833	0.839

Note: Ana = analytical results, Sim = simulation results.

Table 6.1 Comparison of steady state distribution for pre-timed control

Queue size	λ= 0.15		λ= 0.25		λ= 0.35	
	Sim	Ana	Sim	Ana	Sim	Ana
=0	0.707	0.707	0.456	0.456	0.222	0.222
≤1	0.937	0.937	0.774	0.774	0.498	0.498
≤2	0.989	0.989	0.922	0.923	0.717	0.717
≤3	0.998	0.998	0.976	0.976	0.851	0.851
≤4	1.000	1.000	0.993	0.993	0.923	0.924
≤5	1.000	1.000	0.998	0.998	0.961	0.961
≤6	1.000	1.000	0.999	0.999	0.980	0.980
≤7	N/A	N/A	1.000	1.000	0.990	0.990
≤8	N/A	N/A	1.000	1.000	0.995	0.995

Table 6.2 Comparison of steady state distribution for adaptive control

6.5 Summary

This chapter developed a queuing model to analyze the queue distributions at ramp meter, for both pre-timed and adaptive control. Queue distribution can be calculated directly by using this method.

Analytical solutions of the model were compared with results from a simulation

model. The comparison shows that the analytical model matches very well the simulation on both control strategies for different traffic intensities.

This model is an approximation of a real ramp meter system; it includes two simplifying assumptions, (1) the adaptive control does not consider the traffic situation on the mainline freeway, and (2) the detector occupancy does not include the vehicles that pass the detector without stopping, but includes only the vehicles that wait in a queue. These two assumptions should be studied in future research. Also, traffic dynamics should be included in future modeling because vehicles approaching a ramp meter slow down significantly before coming to a full stop.

An assumption, and perhaps the major weakness of the model is the assumption of Poisson arrivals with constant arrival rate. We plan to investigate if such a model can be used for Poisson arrivals with time varying rates, as well as other arrival distributions, in particular an arrival pattern that approximates batch (platoon) arrivals.

CHAPTER 7 CONCLUSIONS AND AREAS OF FUTURE RESEARCH

7.1 Overview

Queuing models are developed to analyze signalized intersections and ramp metering systems with different types of control. This chapter summarizes the highlights of each of the pervious chapters of this dissertation. The contributions and possible applications of the methodology are discussed.

7.2 Chapter 3 — Queuing Models for Adaptive Signal Control

Chapter 3 develops analytical queuing models and a methodology to evaluate traffic adaptive signal control. A simplified adaptive control strategy serves one movement until the queue dissipates, then the second movement is served until the queue dissipates, then the signal goes back to serve the first movement and this cyclic process repeats. An iterative numerical algorithm is developed to compute steady-state performance measures such as average delays and expected queue lengths. The analytically method predicts accurately the simulation results.

7.3 Chapter 4 — Queuing Models for Fixed-time Signal Control

The same intersection analyzed in chapter 3 is modeled with fixed-time traffic signal. Numerical algorithms are developed that derive expressions of queue-based

performance measures. The average delays given by the corresponding numerical method are compared with results obtained in chapter 3. The results show that with Poisson arrivals, adaptive signal always reduces the average vehicular delay compared to the fixed-time signal. And the intersection delays with fixed-time control obtained from the queuing model match the results of Webster's formula well.

7.4 Chapter 5 — A Queuing Model for Actuated Signal Control

In chapter 5 a queuing model is developed for eight-phase actuated signal control. First, an algorithm is proposed to calculate the green phase distribution of an actuated movement. Then, for the case of Poisson arrivals, a queuing model is developed for the eight-phase signal control. An algorithm, derived from the queuing model, was developed to compute steady–state performance measures such as average delays and expected queue lengths. The results obtained matched the simulation results fairly well.

7.5 Chapter 6 — A Queuing Model for Ramp Meter Control

When vehicles arrive at a ramp in a Poisson fashion, the number of vehicles in queue can be modeled as a Markov process. The queue size at the end of each cycle is viewed as the system "state", and a state transition matrix can be developed. Queue length distribution at the end of each cycle can be obtained directly by solving the steady-state equations. From this, we can obtain the queue length distribution over

time. The results given by this analytical model matched well the results from a simulation model. Using this distribution, the arrival rate can be estimated from data of an upstream presence detector.

7.6 Contributions

This dissertation has re-visited queuing theory and queuing models in traffic control, and has developed analytical models to evaluate traffic control strategies. Iterative algorithms were developed to solve each of these models. In particular, the methodology was applied to

(a) two-phase adaptive signal control at an isolated intersection,

(b) two-phase fixed-time signal control for an isolated intersection,

(c) eight-phase actuated signal control at an intersection, and

(d) ramp-metering control strategies.

In all cases, the analytical results matched well the results from simulation models.

Although research on analytical queuing analysis for traffic signal has appeared since 1940s [22], this research has mostly addressed fixed-time signal control and has reported results on expected values of performance measures, such as expected delays and expected queue length. Some research has been reported by Newell [36] on an adaptive control which actually inspired some of the dissertation research. No queuing models on ramp meter control have been reported in the literature.

The models and method developed in this research not only provides expected values of the performance measures, but also their distribution. That is, for example, we can compute the distribution of delays and queue lengths. Using these distributions, we can obtain expected values, as well as we can compute reliability measures such as variances of delays. Knowing variances of queue lengths we can design better the "capacities" at on-ramps or the size of left-turn or right-turn bays. Furthermore, we may be able to use these models to design control strategies that are more "robust" in the sense that the probability of excessive delays are made very small. In short, the complete probabilistic distribution of the queues and delays that result at traffic signals can provide several performance measures and give us insight how we may better design and control the traffic arriving at these signals.

7.7 Area of Future Research

We may divide this into two categories: (A) further modeling and analysis of the scenarios addressed in this dissertation, and (B) modeling and analysis of new scenarios.

With regard to category A, we may use the performance measures results to optimize design parameters in the signal control, for example,

(1) minimum and maximum green times in actuated control,

(2) optimal locations of upstream detectors at the on-ramps, and

(3) probabilistic characterization of departure process of vehicles served by the traffic signal, (this output process will define the input process for the downstream traffic signal).

With regard to category B, we may use the modeling approach exploited here to model other scenarios and applications, such as the follows.

(1) Adaptive signal control for two-way traffic. In this case, say, the E-W/W-E through phase switches to N-S/S-N through phase only when both E-W and W-E queues have been cleared. That is, the E-W/W-E busy period is defined by two arrivals processes, which in turn results two queues to be served in N-S/S-N direction respectively,

(2) Four-phase full-actuated intersection. We studied the actuated left-turn control in Chapter 5, four-phase fully actuated isolated intersections can also be analyzed. Although the iterative algorithm may be more complex, the developed approach may be used for this application,

(3) Different arrival streams. This dissertation considered only two arrivals streams, Poisson and deterministic uniform. It may be possible to extend this modeling approach to other distributions that exhibit some Markovian properties, such as the K^{th} order Erlang,

(4) Comparison with realistic simulation models. The MATLAB simulation models were idealized in the sense that vehicle interactive and driver

behavior (e.g. car following behavior) were ignored except for the start loss time. It would be useful to compare our model results with results from a realistic traffic simulation model, such as VISSIM [39] or AIMSUN [40].

REFERENCES

1. D. Schrank, T. Lomax "The 2005 Urban Mobility Report", Texas Transportation Institute, May 2005.
2. J. H. Kell, I. J. Fullerton, "Manual of Traffic Signal Design", Englewood Cliffs, N.J. Prentice-Hall, 1982.
3. R. L. Shaffer, "Traffic Control Solutions: A Light at the End", Advanced Traffic Control, IMSA Journal Feature Article September/October 2002.
4. C. McShane, "The Origins and Globalization of Traffic Control Signals", Journal of Urban History, March 1999.
5. Federal Highway Administration (FHWA), "Manual on Uniform Traffic Control Devices (MUTCD)", Washington DC, 1988.
6. W. R. McShane, R. P. Roess and E. S. Prassas, "Traffic Engineering", 2nd edition, Prentice Hall, 1998.
7. J. E. Moore, R. Jayakrishnan, M. G. McNally, C. A. MacCarley, "Evaluation of the Anaheim Advanced Traffic Control System Field Operational Test: Introduction and Task A: Evaluation of SCOOT Performance", California PATH Research Report, July 1999
8. S. Chiu and S. Chand, "Self-organizing Traffic Control via Fuzzy Logic", Proceedings of 32nd IEEE Conference on Decision and Control, December 1993
9. P. R. Lowrie, "SCATS: The Sydney Coordinated Adaptive Traffic System". Paper presented at the IEEE International conference on road traffic signaling at London, U.K, 1982.
10. L. K. Head, P. B. Mirchandani and S. Shelby, "The RHODES Prototype: A Description and Some Results", Paper presented at the 77th Transportation Research Board Annual Meeting, Washington, D.C., 1998.
11. J. C. Lee, R. H. Wortman, D. Hook, M. J. Poppe, "Operational Comparison of Leading and Lagging Left Turns", Transportation Research Record No. 1421, 1993.
12. C. R. Wright, J. Upchurch, "Before and After Comparison of Leading Exclusive and Permissive/Exclusive Lagging Left Turn Phasing", Transportation Research Record No. 1368, 1992.
13. http://en.wikipedia.org/wiki/Ramp_meter.
14. Cambridge Systems Inc., SRF Consulting Group Inc., N.K. Friedrichs Consulting Inc., "Twin Cities Ramp Meter Evaluation executive summary", February 1, 2001
15. B. Hellinga, M.V. Aerde, "Examining the potential of using ramp metering as a

component of an ATMS", Transportation Research Record, No. 1494, 1995.

16. P. Kachroo, K. Ozbay, "Feedback Ramp Metering in Intelligent Transportation Systems", Springer, 1st edition, December 2003.
17. M. Papageorgiou, J. Blowweville, H. Hadj-Salem, "Modeling and real-time control of traffic flow on the southern part of Boulevard Peripherique in Paris: Part II: Coordinated on-ramp metering", Transportation Research A, Vol. 24A 1990.
18. S. Kang, D. Gillen, "Assessing the Benefits and Costs of Intelligent Transportation Systems: Ramp Meters", California PATH Research Report, UCB-ITS-PRR-99-19, 1999.
19. F. Ciarallo, P. B. Mirchandani, "RHODES-ITMS-MILOS: Ramp Metering System Test", Report FHWA-AZ02-481, University of Arizona, Tucson, Arizona, April 2002.
20. F. V. Webster, "Traffic Signal Settings", Road Research Laboratory Technical Paper No. 39, 1958.
21. Transportation Research Board, National Research Council, "Highway Capacity Manual", TRB, National Research Council, Washington, D.C, 1985.
22. A. J. Clayton, "Road traffic calculations", Journal of the Institution of Civil Engineers, vol. 16, no. 7, 1941.
23. T. M. Matson, W. S. Smith, F. W. Hurd, "Traffic Engineering", McGraw Hill, New York, 1955.
24. A. J. Miller, "Settings for Fixed-Cycle Traffic Signals". Operational Research Quarterly, Vol. 14, 1963.
25. G. F. Newell, "Approximation Methods for Queues with Applications to the Fixed-Cycle Traffic Light", SIAM Review, Vol. 7, 1965.
26. G. F. Newell, "Properties of Vehicle Actuated Signals: I. One-Way Streets", Transportation Since, Vol. 3, 1969.
27. G. F. Newell and E. E. Osuna, "Properties of Vehicle Actuated Signals: II. Two-Way Streets", Transportation Since, Vol. 3, 1969.
28. V. F. Hurdle, "Signalized Intersection Delay Models - A Primer for the Uninitiated", Transportation Research Record No. 971, 1984.
29. R. Akcelik, "Time-dependent Expressions for Delay, Stop Rate and Queue Length at Traffic Signals", Australian Road Research Board, 1980.
30. R. Akcelik, "Traffic Signals: Capacity and Timing Analysis", Australian Road Research Board, 1981.
31. S. Teply, "Canadian Capacity Guide for Signalized Intersections", Edmonton,

Alberta, Canada: Institute of Transportation Engineers–District 7 and The University of Alberta, 1984.

32. R. Akcelik, “The Highway Capacity Manual Delay Formula for Signalized Intersections” ITE Journal 58 (3), 1988.
33. P. S. Olszewski, “Modeling of Queue Probability Distribution at Traffic Signals”, Proceedings of the 11th International Symposium on Transportation and Traffic Theory, Tokyo, Elsevier, New York, 1990.
34. P. S. Olszewski, “Modeling Probability Distribution of Delay at Signalized Intersections”, Journal of Advanced Transportation, Vol. 28, No. 3, 1994.
35. V. Fadhely, C. Kenneth, A. Donald, “Comparison of queue-length models at signalized intersections”, Transportation Research Record No. 1981, 2000.
36. G. F. Newell, “The Rolling Horizon Scheme of Traffic Signal Control”, Transportation Research Part A, Vol. 32, 1998.
37. L. Kleinrock., “Queueing Systems. Volume 1. Theory”, John Wiley & Sons, 1975.
38. S. Ross, “Stochastic Process, 2nd edition”, John Wiley & Sons, 1996.
39. PTV Planung Transport Verkehr AG., “VISSIM User Manual, Version 3.50”, January 2001.
40. TSS-Transport Simulation Systems, “AIMSUN Version 4.1 User’s Manual”, 2002.