Πολυτεχνείο
Κρήτης

Optimal Traffic Signal Control in a Congested Network with Quadratic Programming

Thesis submitted in partial fulfillment of the requirements for the degree of Master of Science

by

Christos Katrinakis

2009019005

Chania, Crete June 2013

INTENTIONALLY LEFT BLANK

The thesis is approved by the following jury:

Markos Papageorgiou (Supervisor), Professor ____

Ioannis Papamichail (Co-supervisor), Assistant Professor ____________________________

Georgios Stavroulakis, Professor ____

Acknowledgements

I would like to thank my supervisor Professor Markos Papageorgiou for his advice, guidance and patience. I would also like to thank my co-supervisor Assistant Professor Ioannis Papamichail for his guidance and suggestions. I am also grateful to Professor Georgios Stavroulakis for participating in my thesis committee.

I am grateful to my fellow researchers, Dr. Kostas Aboudolas and Mr. Medhi Keyvan Ekbatani for their valuable contribution, support and patience, which made the accomplishment of this thesis possible.

Finally, I would like to thank my friends for their help and support throughout my studies.

Contents

List of Figures

Abstract

Traffic signal control for urban road networks has been an area of intensive research efforts for several decades. Despite the long-lasting research, the various algorithms and tools that have been developed and implemented to increase the network traffic flow efficiency, urban signal control is still an area susceptible of further significant improvements, particularly under saturated traffic conditions.

The purpose of this thesis is to investigate the efficiency of a recently developed computationally feasible technique for real-time network-wide signal control in large-scale urban traffic networks, the rolling-horizon quadratic programming control (QPC), which aims at minimizing and balancing the link queues so as to minimize the risk of queue spillback. The control strategy's efficiency and real-time feasibility is demonstrated and compared to optimized fixed-time control settings and Gating, a new idea of holding traffic back (via prolonged red phases at traffic signals) upstream of the links to be protected from oversaturation, via their simulation-based application to the road network of the city center of Chania, Greece, under a number of different demand scenarios. The comparative evaluation is based on various criteria and tools including the recently proposed fundamental diagram for urban network traffic.

The results show, with no doubt, that in average both real time adaptive compared strategies give much better results from the fixed-time option. Gating, despite its simplicity, excels the QPC strategy, in all the evaluation criteria, but a number of questions are still to be answered in order to derive indisputable conclusions. This type of questions emerged during this conduct of research but further work needs to be done in a future research, in order to be explored.

Περίληψη

Ο έλεγχος φωτεινής σηματοδότησης για αστικά οδικά δίκτυα αποτελεί ένα επιστημονικό πεδίο με εντατικές ερευνητικές προσπάθειες για αρκετές δεκαετίες. Παρά την μακροχρόνια έρευνα, τους διάφορους αλγόριθμους και τα εργαλεία που έχουν αναπτυχθεί και εφαρμοστεί, ώστε να αυξηθεί η κυκλοφοριακή ροή στο δίκτυο, ο έλεγχος φωτεινής σηματοδότησης σε αστικές περιοχές εξακολουθεί να είναι ένας χώρος, όπου μπορούν να υπάρξουν περαιτέρω βελτιώσεις, ιδίως σε συνθήκες κυκλοφοριακής συμφόρησης.

Ο σκοπός αυτής της διατριβής είναι η διερεύνηση της αποτελεσματικότητας μιας πρόσφατα αναπτυγμένης υπολογιστικά εφικτής τεχνικής ελέγχου φωτεινής σηματοδότησης σε πραγματικό χρόνο σε μεγάλης κλίμακας αστικά δίκτυα κυκλοφορίας - η τεχνική του τετραγωνικού προγραμματισμού κυλιόμενου ορίζοντα (QPC) - η οποία αποσκοπεί στην ελαχιστοποίηση και εξισορρόπηση των ουρών των οχημάτων σε κάθε σύνδεσμο, έτσι ώστε να ελαχιστοποιηθεί ο κίνδυνος συμφόρησης.

Η αποτελεσματικότητα και η καταλληλότητα της στρατηγικής ελέγχου σε πραγματικό χρόνο αναλύεται σε σύγκριση με τις βελτιστοποιημένες ρυθμίσεις ελέγχου φωτεινής σηματοδότησης σταθερού χρόνου και με την στρατηγική Gating, μιας νέας ιδέας του περιορισμού της κυκλοφορίας «ανάντη» (μέσω παρατεταμένων κόκκινων φάσεων στους φωτεινούς σηματοδότες) των συνδέσμων, οι οποίοι πρέπει να προστατευθούν από τη συμφόρηση. Η συγκριτική αξιολόγηση γίνεται μέσω προσομοιωμένης εφαρμογής των στρατηγικών στο οδικό δίκτυο του κέντρου της πόλης των Χανίων, στην Ελλάδα, με μια σειρά από διαφορετικά σενάρια ζήτησης και βασίζεται σε διάφορα κριτήρια και εργαλεία συμπεριλαμβανομένου και του βασικού διαγράμματος για την αστική κυκλοφορία του δικτύου, που προτάθηκε πρόσφατα.

Τα αποτελέσματα δείχνουν, χωρίς αμφιβολία, ότι κατά μέσο όρο και οι δύο στρατηγικές πραγματικού χρόνου δίνουν καλύτερα αποτελέσματα σε σύγκριση με τη επιλογή σταθερού χρόνου. Η στρατηγική Gating, παρά την απλότητα, υπερέχει της QPC στρατηγικής, σε όλα τα κριτήρια αξιολόγησης, αλλά μια σειρά από ερωτήματα πρέπει να απαντηθούν προκειμένου να εξαχθούν αδιαμφισβήτητα συμπεράσματα. Αυτού του είδους οι προβληματισμοί αναδείχθηκαν κατά τη διάρκεια της παρούσας έρευνας και χρειάζεται να γίνει μια περαιτέρω εργασία σε μια μελλοντική έρευνα, προκειμένου να διερευνηθούν.

Thesis outline

The organization of the thesis is as follows:

Chapter 1 introduces the need of traffic control in modern society and describes the scope of the comparison of the two traffic control strategies that it is made in this thesis.

Chapter 2 defines the basic concepts for the control of traffic networks through traffic lights and presents a review of advanced urban traffic control. Particular emphasis is given to coordinated real-time systems. Finally, a description of some basic notion is given for the two real-time adaptive control strategies.

Chapter 3 describes the recently introduced traffic control strategy of Gating, its basic theory and the mathematic formulation of the control model.

Chapter 4 describes the basic theory of traffic control with Quadratic Programming, the modification of the problem and the advantages that it provides.

Chapter 5 introduces the field of implementation of the three compared strategies via simulation in the center of the city of Chania and describes some features of the replications that were used. The results are presented and some difficulties that were met are explained.

Chapter 6 presents the conclusions for the three strategies implementation in simulation and the interpretation of the results. It also describes some guidelines on which future work should focus and gives some motives for future research.

1 Introduction

The purpose of this chapter is to mention the need of traffic control in large scale urban networks especially in congested condition. It also defines the scope of this thesis, which aims to find the most feasible technic by comparing two recently developed strategies in the same implementation field.

1.1 The transportation problem

Transportation has always been an important aspect of human civilization, but it is only in the last decades that the phenomenon of traffic congestion has become predominant due to the rapid increase in the number of vehicles and in the transportation demand in virtually all transportation modes. Traffic congestion is a condition on road networks that occurs as use increases, and is characterized by slower speeds, longer trip times, and increased vehicular queuing. Also when vehicles are fully stopped for periods of time, this is colloquially known as traffic jam or traffic snarl-up. It appears when too many vehicles attempt to use a common transportation infrastructure with limited capacity. In the best case, traffic congestion leads to queuing phenomena (and corresponding delays) while the infrastructure capacity ("the server") is fully utilized. In the worst (and far more typical) case, traffic congestion leads to a degraded use of the available infrastructure (reduced throughput that may even lead to a fatal gridlock) with excess delays, reduced safety, and, recently, increased environmental pollution [1].

The increased importance of environmental concerns and the limited economic and physical resources are the most important reasons for which a brute-force approach (i.e., the continuous expansion of the available transportation infrastructure) cannot continue to be the only answer to the ever increasing transportation and mobility needs of modern societies. The efficient, safe, and less polluting transportation of persons and goods calls for an optimal utilization of the available infrastructure via suitable application of a variety of traffic control measures. The rapid developments in the areas of communications and computing played an important role, but it is quite evident that the efficiency of traffic control directly depends mostly on the employed control methodologies [1].

1.2 Scope of the study

As mentioned above traffic congestion in urban road networks is a persisting or even increasing problem of modern society. Congestion can be reduced either by increasing road capacity (supply), or by reducing traffic demand. On the supply side, the provision of new infrastructure is usually not a feasible solution. So it is necessary to focus on a better utilization of the existing infrastructure (e.g. via traffic management), to mitigate congestion and improve urban mobility.

The field of urban traffic control (UTC) has been studied and developed in a variety of ways during the past decades. In fact, the traffic flow conditions in large-scale urban networks depend critically on the applied signal control strategies. It is well known that nowadays, the negative effects of congested transport networks persist or even increase. This is also obvious in urban networks by the excessive delays and wasting time of motorists and passengers, the environmental impact of increased wasted fuel and the reduced safety by higher chance of collisions due to tight spacing and constant stopping-and-going. Introducing improved traffic signal control methods and techniques continues to be a vital issue. In particular, the development of practicable and efficient real-time signal control strategies for urban road networks under saturated traffic conditions is a major challenge with significant scientific and practical relevance. The scientific relevance stems from the increased interest in the specific problem as well as recent, potentially valuable, models and insights that may contribute to improved signal control methods. The practical relevance stems from the congestion, degradation and gridlock problems encountered increasingly in modern urban road networks that could benefit highly from improved signal control under saturated traffic conditions [2].

This thesis purpose is to investigate the efficiency of a recently developed computationally feasible technique for real-time network-wide signal control in large-scale urban traffic networks, the rolling-horizon quadratic programming control (QPC) and compare it to optimized fixed-time control settings and gating, a new idea of holding traffic back (via prolonged red phases at traffic signals) upstream of the links to be protected from oversaturation. This comparison is done via their simulation-based application to the urban network of the city center of Chania, Greece, under a number of different demand scenarios. The comparative evaluation is based on various criteria and tools including the recently proposed fundamental diagram for urban network traffic.

2 Urban Traffic Control

The purpose of this chapter is to define the basic concepts related to the control of traffic networks through traffic lights and review the major advanced urban traffic control. It focuses primarily on the coordinated control systems and presents the main advantages and disadvantages of the most popular systems that are currently used worldwide.

2.1 The need of traffic control

The emergence of traffic (i.e. many interacting vehicles using a common infrastructure) and subsequently traffic congestion (whereby demand exceeds the infrastructure capacity) have opened new innovation needs in the transportation area. City planning and urban design practices can have a huge impact on levels of future traffic congestion.

Nowadays, the increased importance of environmental concerns and the limited economic and physical resources are among the most important reasons why a brute-force approach (i.e. the continuous expansion of the available transportation infrastructure) cannot continue to be the only answer to the ever increasing transportation and mobility needs of modern societies. The efficient, safe, and less polluting transportation of persons and goods calls for an optimal utilization of the available infrastructure via suitable application of a variety of traffic control measures. Very important impacts have the rapid developments in the areas of communications and computing, but it is quite evident that the efficiency of traffic control directly depends mostly on the employed control methodologies [1].

2.2 Control theory

2.2.1 The control loop

The modern intelligent transportation systems are based on the fundamental theory of control (as detailed in [1]). Figure 2-1 illustrates the basic elements of a control loop. The traffic flow behavior in the (road or freeway or mixed) network depends on some external quantities that are classified into two groups:

- Control inputs that are directly related to corresponding control devices such as traffic lights, variable message signs, etc.; the control inputs may be selected from an admissible control region subject to technical, physical, and operational constraints.
- Disturbances, whose values cannot be manipulated, but may possibly be measurable (e.g. demand) or detectable (e.g. incident) or predictable over a future time horizon.

The network's output or performance is measured via suitable indices, such as the total time spent (TTS) by all vehicles in the network over a time horizon. The task of the surveillance is to enhance and to extend the information measurement devices (e.g. loop detectors, cameras) as required by the subsequent control strategy and the human operators. The kernel of the control loop is the control strategy, whose task is to specify in real time the control inputs, based on available measurements/ estimations/predictions, so as to achieve the pre-specified goals (e.g. minimization of TTS) despite the influence of various disturbances. If this task is undertaken by a human operator, we have a manual control system. In an automatic control system, this task is undertaken by an algorithm (the control strategy). The relevance and efficiency of the control strategy largely determines the efficiency of the overall control system. Therefore, whenever possible, control strategies should be designed with care, via application of powerful and systematic methods of optimization and automatic control, rather than via questionable heuristics [1].

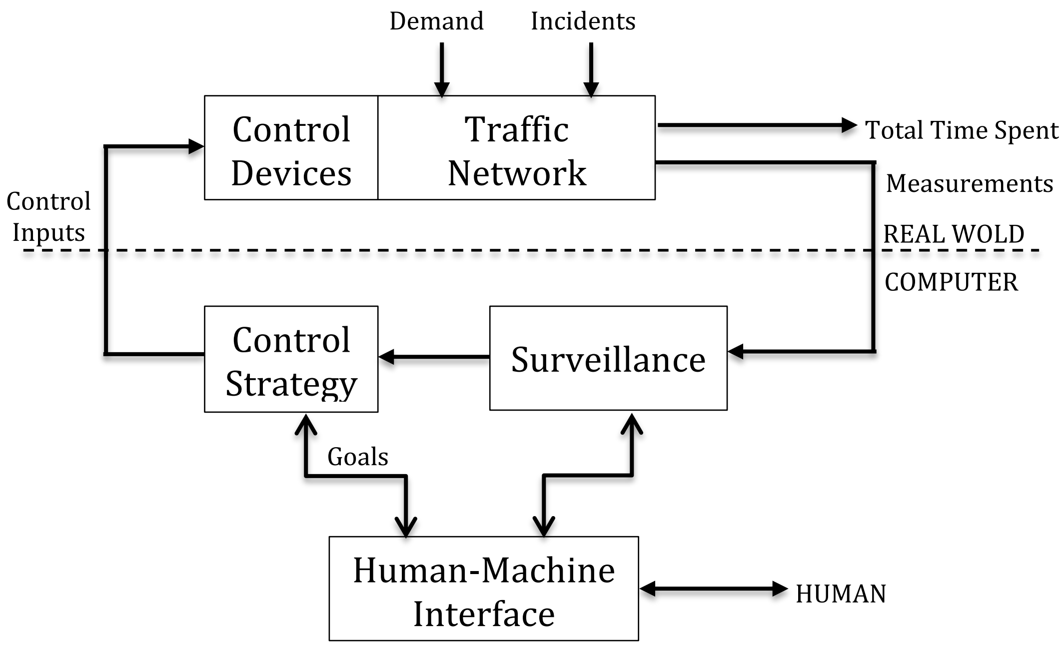


Figure 2-1: The control loop.

2.2.2 A basic property

To explain some basic notions we will use a discrete-time representation of traffic variables with discrete time index $k = 0,1,2 \ldots$ and time interval T. A traffic volume or flow $q(k)$ (in veh/h) is defined as the number of vehicles crossing a corresponding location during the time period $[kT, (k+1)T]$, divided by T. Traffic density $\rho(k)$ (in veh/km) is the number of vehicles included in a road segment of length Δ at time kT, divided by Δ. Mean speed $v(k)$ (in km/h) is the average speed at time kT of all vehicles included in a road segment.

We consider a traffic network Figure 2-2 that receives demands $d_i(k)$ (in veh/h) at its origins $i = 0,1,2 \ldots$ and we define the total demand $d(k) = d_1(k) + d_2(k) + \ldots$. We assume that $d(k),\ k = 0, \ldots, K$ is independent of any control measures taken in the network. We define exit flows $s_i(k)$ at the network destinations $i = 0,1,2 \ldots$, and the total exit flow $s(k) = s_1(k) + s_2(k) + \ldots$ We wish to apply control measures so as to minimize the total time spent T_s in the network over a time horizon K, i.e.

$$T_s = T \sum_{k=0}^{K} N(k) \qquad \text{(2-1)}$$

where $N(k)$ is the total number of vehicles in the network at time k. Due to conservation of vehicles

$$N(k) = N(k-1) + T[d(k) - s(k)] \qquad \text{(2-2)}$$

hence

$$N(k) = N(0) + T \sum_{k=0}^{K} [d(k) - s(k)] \qquad \text{(2-3)}$$

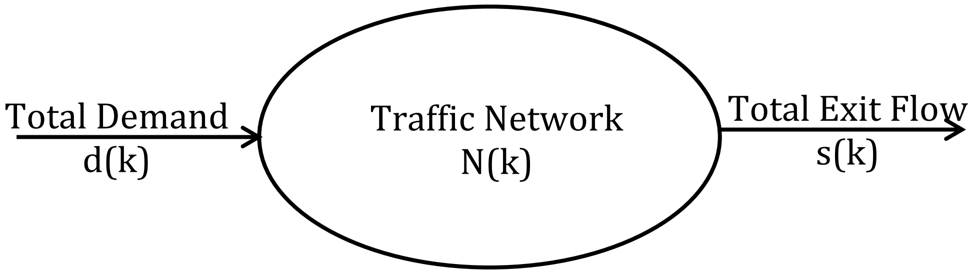


Figure 2-2: A traffic network.

Substituting (2-3) in (2-1) we obtain

$$T_s = T \sum_{k=0}^{K} \left[N(0) + T \sum_{k=0}^{k-1} d(k) - T \sum_{k=0}^{k-1} s(k) \right] \qquad \text{(2-4)}$$

The first two terms in the outer sum of (2-4) are independent of the control measures taken in the network, hence minimization of is equivalent to maximization of the following quantity

$$S = T^2 \sum_{k=0}^{K} \sum_{k=0}^{k-1} s(k) = T^2 \sum_{k=0}^{K-1} (K-k) s(k) \qquad \text{(2-5)}$$

Thus, minimization of the total time spent in a traffic network is equivalent to maximization of the time-weighted exit flows. In other words, the earlier the vehicles are able to exit the network (by appropriate use of the available control measures) the less time they will have spent in the network [1].

2.3 Road traffic control

2.3.1 Basic notions

The basic principles and concepts of managing traffic control are detailed in [1]. Traffic lights at intersections were originally installed in order to guarantee the safe crossing of antagonistic streams of vehicles and pedestrians but there are also the major control measure in road networks. With steadily increasing traffic demands, it was soon realized that once traffic lights exist, they may also lead (under equally safe traffic conditions) to more or less efficient network operations, hence there must exist an optimal control strategy leading to minimization of the total time spent by all vehicles in the network.

Although the corresponding optimal control problem may be readily formulated for any road network, its real-time solution and realization in a control loop like the one of Figure 2-1 faces a number of apparently insurmountable difficulties:

- The red-green switching of traffic lights call for the introduction of binary variables, which renders the optimization problem combinatorial.
- The size of the problem for a whole network is very large.
- Many unpredictable and hardly measurable disturbances (incidents, illegal parking, pedestrian crossings, intersection blocking, etc.) may perturb the traffic flow.
- Measurements of traffic conditions are mostly local (via loop detectors) and highly noisy due to various physical effects.
- There are tight real-time constraints, e.g. decision making within 2s for advanced control systems.

All of these difficulties render the solution of a detailed optimal control problem infeasible for more than one intersection. So, proposed control strategies for road traffic control introduce a number of simplifications of different kinds or address only a part of the related traffic control problems.

An intersection consists of a number of approaches and the crossing area. An approach may have one or more lanes but has a unique, independent queue. Approaches are used by corresponding traffic streams. A *saturation flow* S (veh/h) is the average flow crossing the stop line of an approach when the corresponding stream has right of way (r.o.w.) and the upstream demand (or the waiting queue) is sufficiently large. Two *compatible* streams can safely cross the intersection simultaneously, else they are called *antagonistic*. A *signal cycle* is one repetition of the basic series of signal combinations at an intersection; its duration is called *cycle time* C. A *stage* (or *phase*) is a part of the signal cycle, during which one set of streams has r.o.w. Constant lost times L of a few seconds are necessary between stages to avoid interference between antagonistic streams of consecutive stages (Figure 2-3).

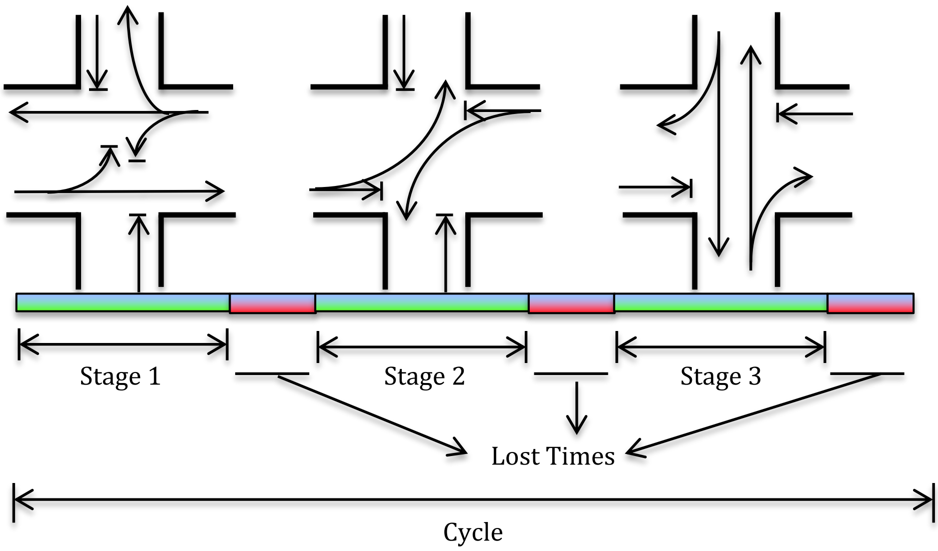


Figure 2-3: Example of signal cycle.

- Stage specification: For complex intersections involving a large number of streams, the specification of the optimal number and constitution of stages is a non-trivial task that can have a major impact on intersection capacity and efficiency.
- Split: This is the relative green duration of each stage (as a portion of the cycle time) that should be optimized according to the demand of the involved streams.
- Cycle time: Longer cycle times typically increase the intersection capacity because the proportion of the constant lost times becomes accordingly smaller; on the other hand, longer cycle times may increase vehicle delays under saturated intersections due to longer waiting times during the red phase.
- Offset: This is the time difference between cycles for successive intersections that may give rise to a "green wave" along an arterial; clearly, the specification of offset should ideally take into account the possible existence of vehicle queues.

2.3.2 Methods of traffic control with traffic lights

The control of traffic light signals can be accomplished in one of four different ways:

- The first way is to determine the number and composition of the phases in each traffic junction (i.e. the selection of movements that take green or red to each phase). The number of phases in a signalized node determined by the number and size of the traffic flows of vehicles and pedestrians to be served

- The second and perhaps most important way is to identify the relative duration of each phase. The relative duration equal to the duration of the phase divided by the period signal.
- The third way is to change the duration of the signal period. Increasing the length of time usually results in steady, increase traffic capacity of the junction due to the relative decrease in the transitional stage. However, long periods create greater delays.
- The fourth way refers only to coordinated systems and to define their temporal displacement between the offset of periods of adjacent intersections. The correct setting of the time shift is essential for the creation of so-called "green wave" along a highway.

2.3.3 Classification of traffic control strategies

Control strategies employed for road traffic control may be classified according to the following characteristics:

- *Fixed-time* strategies use historical data in order to specify, off-line, optimal time-of-day-dependent plans for the traffic lights; *traffic-responsive* strategies use real-time measurements in order to specify in real time suitable signal settings.
- *Isolated* strategies are applicable to single intersections while *coordinated* strategies consider an urban zone or even a whole network comprising many intersections.
- Some strategies are only applicable to *under saturated* traffic conditions, whereby vehicle queues are only created during the red phases and are dissolved during the green phases; other strategies are adapted also for *oversaturated* conditions with partially increasing queues that in some cases may even reach the upstream intersection.

The major drawback of fixed time strategies stems from the fact that these regulations are based on historical measurements and not on real time data. This simplification reduces the effectiveness of the fixed time because:

- The demand is not constant, and can vary from day to day or at different times of year.
- Demand changes and leads to long term "aging" of the optimal settings.
- The turn rates can also change in the same way as demand.
- Incidents and further disturbances can disrupt the prevailing traffic conditions in such a way that it is not possible to predict.

Traffic-responsive systems, if properly designed, are more efficiently, as they face all the above drawbacks. However, they are also more expensive, since they require installation, operation and maintenance of a system which operates in real time and includes measurement devices, communication, local controllers and particularly in the case of the coordinated control, a central control room [1].

2.4 Review of advanced traffic control strategies

2.4.1 Isolated intersection control

Fixed-time strategies: Isolated fixed-time strategies are only applicable to under saturated traffic conditions. *Stage-based* strategies under this class determine the optimal splits and cycle time so as to minimize the total delay or maximize the intersection capacity. *Phase-based* strategies determine not only optimal splits and cycle time but also the optimal staging, which may be an important feature for complex intersections. Well known examples of stage-based strategies are SIGSET (SIGnal SETtings) [3] and SIGCAP (SIGnal CAPacity) [4]. A nonlinear total delay function derived for under saturated conditions is used in SIGSET as an optimization objective and so it solves a linearly constrained nonlinear programming problem to minimize the total intersection delay for given stream demands. On the other hand, SIGCAP maximizes the intersection's capacity (multiplying the real demand by a factor μ) leading to a linear programming problem [1]. Phase-based approaches solve a similar problem, suitably extended to consider different staging combinations. Phase-based approaches consider the compatibility relations of involved streams as pre-specified and deliver the optimal staging, splits, and cycle time, so as to minimize total delay or maximize the intersection capacity. The related computation time is naturally much higher than for stage-based approaches, but this is of minor importance, as calculations are performed offline.

Traffic-responsive strategies: Isolated, traffic-responsive strategies make use of real-time measurements provided by inductive loop detectors that are usually located some 40m upstream of the stop line, to execute some more or less sophisticated vehicle-actuation logic. One of the simplest strategies under this class is the vehicle-interval method that is applicable to two-stage intersections. Minimum-green durations are assigned to both stages. If no vehicle passes the related detectors during the minimum green of a stage, the strategy proceeds to the next stage. If a vehicle is detected, a critical interval (CI) is created, during which any detected vehicle leads to a green prolongation that allows the vehicle to cross the intersection. If no vehicle is detected during CI, the strategy proceeds to the next stage, else a new CI is created, and so forth, until a pre-specified maximum-green value is reached. An extension of the method also considers the traffic demand on the antagonistic approaches to decide whether to proceed to the next stage or not [1]. A more sophisticated version of this kind of strategy is MOVA (Microprocessor Optimized Vehicle Actuation) [5].

2.4.2 Fixed-time coordinated control

The most popular representatives of this class of strategies for urban networks are outlined below. By their nature, fixed-time strategies are only applicable to under saturated traffic conditions.

MAXBAND [6] considers a two-way arterial including several subsequent signals (intersections) and specifies the corresponding offsets so as to maximize the number of vehicles that can travel within a given speed range without stopping at any signal (green wave). A number of significant extensions have been introduced in the original method in order to consider a variety of new aspects such as different bandwidths for each link of the arterial (MULTIBAND) [7].

TRANSYT (TRAffic Network StudY Tool) is the most known and most frequently applied signal control strategy, and it is often used as a reference method to test improvements enabled by real-time strategies. The procedure is an iterative one: For given values of the decision variables (control inputs), i.e. of splits, offsets and cycle time, the dynamic network model calculates the corresponding performance index, e.g. the total number of vehicle stops. A heuristic "hill-climb" optimization algorithm introduces small changes to the decision variables and orders a new model run, and so forth, until a (local) minimum is found [8].

2.4.3 Traffic-responsive coordinated strategies

The extensive worldwide research in the area of traffic control systems in real time resulted in the development of a number of control strategies which adopt different design philosophy and have various common and distinguishing characteristics. Generally, these systems achieve better average travel time on the network from 0 to 20%. However, it has been found that the efficiency is reduced substantially under conditions of traffic saturation, and in some cases lead to the blocking part of the network (gridlocks).

The first urban traffic control strategies in real time appeared in the 1980's with the development and implementation are the British system SCOOT (Split Cycle and Offset Optimization Technique) [9] and the Australian system SCATS (Sydney Coordinated Adaptive Traffic System) [10]. Both SCOOT and SCATS aim in the coordinated control of urban networks.

SCATS adopt a hierarchical structure in the implementation of control in which a higher level is responsible for coordination at the network level, while at a lower level modifies the signaling each node individually to meet the prevailing local traffic conditions. It attempts to equalize the degree of saturation (DS), i.e., the ratio of effectively used green time to the total green time, for all the approaches. The computation of cycle length and split plan is only carried out at the critical junctions. Cycle length and split plan at non-critical junctions are controlled by the critical junctions via offsets. The algorithm involves many parameters, which need to be properly calibrated for each critical junction. In addition, all the possible split plans need to be pre-specified and a voting scheme is used in order to select a split plan that leads to approximately equal DS for all the approaches [11]. SCATS calculates cycle length, splits and offsets cycle-by-cycle and dynamically changes the grouping of signals in as traffic changes. It has been distributed to 141 cities worldwide controlling over 31,700 intersections (SCATS, 2010) and it has been successfully deployed on arterial roads, downtown grid networks, and at small groups of intersections. There are 14 deployments in USA, ranging in size from 11 signals up to 625 in Oakland County, MI [12].

SCOOT was originally designed to control dense urban networks, such as large towns and cities. It is also successful in small networks, especially for areas where traffic patterns are unpredictable. Through SCOOT, by using the saturation level of the highest loaded intersection, the cycle time within the network is determined and through the saturation level of the entries, the green time at the junction is determined. Merely the optimization of the offset is based on a dynamic traffic model, with which the waiting period can be minimized consecutively via the observation of partial networks ("Mini-Areas") at all junctions. The benefits were evaluated by large area field trials in five cities in the UK and three other cities in the USA, respectively Canada (Oxnard, Red Dear, and Toronto), in which the method was compared to an optimized fixed time control. SCOOT achieved average delay savings of about 12% [13]. There are over 200 SCOOT systems worldwide working in large congested cities, small towns and around freeway interchanges (SCOOT-UTC, 2010). There are ten SCOOT installations in North America. SCOOT continually calculates the required coordination pattern for a group of signals in real time and immediately implements the changes.

Both SCOOT and the SCATS operate in real-time small changes in relative duration of phases during the signal period and the offset start period of successive nodes. The result is considered by some to be inadequate, especially in the rapidly changing traffic conditions during peak hours or in the case of events that reduce the traffic capacity of the regulated system. This is probably also one of the reasons why despite the long development, exploration and application, these systems still behave sometimes better and sometimes worse than traditional fixed time control systems.

A number of other advanced urban traffic control systems have also been proposed. PRODYN, uses a set of non-linear discrete time state equations to model traffic. It takes into account the current controller sequence, the time that has been running for each intersection and the loss of capacity on non-priority streams (left turning), the queue length and the number of vehicles on each section of link. Each intersection has its own optimization module. In the case of links inside the network the optimization modules of upstream intersections inform downstream intersection optimization modules of forecast arrivals for the next 75 seconds. Queue lengths in each link are estimated on the basis of the flows [14].

OPAC (Optimized Policies for Adaptive Control) is a real-time demand-responsive traffic signal timing optimization algorithm for individual intersections. OPAC distinguishes itself from traditional cycle-split signal control strategies by dropping the concept of cycle. In OPAC, the signal control problem consists of a sequence of switching decisions made at fixed time intervals. At each decision point the question is whether to extend or terminate current phase. Dynamic programming techniques are used to calculate optimal solutions. OPAC utilizes on-line data obtained from upstream detectors as well as historical data in the optimization. The objective is to minimize performance measures, such as vehicle delays and stops. Each phase is constrained only by the minimum and maximum phase lengths. Consequently, the duration of a phase is never pre-specified. It depends solely on the prevailing traffic flow conditions. The dynamic optimization process is carried out continuously to ensure that the signal control is always up to date [15].

CRONOS shares certain features with PRODYN (or OPAC), but uses a different algorithm. The system makes real time decisions about whether to retain or change the green phases at the intersections in a network. The advantage of this solution is that the algorithm's run time is fast (one second) and the time needed increases in direct polynomially with the number of links and nodes. The forecasting module predicts, for a given time horizon, the future vehicle arrivals on each link entering the zone. This prediction is based on a rolling average of the arrivals in the past; it is used by the modeling module which calculates the value of a chosen traffic criterion for a given sequence of traffic signal states (colors) over the time horizon. These states are provided by an optimization module, which looks for the best sequence which minimizes the traffic criterion. When this sequence is found, the corresponding traffic signal states are applied on the intersection for the next time step, and the whole process is activated again one time step later. The optimized traffic criterion is the total delay on the zone over the time horizon [16].

UTOPIA (Urban Traffic Optimization by Integrated Automation) was designed to apply to large scale systems. The global approach was to decompose the whole control problem in a hierarchical decentralized way, define proper functional for the resulting problems, together with rules for their interaction and define techniques and algorithms for solving these problems. The Control aims to minimize the total time lost to private vehicles, subject to the constraint that public vehicles with priority shall not be stopped at traffic lights [17].

2.5 Store-and-forward modeling

Store-and-forward modeling of traffic networks was first suggested by Gazis and Potts (1963) and has since been used in various works notably for road traffic control. This modeling approach that describes the network traffic flow process in a simplified way, so as to circumvent the inclusion of discrete variables, offers a major advantage: it allows for highly efficient optimization and control methods with polynomial complexity to be used for the coordinated control of large-scale congested urban networks. On the other hand, the introduced modeling simplification allows only for split optimization, while cycle time and offsets must be delivered by other control algorithms.

The main idea when using store-and-forward models for road traffic control is to introduce a model simplification that enables the mathematical description of the traffic flow process without use of binary variables. This is of paramount importance because it opens the way to the application of a number of highly efficient optimization programming, nonlinear and control methods (such as linear programming, quadratic programming, and multivariable regulators) with polynomial complexity, which, on its turn, allows for coordinated control of large-scale networks in real time.

The critical simplification is introduced when modeling the outflow u_i of a stream i . Assuming sufficient demand on the link, the outflow u_i at discrete time k is set

$$u_i = \frac{g_i(k)}{c} s_i \tag{2-6}$$

where $g_i$is the green time duration for this stream and s_i is the corresponding saturation flow. If the time step T is equal to the cycle time c, Figure 2-4 illustrates that u_i in (2-6) is equal to the average flow during the corresponding cycle, rather than equal to s_i during the green phase and equal to zero during the red phase. In other words, (2-6) suggests that there is a continuous (uninterrupted) outflow from each network link (as long as there is sufficient demand). The consequences of this simplification are:

- The time step T of the discrete-time representation cannot be shorter than the cycle time c, hence real-time decisions cannot be taken more frequently than at every cycle.
- The oscillations of vehicle queues in the links due to green/red-commutations are not described by the model.
- The effect of offset for consecutive intersections cannot be described by the model.

Despite these consequences, the appropriate use of store-and-forward models may lead to efficient coordinated control strategies for large-scale networks as demonstrated in simulation studies in some of the aforementioned references [1].

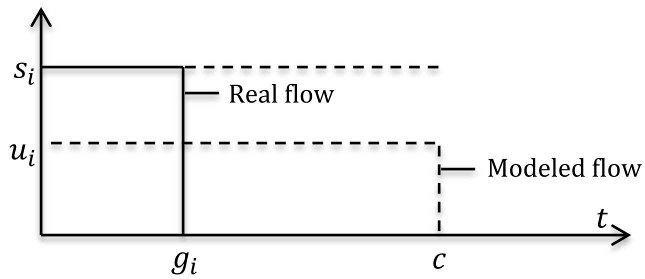


Figure 2-4: Simplified modeling of link outflow u_i.

2.6 The fundamental diagram

Recent experimental analysis has shown that some types of urban networks exhibit a low scatter reproducible relationship between average network flow and density, known as the macroscopic fundamental diagram (MFD). It has also been shown that heterogeneity in the spatial distribution of density can significantly decrease the network flow for the same value of density. Analytical theories have been developed to explore the connection between network structure and an MFD for urban neighborhoods with cars controlled by traffic signals [18].

The notion of a fundamental diagram (e.g. in the form of a flow-density curve) for highways was recently found to apply (under certain conditions) to two-dimensional urban road networks [19]. In fact, a fundamental-diagram-like shape of measurement points was first presented by Godfrey (1969) [20], but also observed in a field evaluation study by Dinopoulou et al. (2005) [21]. The concept is sometimes called MFD (macroscopic fundamental diagram), but since the ordinary fundamental diagram (for highways) is also macroscopic, we prefer to call it NFD (network fundamental diagram) for better distinction.

Figure 2-5 illustrates the typical shape of a fundamental diagram for urban networks, where the vertical axis represents the total flow in the network (the sum of flows leaving the links in the network) or the total flow of vehicles approaching respective destinations of the network, while the horizontal axis represents the number of vehicles within the network. In the case of highways, the fundamental diagram is the result of the network infrastructure, the capabilities of vehicles and driver's behavior, but may also be influenced by the effect of a control measure such as variable speed limits [1].

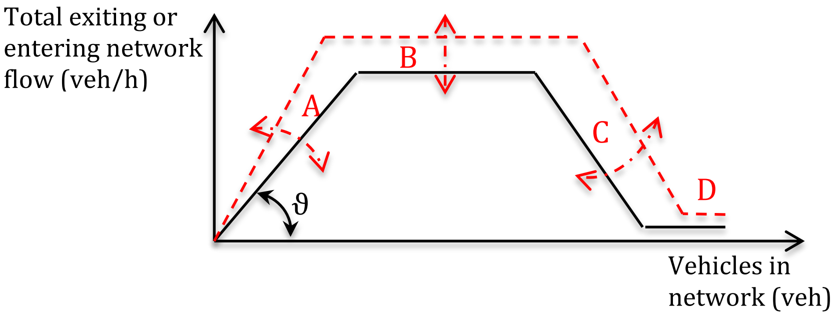


Figure 2-5: The fundamental diagram.

In the case of urban networks, the fundamental diagram also depends on the form of traffic volume (departure-destination routing or vehicles) as well as on the operation of traffic lights. Therefore, considering that traffic volume for specific periods of time is comparable from day to day, the fundamental diagram for urban networks can be used for evaluating different traffic control strategies.

In Figure 2-5, the traffic state on the solid line A represents unsaturated traffic conditions (where the vehicles waiting at signalized junctions are served during the next green stage), with a part of green times partly lost due to reduced traffic demand.

Note that the slope of the line A ($\tan \vartheta$) is proportional to the average speed of vehicles in the urban network. The average speed of vehicles in the urban network can be affected (as shown by the arrows in Figure 2-5) by applying different control strategies of traffic lights (green light relative duration, time period, offset of periods of adjacent intersections). The traffic situation in horizontal line B represents the network's capacity (maximum flow) which may also be affected by the application of different control at traffic lights. Note that the maximum flow in urban road networks can be observed for a range of number-vehicle (hence the horizontal line) as opposed to the traffic flow on highways that occurs for a given density. Traffic states along line B are characterized by partial congestion, i.e. most of the network links have saturation flow during the corresponding green stage, but does not take place severe overflow of vehicles to upstream links.

When the queues on the links of the network begin to overflow and block upstream links, we enter in the saturation region C. In this region, the increasing number of vehicles can lead to extensive overflow of queues, partially blocking of the network, wasting green light phase in the respective nodes and hence lower overall network flow. Properly designed real-time traffic control systems can alter the saturation region C in two ways:

- Increasing the number of vehicles in the urban network when saturation area C starts, i.e. extending the region of partial congestion B for larger number of vehicles in urban road network,

- Increasing the (negative) slope of the saturation region C. Both cases lead to an increase in flow for a larger number of vehicles in the urban network.

Finally, the region D is characterized by blocking part or the entire network with a very high number of vehicles and almost zero flow, a condition that if occurs can hardly be managed from any traffic control strategy.

The NFD concept for urban road network has been an issue of intensive investigations recently; indeed, the conditions under which it appears, the stability of its shape under different O–D patterns or at different peak periods or days-of-the-week, the impact of different signal control strategies, the possible hysteresis between the network filling and emptying phases, are still under the loop of ongoing analytical or empirical investigations and research. Nevertheless, based on what is known or observed in data, it is not too early for the NFD concept to be considered a basis for the derivation of traffic control strategies [2].

3 Traffic Signal Control with Gating

3.1 Introduction

In fact there are few strategies facing the oversaturation problem. Michalopoulos and Stephanopoulos [22] [23] present two in-depth theoretical studies of oversaturated signal systems with queue length constraints, firstly for a single intersection and then for two connected intersections.

The model is based on the earlier work [24] and describes mathematically the optimal control policy minimizing total intersection delay subject to the usual constraints plus the new upper bound on queue lengths. In the second study, where there are connected intersections, the coordination of intersections is also described as a constraint to the problem. It is assumed that the control action taken at the second intersection is an explicit function of the control action taken at the first intersection, with an appropriate time lag or lead.

Clearly if an adaptive algorithm is available, the optimal control policy could be determined and implemented in a real time basis. This is what gating is trying to achieve.

A practical tool, frequently employed against over-saturation of significant or sensitive links or urban network parts, is gating. The idea is to hold traffic back (via prolonged red phases at traffic signals) upstream of the links to be protected from oversaturation, whereby the level or duration of gating may depend on real-time measurements from the protected links. The method is usually employed in an ad hoc way (based on engineering judgment and manual fine-tuning) regarding the specific gating policy and quantitative details, which may readily lead to insufficient or unnecessarily strong gating actions [2]. To address this problem, a new traffic-responsive gating strategy, based on a simple but efficient PI feedback regulator, was developed for UTC (urban traffic control) under saturated conditions via exploitation of the NFD (network fundamental diagram) concept. More specifically, the NFD is used to derive clear gating targets that maximize throughput in the protected network part; moreover, an appropriate simple dynamic model is developed, that allows for the straightforward derivation of simple but efficient feedback regulators, suitable for smooth and efficient operations.

3.2 General gating task

The objective of the presented methodology is to mitigate urban traffic congestion via feedback gating, by exploiting the notion of the network fundamental diagram (NFD) for an urban network part that needs to be protected from the detrimental effects of over-saturation. To gate the traffic flow (usually during the peak periods) in an urban network, the area to be protected from possible congestion and the locations where gating queues will be created, must be defined. The general scheme of gating, including the protected network (PN), is sketched in Figure 3-1.

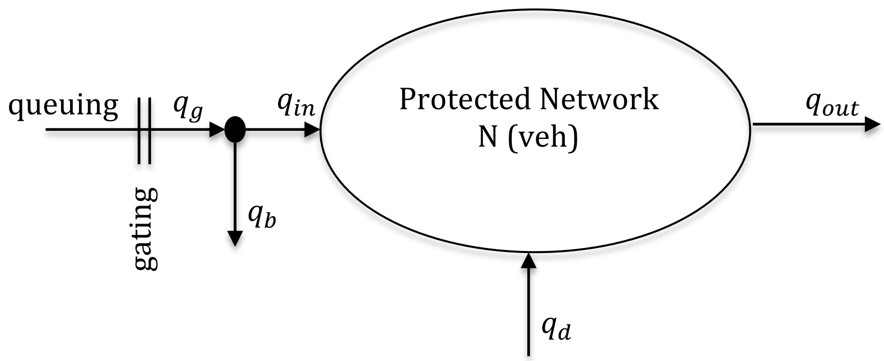


Figure 3-1: The protected network.

To implement gating, the usual traffic lights settings must be modified at (one or more) upstream junctions, which may be located more or less, close to the problematic area. In Figure 3-1, the double line indicates the gating location, upstream of which vehicle queues may grow temporarily faster than without gating; q_g is the gated flow, a part of which (q_b) may not be bound for the protected network (PN); while in q_{in} is the part of the gated flow that enters the protected network; q_d represents other (non-gated or internal) inflows to the PN (disturbances); finally q_{out} and N stand for the PN exit flow (both internal and external) and the number of vehicles included in the PN, respectively.

If N is allowed to grow beyond certain limits, the PN exit flow q_{out} decreases (according to the NFD) due to link queue spillovers and gridlock. To avoid this PN degradation, gating should reduce the PN inflow q_{in} appropriately, so as to maximize the PN throughput. This may incur some temporary vehicle delays in the queues of the gated junctions, which, however, may be eventually offset (at least for the q_{in} portion of the gated flow) thanks to the higher PN exit flow enabled by gating; on the other hand, the flow q_b will experience gating delays without any direct reward; these delays will be generally smaller if the gating junction is closer (or attached) to PN, due to accordingly smaller (or zero) flows q_b. In some situations, e.g. when major problems in PN may cause congestion to spread rapidly to adjoining areas, the use of gating could provide even higher benefits to the overall network [25].

3.3 Fundamental diagram of the PN

As explained earlier the fundamental diagram of the urban network in the method of gating is paramount. So, a network fundamental diagram may be an *ideal* NFD, if based on exact knowledge of the displayed quantities (this is practically only possible in analytic or simulation-based studies) for all links $z \in \mathbb{Z}$, where $\mathbb{Z}$ is the set of all network links; or an *operational* NFD, if based on available (more or less accurate) measurements and estimates at a subset $\mathbb{M}$ of all links, i.e. $\mathbb{M} \subseteq \mathbb{Z}$. An operational NFD is called complete, if the measurements cover all network links, i.e. if $\mathbb{M} = \mathbb{Z}$.

The NFD's y-axis reflects the Total Travelled Distance (TTD in veh·km per h), while the x-axis reflects the Total Time Spent (TTS in veh·h per h) by all vehicles in the PN. TTD and TTS are obtained from the emulated loop measurements via the following equations:

$$TTS(k) = \sum_{z\in\mathbb{M}} \frac{T * \widehat{N}_z(k)}{T} = \sum_{z\in\mathbb{M}} \widehat{N}_z(k) = \widehat{N}(k) \qquad (3\text{-}1)$$

$$TTD(k) = \sum_{z\in\mathbb{M}} \frac{T * q_z * L_z}{T} = \sum_{z\in\mathbb{M}} q_z(k) * L_z \qquad (3\text{-}2)$$

where z is the link where a measurement is collected; $\mathbb{M}$ is the set of measurement links; k = 0, 1, 2,… is a discrete time index reflecting corresponding cycles; T is the cycle time; q_z is the measured flow in the link z during cycle k; L_z is the length of link z; and $N_z(k)$ is the estimated number of vehicles in link z during cycle k, which is derived from measured occupancy measurements via the following equation

$$\widehat{N}_z(k) = L_z * \frac{\mu_z}{100 * \lambda} * o_z(k-1) \qquad (3\text{-}3)$$

where o_z is the measured occupancy (in %) in link z during cycle k; μ_z is the number of lanes of link z; and λ is the average vehicle length (in m). According to the derivations in (3-1) and (3-2), TTS equals the number of vehicles in all PN links equipped with detectors; while TTD is a length-weighted sum of the corresponding PN link flows [2].

3.4 System modeling for feedback control design

Gating may be enabled via very simple, but highly efficient and robust feedback regulators that are well-known in Control Engineering. The regulators are strictly based on real-time measurements, without any need for online model or demand predictions. On the other hand, for a proper choice of the feedback structure (among several offered in classical feedback theory), it is essential to know the basic dynamics of the process under control, and this task is indeed rendered quite simple and easy when using the notion of the NFD.

The developed model and feedback controller structures are summarized in Figure 3-2. The model input is the gated flow q_g (see Figure 3-1); the model output is the PN's TTS; while the main external disturbance is the uncontrolled PN inflow q_d . The model is first developed in a continuous-time environment for convenience. To start with, we have in the general case

$$q_{in}(t) = \beta * q_g(t-\tau) \qquad (3\text{-}4)$$

where β is the portion of gated flow (q_g) that enters the PN; t is the time argument; τ is the travel time needed for gated vehicles to approach the PN (when the gating link is not directly at the PN boundary).

The conservation equation for vehicles in the PN (see Figure 3-1) reads:

$$\dot{N}(t) = q_{in}(t) + q_d(t) - q_{out}(t) \quad (3\text{-}5)$$

As in the discrete-time case, we have also for the ideal values $TTS_{id}(t) = N(t)$ (where N is the real number of vehicles within PN), but TTS in Figure 3-2 denotes the operational value, which differs from the ideal value in two respects: firstly, detectors may not be available in each and every PN link, hence the operational TTS will be smaller by some factor $A \leq 1$; secondly, the occupancy measurement and, most importantly, the estimation (3-3) may not be exact, hence we introduce a measurement/estimation error ε_1; which finally yields

$$TTS(t) = AN(t) + \varepsilon_1(t) \quad (3\text{-}6)$$

From this operational $TTS(t)$, we may derive, using the operational NFD, the corresponding (operational) TTD, i.e.

$$TTD(t) = F[TTS(t)] + \varepsilon_2(t) \quad (3\text{-}7)$$

where $F(.)$ is a nonlinear best-fit function of the operational NFD's measurement points, and ε_2 denotes the corresponding fitting error (due to NFD scatter). Since TTD in (3-4) is the operational quantity, the ideal TTD_{id} (considering all PN links, not just the ones equipped with detectors) will be bigger, i.e.

$$TTD_{id}(t)B = TTD(t) \quad (3\text{-}8)$$

where $B \leq 1$ is the flow-analogous factor of A earlier.

To proceed, we will now introduce the modeling assumption that the PN outflow q_{out} is proportional toTTD_{id}, i.e.

$$q_{out}(t) = \frac{\Gamma}{L} TTD_{id}(t) \quad (3\text{-}9)$$

where Γ is a sort of network exit rate, $0 \leq \Gamma \leq 1$, and L is the average PN link length. Replacing (3-5) in (3-6), we complete the process model derivation according to Figure 3-2. The overall model (from q_g to TTS) turns out to be a time-delayed nonlinear first-order system.

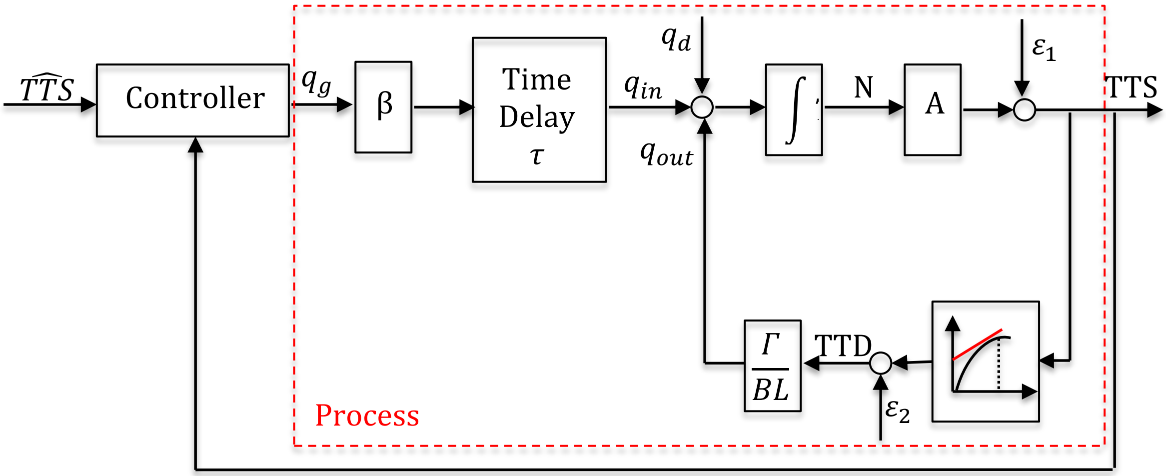


Figure 3-2: Feedback process for gating.

This model may be linearized around an optimal steady state that is within the aforementioned maximum TTD region of the NFD. Denoting steady-state variables with bars, we have

$$\bar{q}_{in} + \bar{q}_d = \bar{q}_{out} \tag{3-10}$$

$$\bar{q}_{out} = \frac{\Gamma}{BL}\overline{TTD} \tag{3-11}$$

while $\bar{\varepsilon}_1$ and $\bar{\varepsilon}_2$ are set equal to zero. With the notation $\Delta x = x - \bar{x}$ used analogously for all variables, the linearization yields

$$\frac{d}{dt}(\Delta TTS) = \left(\Delta q_{in} + \Delta q_d - \frac{\Gamma \bar{F}'}{BL}\Delta TTS\right) * A + \varepsilon \tag{3-12}$$

where ε may be derived from the previous errors $\bar{\varepsilon}_1$ and $\bar{\varepsilon}_2$, and $\bar{F}'$ is the slope of the NFD at the optimal set-point TTS , i.e. $\overline{TTS} = TTS$.

The continuous-time state equation (3-9) of the protected network (using the conservation equation and the NFD) may be directly translated in discrete time as follows:

$$\Delta TTS(k+1) = \mu * \Delta TTS(k) + \zeta * [\Delta q_{in}(k) + \Delta q_d(k)] + \varepsilon(k) \tag{3-13}$$

where $\mu = \exp(-\Gamma FTA/BL)$ and $\zeta = (1 - \mu)\, BL\, /\Gamma \bar{F}'$. It is trivial to include in these models the time delay, by replacing q_{in} from (3-4).

The derived simple model includes a number of parameters that have clear physical meaning; nevertheless, the precise value of some of these parameters may be difficult to obtain in practice, particularly if the PN is a sizeable network (as in the Chania example). However, the main reason for developing the gating model is to deduce the structure of the underlying dynamics, which is essential for the proper choice of the regulator structure [2].

3.5 Controller design

To avoid congestion-caused degradation (i.e. a TTD decrease), the critical value (i.e. the value of TTS at which the maximum TTD is attained) in the NFD is considered as the set value for the controller. The control goal is to keep the traffic state of the PN around the set value, so that TTD is maximized and the network does not enter the over-saturation area in the NFD. To this end, given the derived model structure in the previous section, the following proportional-integral-type (PI) feedback controller is well suitable

$$q_g(k) = q_g(k-1) - K_p[TTS(k) - TTS(k-1)] + K_I[\widehat{TTS} - TTS(k)] \tag{3-14}$$

where K_P and K_I are the proportional and integral gains, respectively. Good regulator gain values may be found with appropriate Control Engineering methods or manual fine-tuning; model parameter estimation (e.g. of μ and ζ in (3-10)), by use of real q_{in} versus TTS measurements, may be useful in this endeavor; in any case, feedback regulators are quite robust to moderate parameter value changes.

If gating is applied at multiple links, the flow calculated by the (unique) regulator (3-14) must be split among the gated links according to some pre-specified policy (e.g. according to the respective saturation flows). The flow calculated by the regulator (3-14) must be constrained by pre-specified minimum and maximum values to account for physical or operational constraints. For the lower bound, one may choose the flow corresponding to the minimum-green settings of the gated links or higher, e.g. if some gated links need to be protected from over-spilling. The upper bound has two components, a constant and a variable one, similarly to ALINEA ramp metering [26], and it is decided in real time which of the two is to be applied at each control step; the constant upper bound may be specified according to the maximum-green settings of the gated links, or lower, e.g. if some downstream links need to be protected from over-spilling; the variable upper bound aims at activating the regulator more promptly under certain circumstances. If the regulator flow distribution is found to violate some of these individual bounds, then the surplus flows are re-distributed among the rest of the gated links.

Gating could be activated only within specific time windows (e.g. at the peak periods) or if some real-time measurement-based conditions are satisfied. After distributing the regulator-ordered flow to the gated links, the individual sub-flows must be converted to appropriate green times by modifying the usual traffic signal settings in the corresponding junctions. Note that the implemented flow may be different than the flow ordered by the regulator for a number of reasons, including limited accuracy of signal specification, low demand, over-spilling downstream link or flow constraints; however, the regulator is largely robust to these occurrences thanks to its feedback structure [2].

4 Traffic Signal Control with Quadratic Programming

4.1 Introduction

This chapter investigates the efficiency of a new signal control methodology (QPC), which offers a computationally feasible technique for real-time network-wide control of the junction green times and is applicable also under congested traffic conditions. This methodology combines traffic flow modeling based on the store-and-forward modeling paradigm (SFM) (see 2.5) and mathematical optimization. More specifically, a generic mathematical model for the traffic flow process in large-scale urban networks is developed first, and a discrete-time optimal control problem is formulated for the design of traffic signal control strategies that aim at minimizing and balancing the link queues so as to minimize the risk of queue spillback. The derived optimization problem is of the quadratic-programming (QP) type, i.e. it involves a quadratic objective function with linear equality and inequality constraints.

The concept of rolling-horizon (model-predictive) control is eventually used to address the signal control problem in an on-line manner (closed-loop), whereby an optimal solution with respect to a fixed-length moving horizon with updated initial conditions (feedback) is calculated at each decision time, and the first-step control action (signal control plan) is applied to the signalized junctions of the traffic network [27].

The main control objective is to minimize the risk of oversaturation and spillback of link queues. To this end, one may attempt to minimize and balance the links' relative occupancies $x_z = x_{z,max}$ (see below). This criterion is physically reasonable as well as convenient from the numerical solution point of view. Alternatively, one may minimize the total time spent, but this may increase the risk of link queue spillback.

4.2 Problem formulation

The urban road network is represented as a directed graph with links $z \in \mathbb{Z}$ and junctions $j \in \mathbb{J}$. For each signalized junction j, we define the sets of incoming I_j and outgoing $\mathbb{O}_j$ links. It is assumed that the offset, the cycle time C_j, and the lost time L_j of junction j are fixed or calculated in real time by another algorithm. In addition, to enable network offset coordination, we assume that $C_j = C$ for all junctions $j \in \mathbb{J}$. Furthermore, the signal control plan of junction j is based on a fixed number of stages that belong to the set F_j, while u_z denotes the set of stages where link z has right of way (r.o.w.). Finally, the saturation flow S_z of link $z \in \mathbb{Z}$, and the turning movement rates $t_{w,z}$, where $w \in \mathbb{I}_j$ and $z \in \mathbb{O}_j$ are assumed to be known and constant but may be time-varying for the QPC approach.

By definition, the constraint

$$\sum_{i \in F_j} g_{j,i} + L_j = (or \ \le) C \tag{4-1}$$

holds at junction j, where $g_{j,i}$, is the green time of stage i at junction j. Inequality in (4-1) may be useful in cases of strong network congestion to allow for all-red stages. In addition, the constraint

$$g_{j,i} \ge g_{j,i,min}, \quad i \in F_j \tag{4-2}$$

where $g_{j,i,min}$ is the minimum permissible green time for stage i at junction $j \in \mathbb{J}$, is introduced to guarantee allocation of sufficient green time to pedestrian phases.

Consider a link z connecting two junctions M and N such that $z \in \mathbb{O}_M$ and $z \in \mathbb{I}_N$ (Figure 4-1). The dynamics of link z are given by the conservation equation

$$x_z(k+1) = x_z(k) + T[q_z(k) - s_z(k) + d_z(k) - u_z(k)] \tag{4-3}$$

where $x_z(k)$ is the number of vehicles within link z at time kT, $q_z(k)$ and $u_z(k)$ are the inflow and outflow, respectively, of link z in the sample period $[kT, (k+1)T]$; with T the discrete-time step and k = 0, 1, . . . the discrete-time index. In addition, d_z and s_z, are the demand and the exit flow within the link, respectively. For the exit flow we set $s_z(k) = t_{z,0} * q_z(k)$, where the exit rates $t_{z,0}$ are assumed to be known.

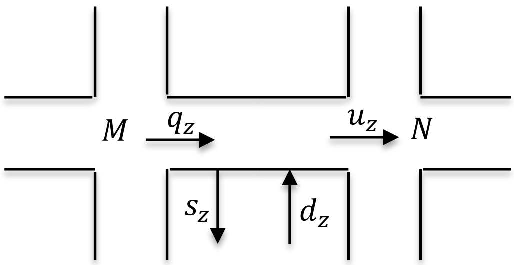


Figure 4-1: An urban road link.

Queues are subject to the constraints

$$0 \le x_z(k) \le x_{z,max} \quad \forall\, z \in \mathbb{Z} \tag{4-4}$$

where $x_{z,max}$ is the maximum admissible queue length. This constraint may automatically lead to a suitable upstream gating in order to protect downstream areas from oversaturation during periods of high demand.

The inflow to the link z is given by $q_z(k) = \sum_{w \in \mathbb{I}_M} t_{w,z}\, u_w(k)$, where $t_{w,z}$ with $w \in \mathbb{I}_M$ are the turning movement rates towards link z from the links that enter junction M.

For the outflow u_z we introduce the suggested modeling approach in 2.5 and from the equation (2-6) we derive

$$u_z = \frac{G_z(k)}{C} S_z \tag{4-5}$$

This equation stands for the discrete time step T, which is equal to C, where S_Z is equal to the saturation flow and G_z is the green time of link z, calculated as $G_z(k) = \sum_{i \in u_z} g_{j,i}(k)$.

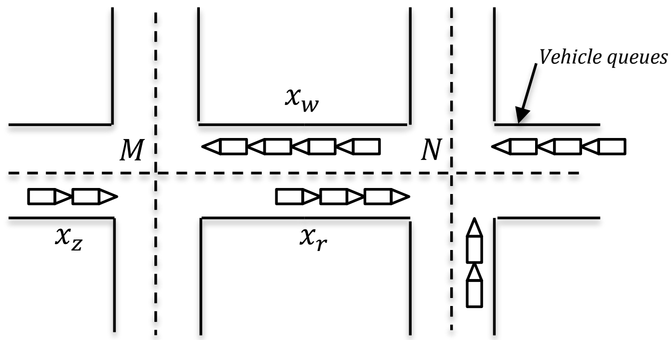


Figure 4-2: A two-way link connecting two junctions M and N

The green times G_z of each link z are introduced as additional independent variables. The introduced link green times G_z are constrained as follows:

$$0 \leq G_z(k) \leq \sum_{i \in u_z} g_{j,i}(k) \;\; \forall j \in \mathbb{J} \tag{4-6}$$

Note that, if the queue x_z is not sufficiently long or even equals zero, or if the downstream link queue is too long to accommodate a high inflow, the constraints in (4-4) will become active and will reduce the corresponding stage greens accordingly. As an illustrative example, assume that at a certain cycle there are two links z and w having r.o.w. simultaneously during a stage (M, i), and that $x_z \approx 0$ while $x_w \gg 0$ (see Figure 4-2). If G_z and G_w are not independently introduced, we have by definition $G_z = G_w = g_{M,i}$. Then, the stage green $g_{M,i}$ will be strictly limited by the constraint $x_z \geq 0$ although link w may need a longer green phase for dissolving x_w. In contrast, by introducing G_z and G_w independently, the algorithm can guarantee $x_z \geq 0$ by choosing G_z accordingly short without constraining G_w and the stage green. Similarly, if the link r downstream of link z is close to spillback (see Figure 4-2), the constraint $x_r \leq x_{r,max}$ can be guaranteed by choosing G_z accordingly short without constraining the green time of other links that have r.o.w. during the same stage.

Of course, this manipulation is necessary only within the control algorithm in order to preserve the validity of the modeling equation (4-5) and the overall model consistency. When applying the control results in real life, any G, that have been restricted to guarantee $x_z \geq 0$, may be switched to $G_z = g_{N,i}$.

In view of the above modification, replacing (4-5) in (4-3)

$$x_z(k+1) = x_z(k) + T[q_z(k) - s_z(k) + d_z(k) - \frac{G_z(k)}{C} S_z] \Rightarrow$$

$$x_z(k+1) = x_z(k) + \frac{T}{C}[(1 - t_{z,O}) \sum_{w \in I_M} t_{w,z}\, G_w(k) S_w - G_z(k) S_z] + T d_z(k)$$

and so it leads to a linear state-space model for road networks of arbitrary size, topology, and characteristics

$$x_z(k+1) = x_z(k) + \overline{\boldsymbol{B}}(k)\boldsymbol{G}(k) + T\boldsymbol{d}(k) \tag{4-7}$$

where $\boldsymbol{G}(k)$ is the link control vector consisting of the green times G_z of each link $z \in \mathbb{Z}$; $\overline{\boldsymbol{B}} \in \mathbb{R}^{\boldsymbol{n*n}}$ is a matrix of appropriate dimensions reflecting the network characteristics. Note that $\overline{\boldsymbol{B}}$ may be time-variant, if the involved saturation flows or turning movement rates are time-variant.

In this approach, the employed finite-horizon quadratic criterion that addresses the control objective has the form

$$\mathcal{J} = \frac{1}{2}\sum_{k=0}^{K}\sum_{z \in Z}\frac{x_z^2(k)}{x_{z,max}} \tag{4-8}$$

On the basis of the linear model (4-7); the constraints (4-1), (4-2), (4-4), (4-6) and the quadratic cost criterion (4-8) a (dynamic) optimal control problem may be formulated over a time-horizon K, starting with the known initial state x(0) in the state equation (11).

More precisely, the resulting QP problem reads: Minimization of the cost criterion (4-8) subject to (4-1), (4-2), (4-4), (4-6) and (4-7). This optimization problem has three types of time-dependent decision variables, namely the state variables $x_z(k)$, the stage green times $g_{j,i}(k)$, and the link green times $G_z(k)$. This QP problem (with very sparse matrices) may be readily solved by use of broadly available codes or commercial software within few CPU-seconds even for large-scale networks and long time-horizons [27].

By modifying the capacity x_{max} in the cost criterion of each link inside the PN (Protected Network) QPC model will consider congested the protected links for less number of vehicles x inside the link. So, it will provide such green times to the upstream junctions, regarding that the protected link has reached its capacity (even though it can hold more vehicles in reality) The traffic will be hold back upstream of the links to be protected from oversaturation succeeding the Gating concept.

4.3 The rolling horizon framework

For the application of the proposed QPC methodology in real time, the corresponding algorithm is embedded in a rolling-horizon (model-predictive) scheme. More precisely, the optimal control problem is solved on-line once per cycle for a large optimization horizon using the current state (current estimates of the number of vehicles in each link) of the traffic system as the initial state $x(0)$ and predicted demand flows over the horizon K; the optimization yields an optimal control sequence for K future cycles whereby only the first control (signal control plan) in this sequence is actually applied to the signalized junctions of the traffic network. The general algorithmic scheme of the rolling horizon framework is as follows:

> At time step k_0, the QP problem is solved, based on a measured initial condition $x(k_0)$ and on available demand predictions $\boldsymbol{d}(k)$, $k = k_0, \dots, k_0 + K - 1$, where K is the optimization horizon, to obtain the controls $\boldsymbol{g}^*(k)$ and states $\boldsymbol{x}^*(k+1)$, $k = k_0, \dots, k_0 + K - 1$. However, only a part of the control trajectory is actually applied to the process, namely $\boldsymbol{g}^*(k)$, $k = k_0, \dots, k_0 + k_R - 1$, where $k_R \ll K$ (e.g. $k_R = 1$). Then, at time step $k_0 + k_R$, based on the new measured initial condition $\boldsymbol{x}(k_0 + k_R)$ (feedback) and updated demand predictions $\boldsymbol{d}(k)$, $k = k_0 + k_R, \dots, k_0 + k_R + K - 1$, the QP problem is solved again to obtain the controls $\boldsymbol{g}^*(k)$ and states $\boldsymbol{x}^*(k+1), k = k_0 + k_R, \dots, k_0 + k_R + K - 1$, but only $\boldsymbol{g}^*(k)$, $k = k_0 + k_R, \dots, k_0 + 2k_R + K - 1$, is actually applied to the process, and so forth.

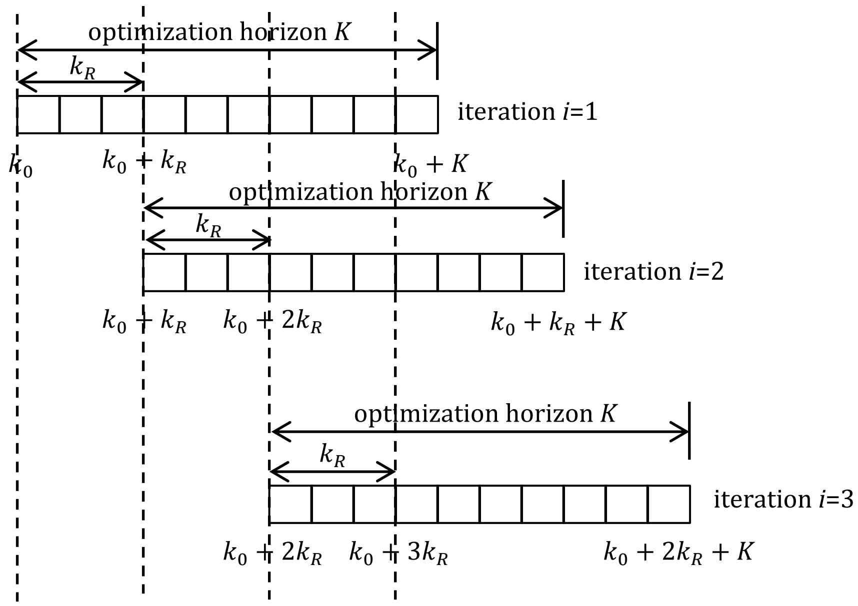


Figure 4-3: The rolling horizon figure.

There are several important issues that are associated with the rolling horizon framework just described:

- The saturation flows S_z and the turning movement rates $t_{w,z}$, may be time-variant, e.g. estimated or predicted in real time by well-known recursive estimation schemes; in addition, the predicted demand flows $d(k)$ may be calculated by use of historical information or suitable extrapolation methods (e.g. time series or neural networks).

- A satisfactory optimization horizon K should be in the order of the time needed for the network to be emptied. A much shorter optimization horizon may lead to "myopic" control actions.
- The computation time needed for the numerical solution of the QP problem must be short enough to permit the outlined repetitive on-line solution of the optimization problem. This is guaranteed for the present optimization method.
- The state variables $\boldsymbol{x}$ (the number of vehicles in each link) must be measurable or be estimated in real time. Occupancy measurements collected via traditional detector loops may be utilized to estimate the numbers of vehicles within links via suitable nonlinear functions [28]. The detector locations within links may be arbitrary, although the quality of estimation may be improved if the detectors are located around the middle of the link.

The outlined rolling-horizon procedure avoids "myopic" control actions while embedding a dynamic optimization problem in a traffic-responsive environment.

5 Microscopic Simulation

5.1 Introduction

To preliminarily investigate the comparative efficiency and real-time feasibility of the developed approaches to the problem of urban signal control, the urban network of the city center of Chania, Greece, is considered. For this network, we compare the optimized fixed time traffic signals with the closed-loop behavior of Gating and with the open-loop behavior of QPC methodology.

A greater part of the Chania urban road network is modeled in the AIMSUN microscopic simulation environment (TSS, 2008), according to Figure 5-1. The microscopic simulator AIMSUN is stochastic, thus different simulation runs (replications) with different random seeds may lead to different results. For this reason, it is common to use a number (10 in this work) of replications (4hour duration) for each investigated scenario and then calculate the average value of the these (10) runs for each evaluation criterion in order to compare different control cases. In these scenarios the number of vehicles is increased gradually (with different rates), reaching a peak congested period from 1,2 to 3h of simulation and then reduced, resulting in an empty PN, if no gridlock is apparent.

The PN consists of 165 links. In the middle of every link inside the red border line, a loop detector has been installed, and the related measurements are collected at every cycle (in this case 90sec). As indicated with small circled links in Figure 5-2, multiple origins and destinations are introduced at the network boundaries, but also at internal network locations, including the PN area. These origins and destinations (O–D) account for various corresponding in- and outflows, including on-street and off-street parking arrivals and departures, that may partially affect the PN area. The introduced O–D flows are realistic (based on real measurements) but not exact (particularly with regard to the used O–D rates). When running AIMSUN, the tool's embedded real-time dynamic traffic assignment option is activated, as this is deemed to lead to a more realistic distribution of the demand within the network [2]. This means that each vehicle re-decides which route to choose, depending on the current conditions of the network.

5.2 Evaluation criteria

Three performance indexes are utilized for the evaluation of each of the ten replications (as provided by AIMSUN):

- the average vehicle delay per km
- the mean speed, both for the entire Chania urban network (not only the PN)
- the total number of vehicles that exit the overall network during the whole scenario.

In addition to the above criteria, we also compare some quality criteria like TTS, TTD in the PN and the fundamental diagram provided from these data, as explained in 3.3.

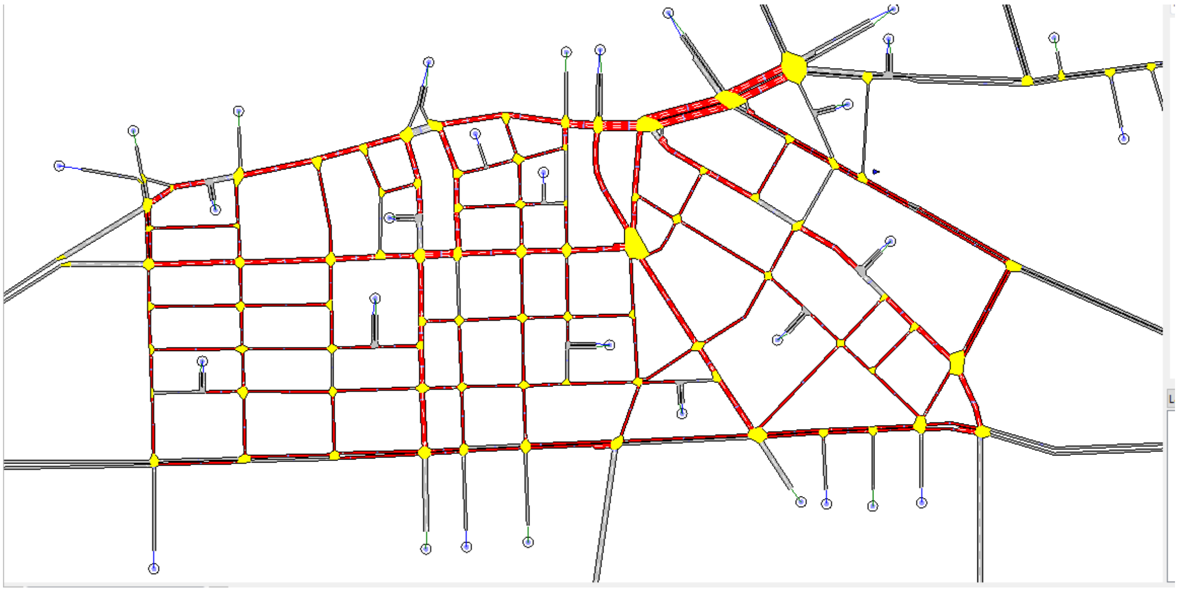


Figure 5-1: Chania urban road network modeled in AIMSUN.

5.3 Fixed-time control

In this case, AIMSUN uses the fixed-time settings, as in the real network of Chania. No other control action was used and the results for all 10 replications are showed in Table 1.

5.4 Gating strategy

For gating the PN is separated from the rest of the network by the red border in Figure 5-2. Eight gating links are specified exactly at the border of the protected network, indicated by arrows. The gating links have been chosen to provide sufficient space for vehicle queuing, so that further upstream junctions are not significantly obstructed. With dynamic traffic assignment option activated, if gating measures create long queues and delays at the gated links, alternative routes (if available) may be selected by the drivers towards their respective destinations; clearly, this reflects the medium-term routing behavior of drivers to any introduced gating measures. Note also that this diversion may jeopardize to some extent the intended gating impact if drivers divert and enter the PN via non-gated links; therefore, the choice of gating links should also consider the availability and potential attractiveness of alternative routes that bypass the gating location [2].

The parameterization and the methodology that was used for this type of control is not part of this study, but an extensive research was made in [25], [2] where the results are taken from, just for comparison reasons, as seen in Table 1. Note that a set point of TTS=600veh*h per h is selected for the gating operation.

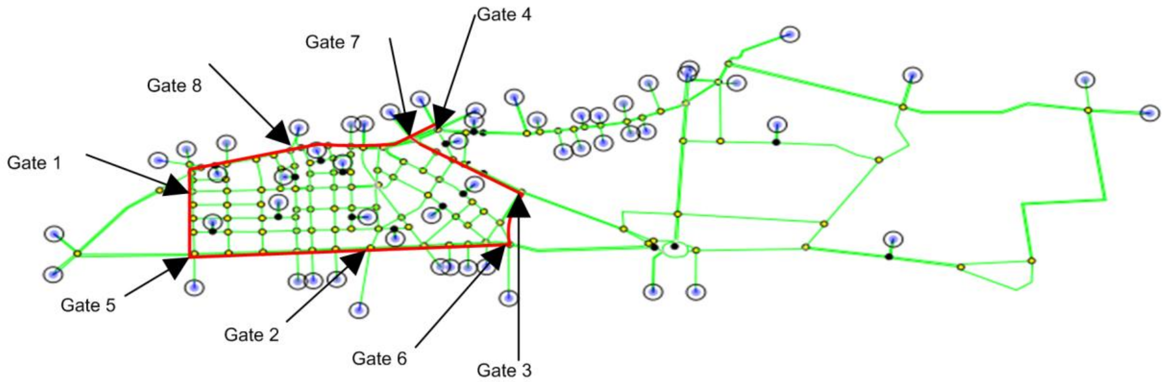


Figure 5-2: Chania urban road network modeled in AIMSUN for gating.

5.5 QPC strategy

For QPC, Figure 5-3 shows that the control model consists of 17 tagged nodes and 64 links. According to the methodology presented in chapter 4 the following sets for this network are defined:

- The set of junctions $\mathbb{J} = \{1,2, \ldots ,24\}$
- The set of controlled urban links $\mathbb{Z} = \{1,2, \ldots ,102\}$

Note that:

- The nodes 1a, 1b and 1c are controlled based on common signaling plans. The same applies to the nodes 2a and 2b, 3a and 3b, 4a and 4b, 16a and 16b. Therefore, from the control point of view these are considered as five nodes, and not as 11 separate nodes.
- Pairs of links L54 and L55, L37 and L38, L39 and L40, O13 and O23, L24 and L25, and the trio of links L15, L16 and L63 are different approaches to the same links at the same time because they do not take priority. For this reason, for the QPC strategy, are considered as different links.
- The detectors that the QPC model uses for collecting data are those used in the real urban network of Chania. The detectors located in the middle of each link in the simulated network in AIMSUN (as in gating) are used for collecting data for the evaluation criteria.
- The cycle time in the network is $C = 90sec$ and $T = C$ is taken as a control interval [29].

The implementation of QPC algorithm under control model developed in Chapter 4 requires the following data:

- For urban links: the capacity (x_{max}), the saturation flow (S_z), the turning rates ($t_{w,z}$), the duration of the fixed green and the stages at which the vehicles have r.o.w.
- For urban junctions: the signaling period (C), lost time (L), the number of stages (phases), the nominal green times (g_N) and the minimum green times (g_{min}).

Most of the above data are presented in the Table 4 Table 5 in Annex A.

Furthermore the optimization horizon for each scenario is 900sec (10 cycles).

To begin with, the first thing was to design the strategic network of QPC algorithm as seen in Figure 5-3. Note that this network is an extension of the network that was used in previous researches [29][30][31] and represents the real traffic network of the center City of Chania, Greece. By extending the strategic network, the input data of the algorithm also needed to be modified.

After that, in order to export some quality indexes (see below section 6.2) with the same way as gating (except from those export directly from AIMSUN), a subroutine was added in the QPC algorithm. This routine concerns the TTS and TTD output data that are exported by measurements from the detectors placed in the middle of each link in the protected network (PN).

In order to evaluate that all modifications were done in the right way, the algorithm was forced to run for all 10 replications, with fixed-time plan and the results were compared with those exported directly from AIMSUN; the results were identical.

Then, a survey was made by the running multiple times the replications with QPC enabled and different weight W (trial and error method). This parameter defines how much the values of green phases of traffic lights can change compared to the nominal ones. Small W values mean independent green phases, while big values indicate green times closer to nominal. The results from these runs are showed in Table 2. The best value derived from the average results is $W=10^{-7}$.

After finding the best weight minimizing the evaluation criteria as exported from AIMSUN the next step was to modify (reduce) the capacity of links inside the PN. The links are showed in Figure 5-3 with red arrows. The concept is to protect these links from oversaturation by “deceiving” the algorithm, as it tries to find a solution knowing that fewer vehicles can be served. This leads to hold vehicles upstream of the link with the modified capacity, as the strategy “sees” that the link has reached its capacity even with fewer vehicles. Here, 3 different values of capacities were tested for all 10 replications. These values are $(\frac{x_{max}}{2}, \frac{x_{max}}{3}, \frac{x_{max}}{4})$ and the results are presented in Table 3, annotated below.

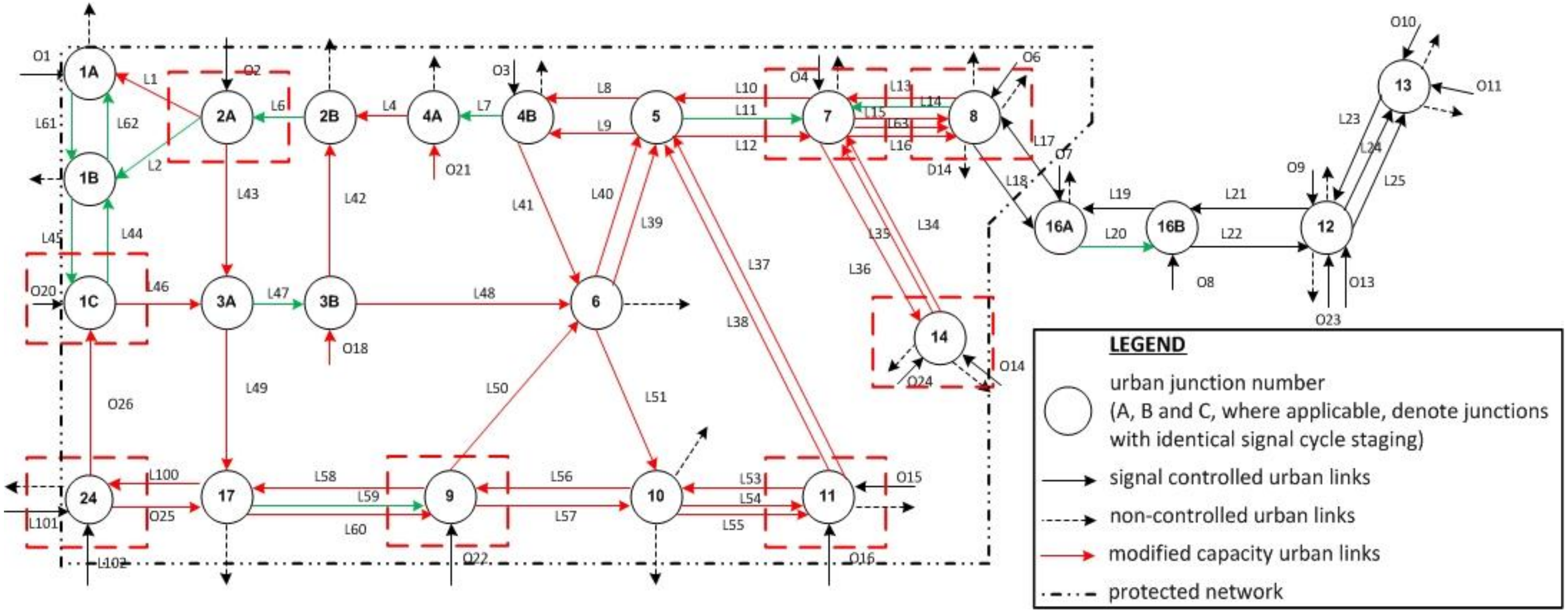


Figure 5-3: Schematic map of the model for the network of Chania.

5.6 Results of simulation

After implementing both strategies in the simulated network in AIMSUN, the final results are shown in the following tables and figures. Note that the 10 replications are identical for all tested plans. The first evaluation criteria that are examined are those derived directly from AIMSUN and concern the whole network (not only the PN) and give a general picture of how each strategy "reacted" in the different demand scenarios. The two real-time adaptive strategies are also compared using some quality criteria as derived from each algorithm and concern the values of TTS, TTD data in time for the PN and the fundamental diagram as resulting from these values. The conclusions of the comparison are detailed in the next chapter.

	FIX-TIME			**GATING**		
Replication	Delay (s/km)	Speed (km/h)	Vehicles out	Delay (s/km)	Speed (km/h)	Vehicles out
1	427.7	7.2	15643.0	244.4	11.3	15769.0
2	279.3	10.2	15697.0	234.4	11.7	15592.0
3	438.2	7.0	16033.0	239.9	11.5	15930.0
4	458.9	6.8	15693.0	257.2	10.9	15799.0
5	271.8	10.4	15696.0	223.5	12.1	15585.0
6	327.5	8.9	15712.0	239.9	11.5	15710.0
7	352.7	8.4	15801.0	241.8	11.4	15808.0
8	289.8	9.9	15829.0	249.7	11.1	15949.0
9	321.8	9.1	15921.0	242.7	11.4	16042.0
10	253.0	11.0	15859.0	236.3	11.6	15959.0
Average	**342.1**	**8.9**	**15788.4**	**241.0**	**11.5**	**15814.3**

Table 1: Fix-time and Gating Results.

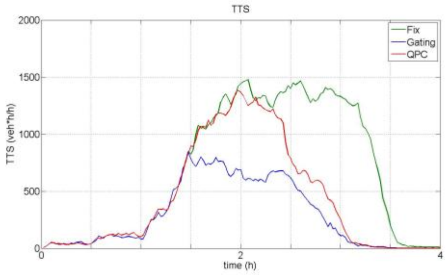
Weight	QPC (W=1e-14)			QPC (W=1e-8)			QPC (W=1e-7)			QPC (W=1e-6)		
Replication	Delay (s/km)	Speed (km/h)	Vehicles out	Delay (s/km)	Speed (km/h)	Vehicles out	Delay (s/km)	Speed (km/h)	Vehicles out	Delay (s/km)	Speed (km/h)	Vehicles out
1	315.2	9.3	15896.0	292.2	9.8	15878.0	312.4	9.3	15821.0	273.3	10.4	15736.0
2	289.3	9.9	15874.0	326.2	9.0	15825.0	308.4	9.4	15622.0	309.6	9.4	15720.0
3	333.6	8.8	15847.0	294.6	9.8	15637.0	282.4	10.1	15826.0	272.6	10.4	15632.0
4	324.1	9.0	15585.0	293.6	9.8	15822.0	276.8	10.3	15664.0	292.7	9.8	15864.0
5	340.6	8.7	15951.0	297.4	9.7	15812.0	297.0	9.7	15834.0	374.4	8.0	15796.0
6	304.1	9.5	15735.0	283.0	10.1	15596.0	266.2	10.6	15668.0	339.5	8.7	15662.0
7	240.7	11.4	15608.0	290.3	9.9	15670.0	314.3	9.3	15676.0	295.8	9.7	15715.0
8	408.2	7.5	15731.0	300.5	9.6	15701.0	286.5	10.0	15750.0	331.2	8.9	15599.0
9	391.9	7.7	15700.0	328.7	9.0	15841.0	305.3	9.5	15660.0	331.7	8.9	15886.0
10	316.3	9.2	15725.0	305.2	9.5	15744.0	307.0	9.5	15697.0	290.4	9.9	15681.0
Average	**326.4**	**9.1**	**15765.2**	**301.2**	**9.6**	**15752.6**	**295.6**	**9.8**	**15721.8**	**311.1**	**9.4**	**15729.1**

Table 2: QPC Replication Results for multiple weights W and $x=x_{max}$.

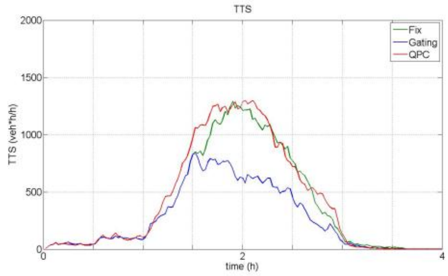
	QPC (W=1e-7)			QPC (W=1e-7)			QPC (W=1e-7)		
Capacity	**Xmax/2**			**Xmax/3**			**Xmax/4**		
Replication	Delay (s/km)	Speed (km/h)	Vehicles out	Delay (s/km)	Speed (km/h)	Vehicles out	Delay (s/km)	Speed (km/h)	Vehicles out
1	255.9	10.9	15722	282.8	10.1	15568	356.6	8.4	15678
2	305.5	9.5	15778	314.0	9.3	15977	327.0	9.0	15754
3	271.2	10.4	15577	326.1	9.0	15915	325.4	9.0	15732
4	316.2	9.2	15682	313.8	9.3	15612	400.7	7.6	15758
5	326.5	9.0	15791	412.7	7.4	16178	370.9	8.1	15811
6	317.0	9.2	15757	295.1	9.8	15754	326.3	9.0	15750
7	308.9	9.4	15630	291.3	9.9	15527	367.8	8.2	15907
8	312.8	9.3	15822	336.4	8.8	15769	306.6	9.5	15724
9	304.4	9.5	15900	292.0	9.9	15872	358.2	8.4	9358
10	360.9	8.3	15718	237.3	11.6	15782	351.8	8.5	15733
Average	**307.9**	**9.5**	**15738**	**310.1**	**9.5**	**15795**	**349.1**	**8.6**	**15121**

Table 3: QPC Results for modified capacity of links inside the PN.

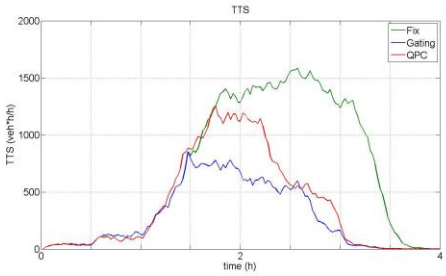
Replication 1



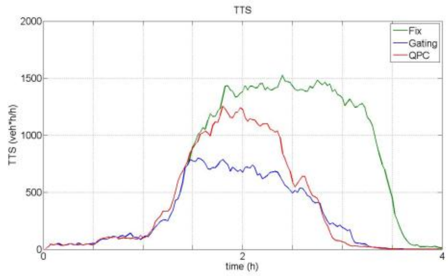
Replication 2



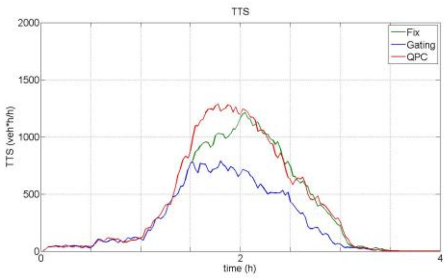
Replication 3



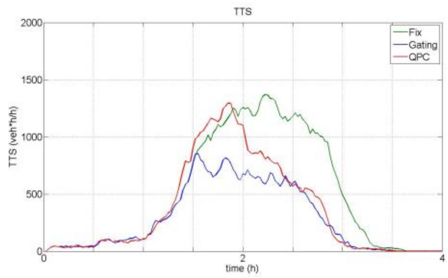
Replication 4



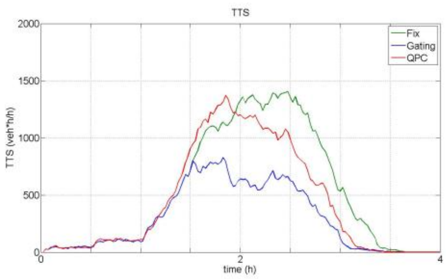
Replication 5



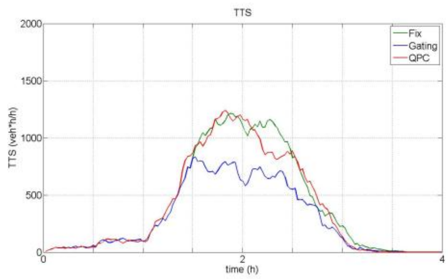
Replication 6



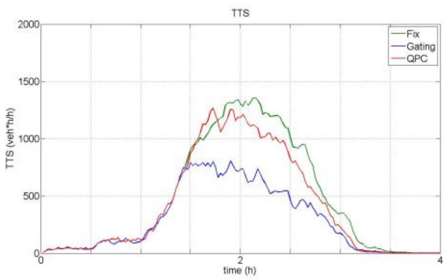
Replication 7



Replication 8



Replication 9



Replication 10

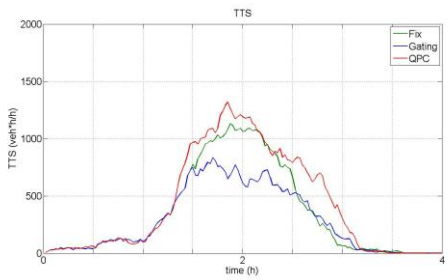
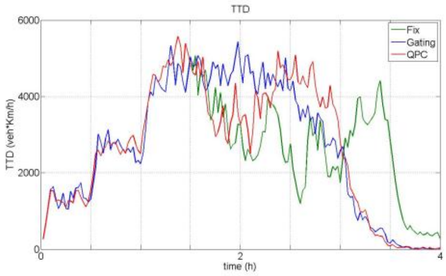
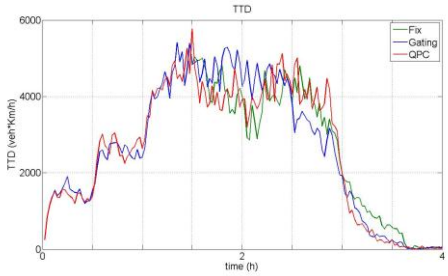


Figure 5-4: TTS of PN for all 10 replications

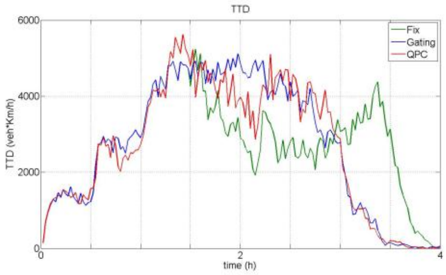
Replication 1



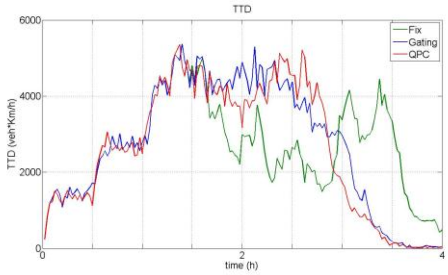
Replication 2



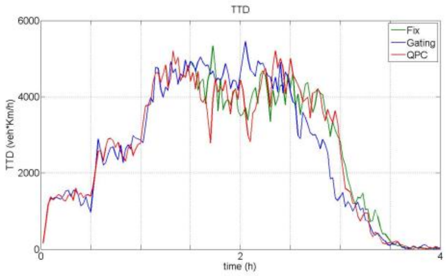
Replication 3



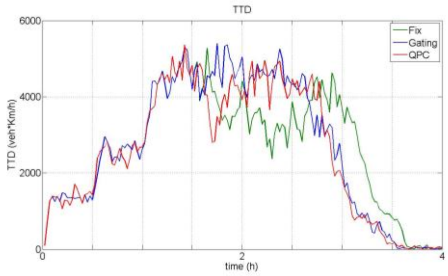
Replication 4



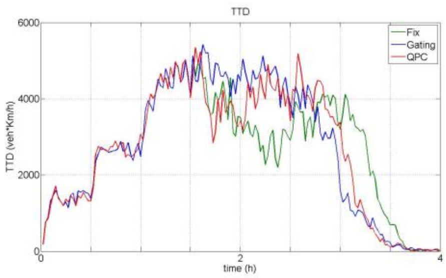
Replication 5



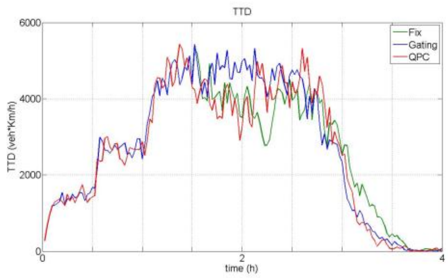
Replication 6



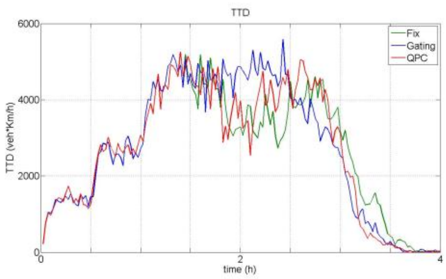
Replication 7



Replication 8



Replication 9



Replication 10

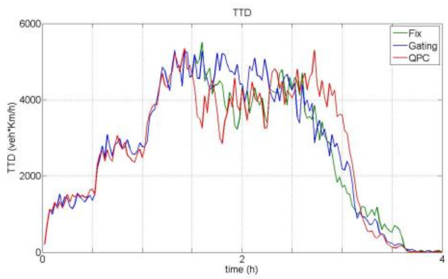
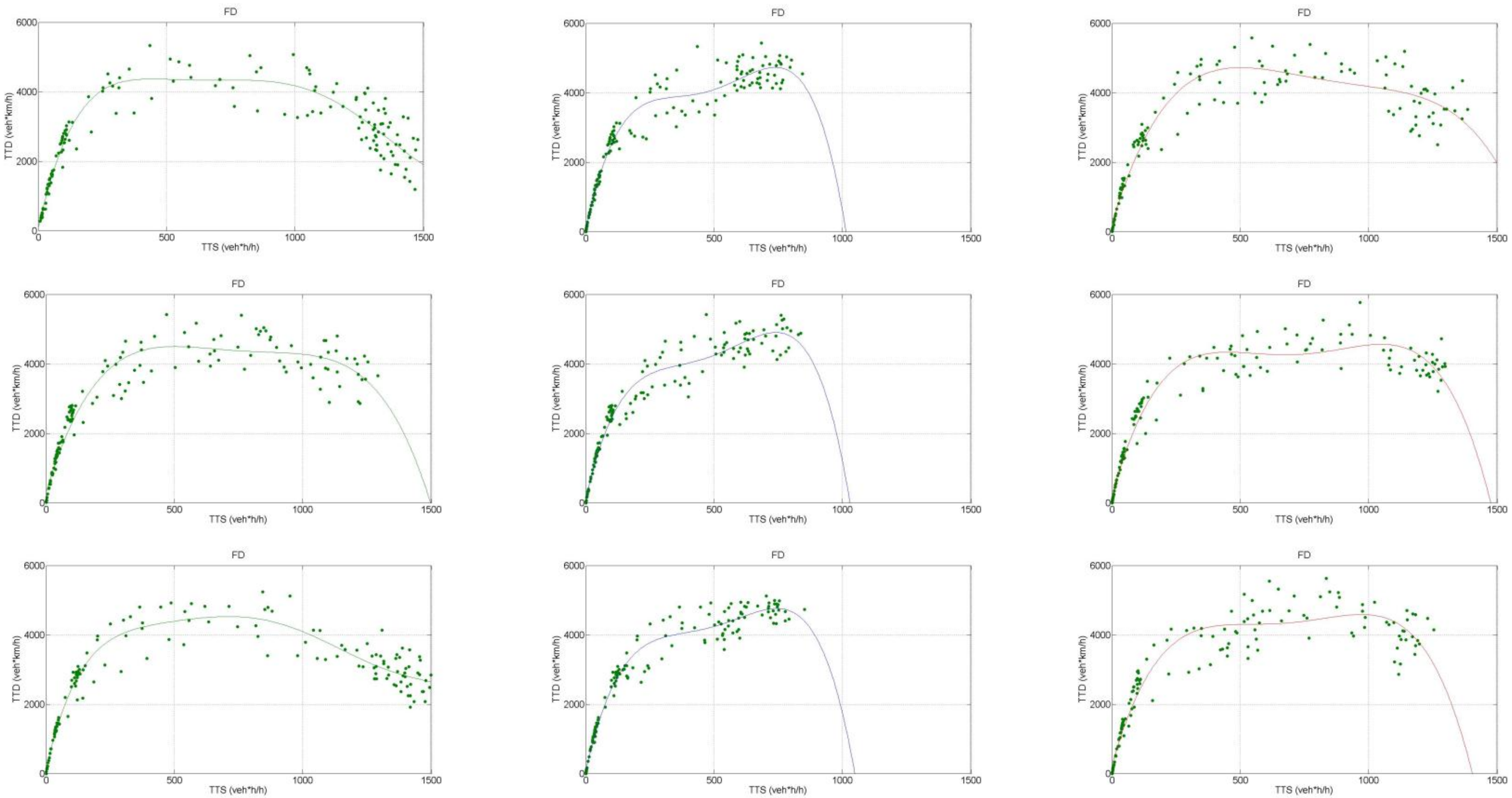


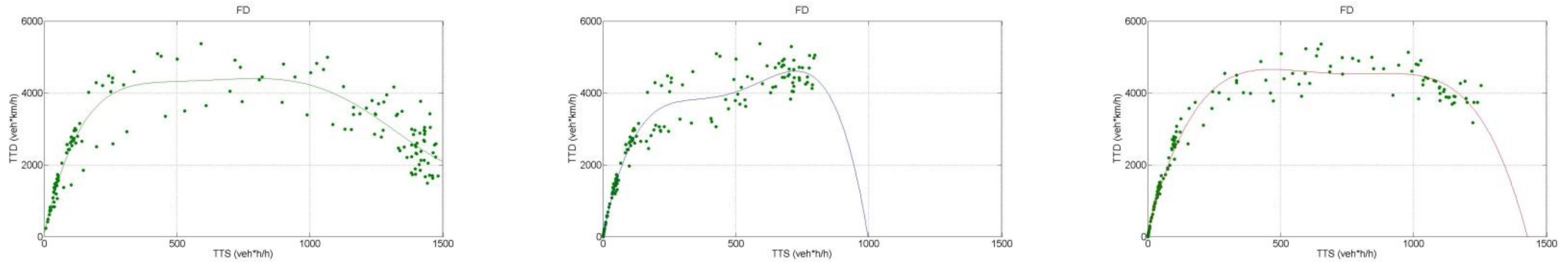
Figure 5-5: TTD of PN for all 10 replications

Fix | **Gating** | **QPC**

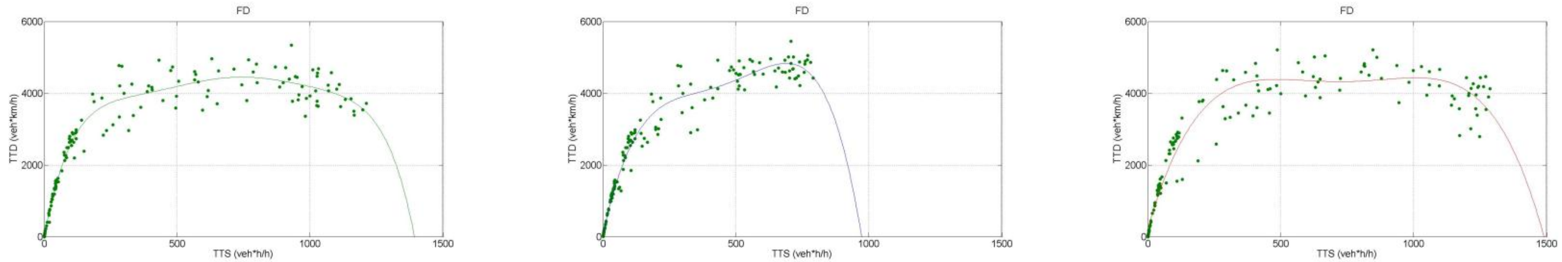
Replication 1 | **Replication 2** | **Replication 3**



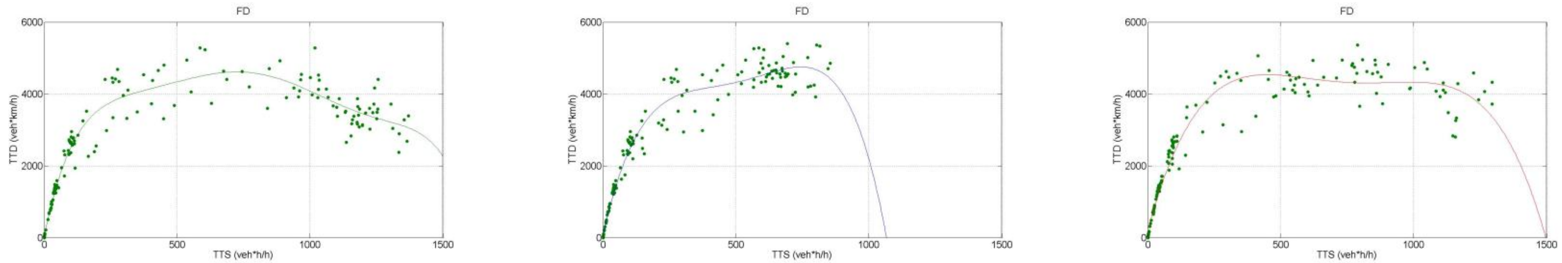
Replication 4



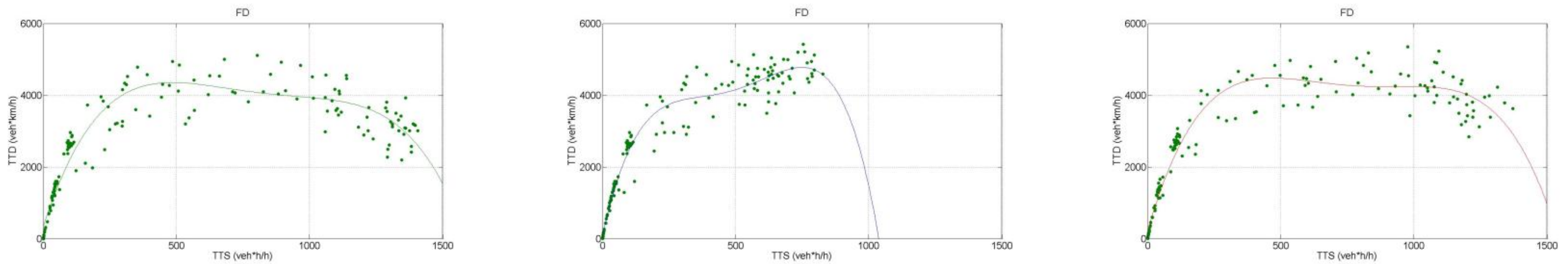
Replication 5



Replication 6



Replication 7



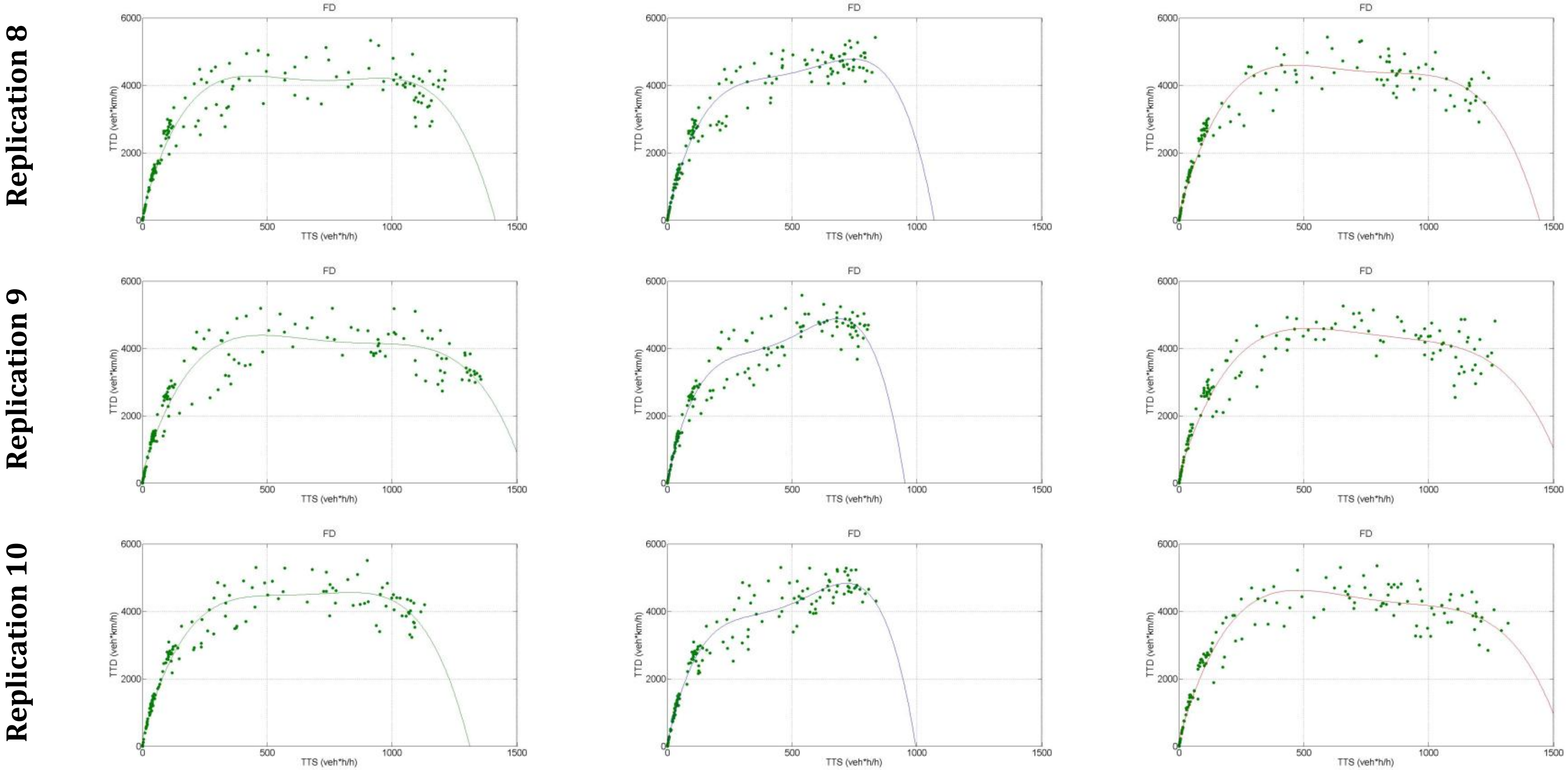


Figure 5-6: FD (TTD Vs TTS) of PN for all 10 replications

6 Conclusions

6.1 General

The results presented in Table 1 and Table 2, show by far that both real adaptive strategies excel the fixed-time one. The average values of the three performance indexes mentioned in section 5.2 are improved by 29.5% and 13.6% for gating and QPC respectively. Gating strategy has better indexes in all replications.

Note that QPC is tested for various values of the weight W, all resulting in better average, but in some replications, like the 5th one (see Table 3), the Fixed-Time plan has better performance. The best results for QPC are obtained with W=10^{-7} without modifying the PN links' capacities ($x = x_{max}$).

6.2 Comparing the real-time adaptive strategies

As mention above, results from gating strategy are better. Beyond this, some quality criteria must be examined in order to determine the behavior of each real-time adaptive strategy. These criteria must show clearly how the PN is protected from oversaturation when demand is increased, surpassing the capacity of the network. TTS and TTD data versus time, as well as the fundamental diagram of the urban protected network (PN), as presented in sections 2.6 and3.3, provide this kind of information.

For gating, it is obvious that the regulator manages to control the increased demand in all scenarios, protecting the network from congestion and finally "serve" more vehicles during the replication time. As shown in Figure 5-4, TTS is hold near the chosen set point (see section 5.4) and as a consequence, TTD is maintained at high levels, respectively (Figure 5-5).

For QPC we can conclude from Figure 5-4 that in most cases, a kind of protection is achieved compared to Fixed-Time control, but lacks significantly compared to gating strategy. Furthermore, as derived from Table 3, by reducing the capacity of links inside the PN, the QPC strategy deteriorates the results. This is unlike than what was expected because as mentioned in section 5.5 the idea was that the protection of the PN could be achieved by reducing the capacity of each link in the cost criterion in QPC algorithm (see (4-8)).

Also, by comparing 2 characteristic examples in Table 1 and Table 3 it can be seen that a solid conclusion cannot be derived regarding QPC behavior in various demand scenarios. In replication 1, QPC outmatches Fixed-Time strategy in all cases of modified capacities in PN ($x_{max}, \frac{x_{max}}{2}, \frac{x_{max}}{3}, \frac{x_{max}}{4}$) but in replication 5 the results are vice versa.

From Figure 5-6 it can be noticed that gating strategy manages to reach the best values of TTD for less number of vehicles (TTS) and keep it even during high demand period of time. QPC is managing better the high demand scenarios compared to Fixed-Time, as in most cases achieves better values of TTD for lower values of TTS (Area C in Figure 2-5).

It is clear that a further investigation needs to be done in order to determine why QPC does not have the expected results. Some reasons may be:

- QPC works considering the future demand is zero, i.e. has no information about the future and the results seem to be capturing myopic control decisions.
- QPC works assuming constant turning rates and saturation flows, which are used in the real implementation. Given O-D scenarios of demand and the activation of dynamic traffic assignment every 30sec, the above values shift significantly and are certainly different from those used in the real implementation (unlike the saturation flows should not alter significantly). Logically, the more frequently the traffic assignment algorithm is activated the more the turning rates vary.
- The strategic network of QPC differs from the network in AIMSUN because many links are added between the major junctions. These links, with dynamic traffic assignment activated, are preferred from the vehicles leading to saturated side-roads that QPC cannot observe. This behavior is not noticed in real life, as most drivers would prefer main-roads to side-roads even in congested situations.
- Measurements of QPC are taken from detectors located in real position of network and differs a lot from the data used in gating, deteriorating the performance of the algorithm

Obviously QPC is implemented without the most favorable conditions as gating,

6.3 Motives for future research

This thesis focused on the comparison of the QPC algorithm to the new real-time adaptive strategy of gating and the fixed-time control plan by the implementation in a modeled road network of Chania City in the AIMSUN microscopic simulation environment (TSS, 2008). It is clear that in general both strategies excel the fixed-time plan in congested scenarios, but despite the extend research much more need to be investigated regarding the QPC strategy. Issues for future research may include the following:

- Implementation of strategies in a less complex urban network, in order to examine how each plan manages the increased demand, with vehicles having fewer choices to alter their routes when enabling the dynamic traffic assignment option. Additionally, strategies will use data provided by the same detector as the network will be identical.

- Predict the demand by programming an additional model in QPC. This can be done by a normalization model estimating the number of vehicles entering a link each cycle by using the current measurements and the previous estimation.
- Calculate the changed turning rates by dynamic traffic assignment and provide the new calculations to QPC algorithm, so as to use more accurate data.

7 Annex A

Description of links in QPC Model

Name link	Name junct	Sat Flow	Leng th	Num Lanes	Capacity	Speed	Dist Detect Stop line	Num Stag	Stag
		(veh/h)	(m)		(veh)	(km/h)	(m)		
z	j	S_z	L_z	μ_z	x_{max}	v			
O1	1	1800	66	1	13	45	21	1	3
O2	2	1800	50	1	10	45	50	1	2
O3	4	1800	50	1	10	45	50	1	3
O4	7	3600	80	2	30	45	40	1	2
O6	8	3600	90	1	22	45	40	1	2
O7	16	1800	100	1	20	45	30	1	2
O8	16	1800	60	1	12	45	3	1	2
O9	12	1600	80	1	16	45	40	1	3
O10	13	1900	200	1	40	45	100	1	1
O11	13	2000	400	1	80	45	190	1	2
O13	12	1850	100	1	20	45	65	2	1,3
O14	14	1800	160	1	32	45	80	1	1
O15	11	1800	120	1	24	45	65	1	3
O16	11	3000	170	2	68	45	85	1	1
O18	3	1800	120	1	24	45	56	1	2
O20	1	3400	150	2	60	45	60	1	1
O21	4	1575	100	1	20	45	82	2	2,3
O22	9	3600	200	2	80	45	90	1	1
O23	12	1650	100	1	20	45	65	1	3
O24	14	1800	118	1	24	45	50	1	2
O25	17	1800	305	1	61	45	96	1	2
O26	1	1800	210	1	42	45	40	1	2
L1	1	4000	300	2	60	45	95	2	1,2
L4	2	3600	110	2	44	25	90	1	1
L8	4	3600	40	3	14	45	30	1	1

L9	4	1800	40	1	8	45	30	1	1
L10	5	3600	60	3	36	25	44	1	1
L12	7	4000	60	3	24	30	50	1	1
L13	7	3600	64	2	26	30	40	1	1
L15	8	1575	64	1	25	45	33	1	3
L16	8	1800	64	1	13	30	33	3	1,2
L17	8	2500	248	1	50	45	110	2	1,4
L18	16	2000	248	1	50	45	40	1	1
L21	16	1800	384	1	77	40	50	1	1
L22	12	2150	384	1	77	40	90	1	2
L23	12	2200	190	1	38	45	90	1	1
L24	13	1950	190	1	20	45	60	1	1
L25	13	2400	190	1	20	45	60	2	1,2
L34	7	1575	360	1	38	45	40	1	3
L35	7	2400	360	1	38	45	40	1	3
L36	14	1575	360	1	72	45	40	1	1
L37	5	1800	540	1	20	45	90	1	3
L38	5	1800	540	1	60	45	90	1	3
L39	5	1800	118	1	24	40	40	2	1,2
L40	5	1800	118	1	24	40	40	1	2
L41	6	3600	106	2	42	45	50	1	2
L42	2	1800	126	1	25	45	70	1	2
L43	3	2500	138	2	55	45	92	1	2
L46	3	3600	312	2	90	45	220	1	1
L48	6	3600	188	2	75	35	90	1	1
L49	17	3600	220	2	88	45	54	1	1
L50	6	1800	210	1	42	45	54	1	3
L51	10	3600	224	2	90	35	90	1	1
L53	10	1600	244	1	49	45	150	1	2
L54	11	1800	244	1	25	45	98	1	2
L55	11	1800	244	1	25	45	98	2	3,2

L56	9	1800	140	1	28	45	70	1	2
L57	10	1800	140	1	28	45	70	1	3
L58	17	1575	222	1	35	45	115	1	2
L60	9	1575	222	1	35	45	105	1	2
L63	8	1800	64	1	25	45	33	2	2,3
L100	24	1800	305	1	61	45	42	1	1
L101	24	1800	305	1	61	45	150	1	1
L102	24	1800	305	1	61	45	150	1	2

Table 4: Description of links in the QPC Model

junction	Cycle	Lost Time	Num Stages	G Nominal				Gmin			
j1	90	23	3	35	14	18	-	7	7	7	-
j2	90	32	2	46	12	-	-	7	7	-	-
j3	90	24	2	53	13	-	-	7	7	-	-
j4	90	19	3	57	7	-	-	7	7	7	-
j5	90	25	3	35	8	22	-	7	7	7	-
j6	90	33	3	37	10	10	-	7	7	7	-
j7	90	37	3	25	12	16	-	7	7	7	-
j8	90	32	4	30	9	7	12	7	7	7	7
j9	90	35	2	34	21	-	-	7	7	-	-
j10	90	19	3	44	13	14	-	7	7	7	-
j11	90	24	3	46	8	12	-	7	7	7	-
j12	90	33	3	24	15	18	-	7	7	9	-
j13	90	34	2	20	36	-	-	7	10	-	-
j14	90	30	2	48	12	-	-	7	7	-	-
j16	90	32	2	51	7	-	-	7	7	-	-
j17	90	16	2	51	23			7	7	-	-
j24	90	10	2	65	15	-	-	7	7	-	-

Table 5: Description of junctions in the QPC Model

8 Bibliography

[1] M. Papageorgiou, C. Diakaki, and V. Dinopoulou, “Review of Road Traffic Control Strategies,” *Proceedings of the IEEE*, vol. 91, pp. 2043–2067, 2008.

[2] M. Keyvan-Ekbatani, A. Kouvelas, I. Papamichail, and M. Papageorgiou, “Exploiting the fundamental diagram of urban networks for feedback-based gating,” *Transportation Research Part B: Methodological*, vol. 46, pp. 1393–1403, Dec. 2012.

[3] R. B. Allsop, “SIGSET: A computer program for calculating traffic capacity of signal-controlled road junctions,” vol. 12, Traffic Eng. Control, pp. 58–60, 1971.

[4] R. E. ALLSOP, “SIGCAP: A computer program for assessing the traffic capacity of signal-controlled road junctions,” Traffic Engineering & Control, pp. 338–341, 1976.

[5] R. A. Vincent and C. P. Young, “Self-optimizing traffic signal control using microprocessor: The TRRL ‘MOVA’ strategy for isolated intersections,” vol. 27, Traffic Eng. Control, pp. 385–387, 1986.

[6] N. H. G. D. C. Little, M. D. Kelson, “MAXBAND: A Program for Setting Signals on Arteries and Triangular Networks,” *U.S. Dept.Transp.,Washington, DC,Transp. Res. Record 795*, 1981.

[7] C. S. and N. H. Gartner, “MULTIBAND-96: A program for variable bandwidth progression optimization of multiarterial traffic networks,” *U.S. Dept. Transp.,Washington, DC,Transp. Res. Record 1554*, 1996.

[8] D. I. Robertson, “TRANSYT method for area traffic control,” vol. 10, Traffic Eng. Control, pp. 276–281, 1969.

[9] M. C. Hunt, P. B., Robertson, D. L., Bretherton, R. D., & Royle, “The SCOOT on-line traffic signal optimisation technique,” vol. 23, Traffic Engineering and Control, pp. 190–199, 1982.

[10] P. R. Lowrie, “SCATS: The Sydney co-ordinated adaptive traffic system-principles, methodology, algorithms,” Proceedings of the IEE international conference on road traffic signalling, pp. 67–70, 1982.

[11] T. Wongpiromsarn, T. Uthaicharoenpong, Y. Wang, E. Frazzoli, and D. Wang, “Distributed Traffic Signal Control for Maximum Network Throughput.”

[12] K. Fehon and J. Peters, “Adaptive Traffic Signals , Comparison and Case Studies,” 2010.

[13] P. B. Friedrich, “Adaptive Traffic Control In Metropolitan Areas,” *Institute of Transport Engineering and Planning*.

[14] B. L. Pierre-Yves Texier, "The State of the Art in French Urban Traffic Control," pp. 1–8.

[15] L. C. Liao, "A Review of the Optimized Policies for Adaptive Control Strategy (OPAC)."

[16] F. Boillot, S. Midenet, and J.-C. Pierrelée, "The real-time urban traffic control system CRONOS: Algorithm and experiments," *Transportation Research Part C: Emerging Technologies*, vol. 14, pp. 18–38, Feb. 2006.

[17] S. Shepherd, "A Review of Traffic Signal Control," *Institute of Transport Studies, University of Leeds. Working Paper 349*, 1992.

[18] N. Geroliminis and B. Boyacı, "The effect of variability of urban systems characteristics in the network capacity," *Transportation Research Part B: Methodological*, vol. 46, pp. 1607–1623, Dec. 2012.

[19] N. Geroliminis and C. F. Daganzo, "Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings," *Transportation Research Part B: Methodological*, vol. 42, pp. 759–770, Nov. 2008.

[20] J. W. Godfrey, "The mechanism of a road network," vol. 11, Traffic Engineering and Control, pp. 323–327, 1969.

[21] V. Dinopoulou, C. Diakaki, and M. Papageorgiou, "Application and Evaluation of the Signal Traffic Control Strategy TUC in Chania," *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, vol. 9, pp. 133–143, Jul. 2005.

[22] G. Michalopoulos, Pg and Stephanopoulos, "Over-saturated signal systems with queue length constraints - I. Single intersection," *Transportation Research*, vol. 11, pp. 413–421.

[23] G. Michalopoulos, Pg and Stephanopoulos, "Over-saturated signal systems with queue length constraints - II. Systems of intersections," *Transportation Research*, vol. 11, pp. 423–428.

[24] D. and G. A. GAZIS, "Optimal Control of Oversaturated Store-and-Forward Transportation Networks," *IBM*, 1975.

[25] M. K. Ekbatani, A. Kouvelas, I. Papamichail, and M. Papageorgiou, "Congestion Control in Urban Networks via Feedback Gating," *Procedia - Social and Behavioral Sciences*, vol. 48, pp. 1599–1610, Jan. 2012.

[26] W. Wang, Y., Papageorgiou, M., Gaffney, J., Papamichail, I., Rose, G., and Young, "Local ramp metering in random-location bottlenecks downstream of a metered on-ramp," *Transportation Research Record*, pp. 90–100, 2010.

[27] K. Aboudolas, M. Papageorgiou, A. Kouvelas, and E. Kosmatopoulos, "A Rolling-Horizon Quadratic-Programming Approach To The Signal Control Problem In Large-Scale Congested Urban Road Networks," pp. 1–15, 2008.

[28] C. Diakaki, "Integrated control of traffic flow in corridor road networks," *Phd Thesis*, 1999.

[29] K. Aboudolas, M. Papageorgiou, and E. Kosmatopoulos, "Store-and-forward based methods for the signal control problem in large-scale congested urban road networks," *Tramsportation Research Part C: Emerging Technologies*, vol. 17, pp. 163–174, Apr. 2009.

[30] K. Aboudolas, M. Papageorgiou, and E. Kosmatopoulos, "Control and Optimization Methods for Traffic Signal Control in Large-scale Congested Urban Road Networks," *2007 American Control Conference*, vol. 10, pp. 3132–3138, 2007.

[31] V. Dinopoulou, C. Diakaki, and M. Papageorgiou, "Applications of the urban traffic control strategy TUC," *European Journal of Operational Research*, vol. 175, pp. 1652–1665, Dec. 2006.