Simulation-based Optimization of Actuated Traffic Signal Plans

Tamir Balasha
Ph.D candidate
Faculty of Civil and Environmental Engineering
Technion –Israel Institute of Technology
Haifa 32000, Israel
Tel: +972-4-829-3608; Fax: +972-4-829-5708; Email: tbalasha@technion.ac.il

Tomer Toledo
Faculty of Civil and Environmental Engineering
Technion - Israel Institute of Technology
Haifa 32000, Israel
Tel: +972-4-829-3080; Fax: +972-4-829-5708; Email: toledo@technion.ac.il

August 2014

ABSTRACT

The design of traffic signal control has a profound impact on performance of urban traffic systems. Current traffic signal plans involve complex control logic and a large number of parameters that need to be set. However, little attention has been given to optimization and evaluation of these plans. Simulation-based signal optimization has been limited mainly due to the heavy computational burden associated with it.

This paper reports on the overall structure and the various components of a simulation-based optimization system to optimize the parameters of complex actuated traffic signal plans. It focuses on a development of a mesoscopic traffic simulation - MESCOP (Mesoscopic Evaluation of Signal COntrol Plans). MESCOP is detailed enough to represent the characteristics of actuated traffic signal plans, including representation of the intersection layout and the detectors. The stochastic processes of arrival to the intersection and movement within it are also modeled in detail. The model represents passenger cars, transit vehicles and pedestrians. The system framework incorporates connection between MESCOP, traffic signal design and Genetic Algorithm as the optimization method.

The Integrated system to optimize traffic signal plans is demonstrated with an application to a signalized intersection in Haifa, Israel. This intersection is controlled by an actuated traffic signal with transit priority and compensation and queue override mechanisms. The results indicate a large potential to improve the intersection performance, with a reduction of 28% in traffic delays compared to the parameter values set in the original design. Computationally, the results show that MESCOP is very efficient compared to microscopic traffic simulation models, which are often used for similar evaluations. It highlights the benefit of a mesoscopic models especially for large scaling networks and for optimization processes which require high number of simulation replications

INTRODUCTION

Transportation systems face continuous increases in congestion. Congestion limits mobility and results in negative economic impacts. Traffic signal control is the main tool for operators and managers of transportation systems to allocate capacities and affect the state of the system and its performance. The efficient design of intersection traffic signal control has been recognized as one of the most cost useful methods to improve accessibility and mobility in urban networks [1]. However, inadequate design of traffic signal timing plans may prohibit realizing their potential to alleviate congestion [2].

Traffic signal timing has advanced dramatically since Webster [3] developed the basic principles and theory of traffic signal optimization. Over the years, signal plans have evolved from pre-timed plans to actuated plans that utilize detection technologies and are sensitive to variations in traffic demand. The complexity of traffic signal plans has increased further with the introduction of additional features, such as transit priority or pedestrian and bicycle phases and actuation. Thus, signal timing plans are increasingly complex with more sophisticated logical conditions and constraints and contain a large number of parameters that need to be carefully set and fine-tuned. As a result, solutions for setting optimal parameter values are becoming analytically intractable, which further contributes to the difficulty of design and evaluation of traffic signal plans.

A variety of tools and methods have been developed in order to optimize traffic signal plans. Examples of these optimization tools include HCM [4], SYNCHRO [5], TRANSYT-7F [6], PASSER II [7], PASSER V [8] and MAXBAND [9]. They are suitable for pre-timed plans, and in some cases also for traffic actuated signal plans. They embed analytic or macroscopic traffic models to predict the value of a measure of the performance of an intersection, under a given demand scenario and a set of signal timing parameter values. The parameter optimization commonly considers four basic groups of parameters: cycle length, green splits, phase sequence and offsets. However, actuated traffic signal plans may include many other parameters related to signal timing (e.g., minimum and maximum green for each phase), detectors (e.g. minimum gaps), pedestrians (e.g. maximum waiting time) and transit priority. As a consequence, analytical solutions to optimize parameter values become intractable. Thus, there is a need for reliable simulation-based tools to optimize or fine-tune complex traffic signal plans.

Despite the progress in development of sophisticated signal plan designs, little attention has been given to the optimization of these plans. Signal plan optimization methods require a level of detail in the movement of vehicles that is supported by microscopic traffic simulation models. These models also provide the ability to account for system variability that stems from heterogeneity in driving behavior, existence of different vehicle classes with different capabilities and characteristics, and fluctuations in demand. However, the use of simulation-based signal plan optimization has been very limited, mainly due to the heavy computational burden [10]. Foy et al. [11], Hadi and Wallas [12], Rouphail et al. [13], Park et al. [14], Park and Schneeberger [10] and Stevanovic et al. [15] used microscopic traffic simulation models within stochastic optimization algorithms for the four basic parameters of traffic signal plans (cycle length, green splits, phase sequence and offsets). A few other studies ([1], [16], [17], [18], [19]) expanded these works to include additional traffic signal control setting, such as minimum green, maximum green and detector placements. These studies demonstrated substantial potential for improvements to the intersection operations. However, they were also computationally demanding. To curb the computational efforts, researchers limited the number of parameters that were optimized, used sequential rather than joint optimization of the parameters, or reduced the number of simulation replications that were used to evaluate the objective function. These may all yield sub-optimal solutions. Hence there is a need for reliable and efficient optimization tools for complex traffic signal plans.

This paper presents a simulation-based optimization system for actuated traffic signals. It incorporates a mesoscopic traffic simulation model that supports evaluation and optimization of complex actuated traffic signal plans, which is called MESCOP (Mesocopic Evaluation of Signal COntrol Plans). This model is computationally efficient compared to the microscopic models that have been used for this purpose in the past. At the same time, it maintains the level of detail required in order to model the characteristics of actuated traffic signal plans, including features of transit priority and pedestrian actuation. Optimization of complex actuated traffic signal parameters can be achieved simply and in a reasonable run time through the using of this model.

The rest of this paper is organized as follows: the next two sections describe the overall structure of the integrated simulation-based optimization system and the details of the various components within it, especially the developed mesoscopic simulation. Next, evaluation of traffic signal performance, optimization of signal plan and the computational characteristics of the system are demonstrated with an application to a signalized intersection in Haifa, Israel. The potential to generate superior traffic signal plans with significant run time saving are discussed. Finally, a summary and discussion of the results it presented.

OVERALL STRUCTURE

The simulation–based optimization framework consists three main components that interact with each other, as shown in FIGURE 1: traffic simulation model, signal control and optimization algorithm.

The traffic simulation model explicitly represents the movement of individual road users including passenger cars, transit vehicles and pedestrians that pass through the intersection. It contains the characteristics of an intersection such as geometric layout, vehicle and pedestrian detectors, pedestrians crosswalks and more. The simulation model is connected to the signal control, and movements through the intersection occur according to signal indications.

The signal control component implements the control logic for the intersection and the parameters associated with this logic. It is run every second to determine the light indications in the next second. In determining the light indications, the logic may use information on the current and previous indications (e.g. how long a certain light has been green) that it stores, and information on the states of the detectors in the system (e.g. detectors activated) that are received from the vehicle movement simulator.

From the simulation results, values of performance measures are estimated. The signal plan optimization defines a performance measure, as calculated from the simulation model, as the objectives of the signal control (e.g. minimize delays or queue length, maximize throughput). The plan optimization may be constrained by various control requirements (e.g. maximum allowed queue lengths or waiting times, minimum green time to certain movements). The satisfaction of these constraints is also evaluated with values estimated from the simulation model. The decision variables in this optimization problem may include the parameters of the control plan (e.g. green times, extension parameters of the transit priority rules) and system layout parameters (e.g. locations of detectors). The type and number of signal parameters that are optimized may vary depending on the intersection and plan. The optimization algorithm, regardless of its specific details, would require making multiple runs of the simulation with different values of the control plan parameters and calculating the resulting values of the objective function and control constraints. Thus, the optimization module is able to set parameter values for the simulation model, to run the model, and to extract the values of performance from the simulation results. This process continues iteratively until an optimal solution is found.

In the next sections, the components of the simulation-based optimization system are described in further detail.

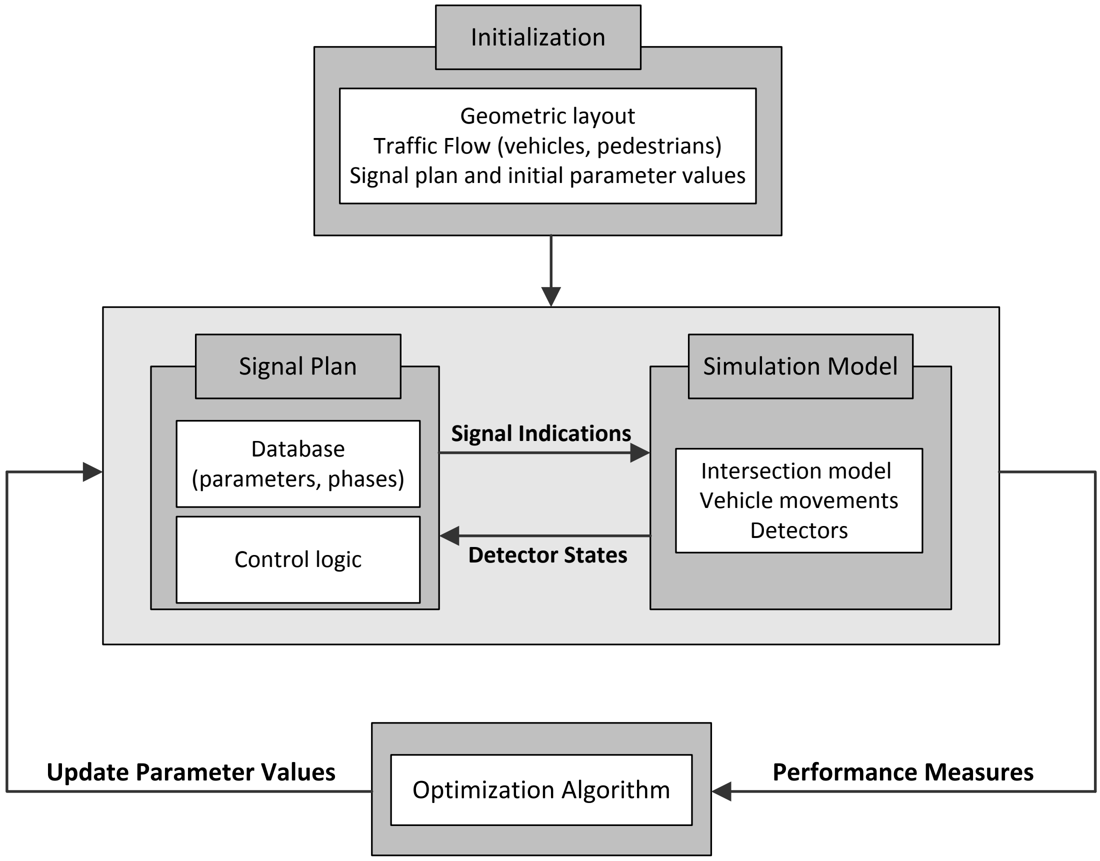

FIGURE 1: The integrated simulation - based optimization overall structure

SYSTEM COMPONENTS

Mesoscopic Simulation Model

The developed mesoscopic simulation component, which is called MESCOP, explicitly represents the individual road users including passenger cars, transit vehicles and pedestrians that pass through the intersection.

Passenger cars and transit vehicles movements are modeled by events occurring at detector locations and the stop line. The movement model comprises of three stages: the initial approach to the intersection, the movement to the stop line and crossing the intersection. A vehicle enters the system at the time it arrives at the furthest detector location (or at the stop line if there are no detectors). The arrival of vehicles at this point is modeled as a stochastic process which is based on an input distribution of inter-arrival times and the assumed mean traffic arrival rate. The travel time from the initial detection point to the stop line (or additional detectors upstream of the stop line, if any) is also modeled as a stochastic process based on the assumed approach speed. Vehicles that arrive at the stop line enter a lane-specific vertical queue. The lane they are allocated to is determined based on their intended turn movement. They are released from this queue when the light indication, received from the traffic control component, for their movement is green, according to the first-in-first-out (FIFO) rule and at a rate based on the approach saturation flow. Pedestrians arrive at the crossing line randomly with an arrival rate provided as input. They are assumed to activate pedestrian crossing buttons, if they exist, at the time of arrival. The simulation implementation is time-based with a step size of 1 second in order to fit with the resolution of the control logic.

Vehicles

Vehicles are represented explicitly in the model as individual entities. FIGURE summarizes the movement of vehicles in the approach to the intersection and until their release from it.

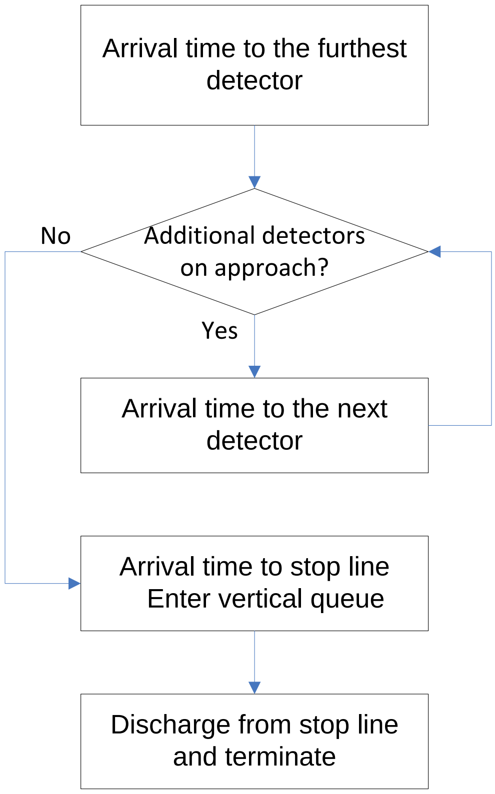

FIGURE 2: Vehicle movement model

Vehicles are generated in the model when they first arrive at the location of the furthest detector in their approach to the intersection. On their arrival, the vehicles activate the detector. The arrivals are modeled as a stochastic process. In the default implementation, the inter-arrival times are assumed to follow a negative exponential distribution. In the case that an approach does not have any detectors, the arrival occurs at the stop line. To prevent a situation that two or more vehicles arrive in the same lane at the same time interval, the headway between vehicles is set to a minimum value of one second and the arrival time of the second vehicle is adjusted accordingly. On arrival, vehicles are allocated to a specific lane according to their specific turning movement at the intersection. A lane changing model, which is commonly implemented in microscopic traffic simulation models, is not implemented in this model. If more than one lane is appropriate for a certain movement, the proportions of vehicles allocated to each of the lanes is calculated according to the method of critical lane flows, which aims to equalize flows on all lanes in a specific approach.

The next event that the vehicle will experience is the arrival to the next detector downstream of the current one (or the stop line, if no additional detectors exist). The arrival time at this detector is given by:

$$t_n(i) = \max\left(t_n(i-1) + \frac{d(i,i-1)}{v_n} + \varepsilon_{ni}, t_{n-1}(i) + h^{\min} \right) \tag{1}$$

Where, $t_n(i)$ is the arrival time of vehicle n to detector (or the stop line) i. $d(i,i-1)$ is the distance between detectors i and i-1. v_n is the approach speed of vehicle n, which depends on the turning movement at the intersection. $h^{\min}$ is a minimum headway between consecutive vehicles on the same lane. ε_{ni} is a random error term.

Vehicles that arrive to a detector activate it. The activation information is passed to the control logic. Vehicles advance this way from one detector to the next and to the stop line. At the stop line they enter a vertical queue. The queue length in each lane and at each time interval is monitored during the simulation. If, during any time interval, the queue length exceeds the distance between the stop line and one (or more) of the upstream detectors, the relevant detectors are activated for that interval. During the effective green time, vehicles are discharged from the queue at the saturation flow rate deterministically and according to the first-in-first-out (FIFO) rule. The saturation flow is an input to the model and may be different for each lane depending on the turning movement.

Pedestrians

FIGURE summarizes the movement of pedestrians through the intersection.

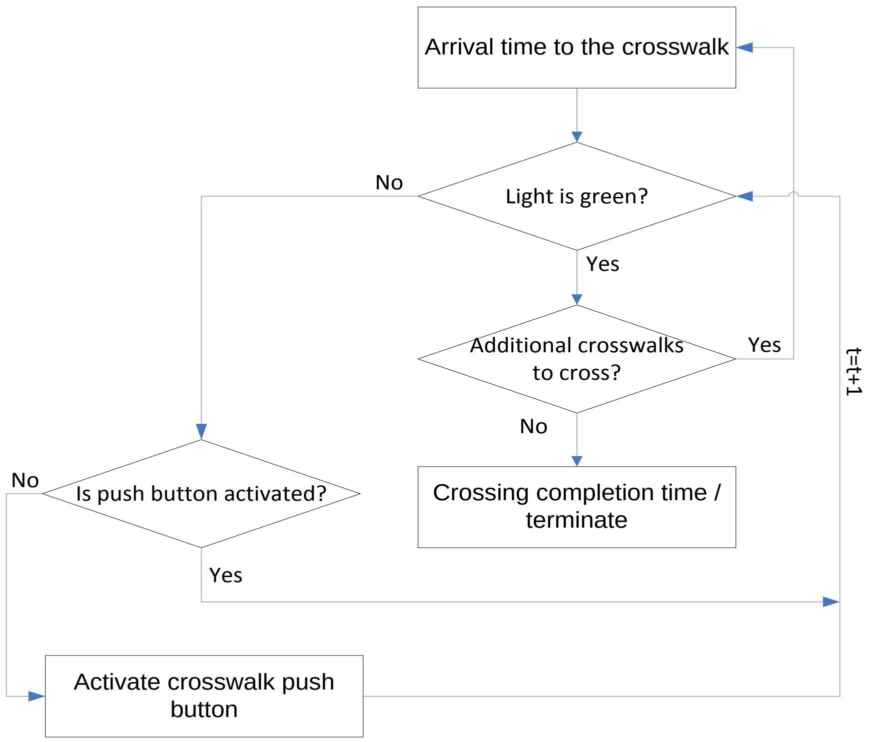

FIGURE 3: Pedestrians movement model

The numbers of pedestrians arriving at crosswalks during a simulation time step follow the Poisson distribution according to the assumed mean flow. During red light phases for a given crosswalk, the first arriving pedestrian activates the relevant push button, if one exists. Once the light turns green, the waiting pedestrians start to cross the intersection. Pedestrians that arrive during the green light phase cross the intersection without any delay. Crossing times are calculated assuming constant pedestrian speeds:

$$t_p^c(i) = \max\left(t_p^a(i), t_p^g(i)\delta_p^{ar}\right) + \frac{l(i)}{v_p} \tag{2}$$

Where, $t_p^c(i)$ and $t_p^a(i)$ are the crossing completion time and the arrival time to the intersection of pedestrian p at crosswalk i, respectively. $t_p^g(i)$ is the beginning of the green phase after the arrival of the pedestrian at the crosswalk. δ_p^{ar} is an indicator that takes the value 1, if pedestrian p arrives at the crosswalk during a red light phase, and 0 otherwise. $l(i)$ is the length of crosswalk i. v_p is the walking speed of pedestrian p.

In the current implementation, it is assumed that all pedestrians have the same walking speed. In case of multiple crosswalks in certain intersection leg, the time of crossing completion of the first crosswalk is the arrival time to the next one.

Detectors System

Actuated traffic signal plans use information about traffic flow in order to allocate green times. Various detection technologies, such as loop detectors and video are implemented to support vehicle presence detection tasks. Presence detection is activated when a vehicle is within a detection zone. Presence information is used to identify vehicle demand, as part of phase skipping logic or to initiate calls for extension of the green time of a phase. With transit priority plans detectors are used to identify approaching transit vehicle, to help predict its arrival time at the stop line and to detect the release of the vehicle from the stop line. The detection of pedestrians is limited in most cases to push buttons.

Detectors in MESCOP may be placed at any location on the approach to the intersection or downstream of the stop line. The representation is not sensitive to the detection technology.

Control Logic

The traffic control logic implementation incorporates functions for both detection and control tasks that support representation of various traffic control plans. The detection functions include:

- Presence detection – a function that queries the simulation model for presence of a vehicle on the detection zone at a specific time point.
- Demand detection – a function to determine whether or not the detector has been activated over a period of time.
- Queue detection – a function to determine the occupancy of a detector over a time interval.
- Gap detection – measuring the time that has passed since the last presence detection.

The detection results are used by the various control functions to adjust the signal timings. Implemented control functions include both actuated and transit priority plans. The actuation functions are:

- Phase skipping – a function that enables to skip a specific phase when a negative result is returned by the demand detection function on the relevant detectors.

- Phase extension – a function to extend the green light of an active phase if certain conditions are met. These conditions may be that gap values are below a pre-specified threshold or that queue values exceed their thresholds. The gap threshold may be different for each movement.

Transit priority control logic functions include:

- Arrival expectation – a function that estimates the arrival time of the transit vehicle to the stop line. Time estimation is based on the time that the presence of the vehicle was detected and an assumed approach speed.
- Phase early termination – Active phases may be shortened in order to allow a transit phase that follows to start sooner.
- Early transit phase start – Changes in the phase timing is planned such that the transit phase would start a few seconds before the expected arrival of the transit at the intersection.
- Phase insertion – Activating a transit phase out of the normal phase sequence.
- Priority cancelation – occurs when the transit vehicle is discharged from the intersection and its presence is detected on a checkout detector located downstream of the stop line. Priority may also be canceled if a transit vehicle is not detected at the stop line a certain time after it was expected to.
- Compensation – a function that guarantees a minimum green time to certain movements or phases. It measures the cumulative green time provided to a movement or phase within a certain period of time. If needed, the green time is extended to meet a minimum threshold.
- Queue length override – a function that aims to avoid long queues in the minor approaches. When the phase of that approach is active, transit priority functions are disabled, if the relevant queue detection value exceeds a certain value over a period of time.

Performance Measures

A variety of measures of performance can be derived from the model output. The base performance model used with the model is average person delay. The delay for each vehicle is computed by the time that lapses from the time it enters the queue and the time it is discharged from it. Each vehicle is assumed to have a certain number of passengers in it, depending on the vehicle type. Pedestrian delays are computed by the time that lapses from the arrival time to the first crosswalk they need to cross, to the crossing start time (at the beginning of green light to the crosswalk or the time of arrival at the crosswalk, if the light indication is green) of the last crosswalk they need to cross. Considering multiple simulation runs, the average person delay at the intersection is given by:

$$d = \frac{1}{R}\sum_{r}\frac{\sum_{i}\sum_{n} d_{nr} N_i \delta_{ni}}{\sum_{i}\sum_{n} N_i \delta_{ni}} \quad (3)$$

Where, d is the average person delay. d_{nr} is the delay for vehicle (or pedestrian) n in simulation run r. N_i is the number of travelers in a vehicle of type i (by definition 1 for a pedestrian). δ_{ni} is an indicator variable that takes the value 1 if vehicle n is of type i (car, various bus types, pedestrian). R is the number of replications made.

Optimization

The optimization was performed by using a Genetic Algorithm [20]. Genetic Algorithms (GA) is a heuristic approach that is widely used to solve problems with objective functions that are difficult or infeasible to formulate analytically. It has proved to be an effective method to optimize signal plans ([1], [2], [10] ,[11], [12], [15], [17], [18], [19]). GA work with a population of points (i.e. values of control parameters), called individuals. Each individual represents a possible solution to the problem. The evolution usually starts from a population of randomly generated individuals. The fitness of every individual, i.e. the value of the objective function at that point, is evaluated. The next generation of individuals to be evaluated in the optimization process, are selected based on the fitness of the current individuals and using evolutionary principles. Three basic operators are used: reproduction, crossover and mutation. Reproduction involves generating an individual in the new generation that is similar to an one in the current generation. In crossover, new individual is created as a mix of two current individuals. Individuals with high fit have higher probabilities to participate in reproduction or cross-over. In mutation, random values are inserted in new individuals. This is used to support exploration of areas that have not been searched previously in order to avoid local optimum. Parameters of the GA algorithm include selecting the number of individuals in each generation, the number of generations and the probability values for the operators. Selection of proper values is essential for efficient GA operations [21]. Termination criteria of the GA algorithm often consider convergence thresholds for the fitness value (e.g. the difference between the best solutions in consecutive generations or the difference between the best and the average solution in the current population).

SYSTEM APPLICATION

Intersection layout

The integrated system is demonstrated with an application to the planned control in the intersection of Haatzmaut Avenue and Hayat Street in Haifa, Israel. The intersection is shown schematically in FIGURE . The planned control of this intersection is fully actuated and incorporates transit priority for a Bus Rapid Transit (BRT) line that crosses the intersection in both main directions (movements 2 and 6). In addition to the BRT movements, there are four vehicle movements (1, 3, 5 and 7) and seven pedestrian signalized crosswalks (a through f). The movements in the intersection are organized in three signal phases (A, B, C) as shown in FIGURE . Phase A is the main one. It provides green light to the BRT vehicles and to through vehicles from both main directions. Phases B and C provide green time to movements on the minor approach from the south and from the north, respectively. The right turn movement in phase B must yield to pedestrians crossing crosswalk g. Presence detectors are located on the minor approaches. These are used for demand (D1 and D5), extension (E1 and E5) and queue detection (Q1) tasks on the relevant phases. The demand detectors are located at the stop line. The Extension detectors are located 10 meters upstream of the intersection. There are two detectors on the eastbound BRT approach (DPT21, DPT22) and one of the westbound BRT approach (DPT62). These are used to identify an approaching BRT vehicle and in the arrival expectation task. An initial arrival expectation is estimated when the vehicle arrives to DPT21. This expectation is updated when the transit vehicle is detected at DPT22. The signal plan and timing are adjusted accordingly to provide priority to approaching BRT vehicles. DPT21 is located 300 meters upstream the intersection. DPT22 and DPT62 are located 100 meters upstream the intersection. There is only one detector on the westbound approach because of the presence of a BRT station close to the intersection. Both approaches also include priority cancellation detectors downstream of the intersection (DPT23 and DPT63). The four Pedestrian crosswalks on the major approach are activated by push buttons (P_d, P_e, P_f, P_g). The design traffic flows in the intersection were estimated from traffic counts and updated with results from a traffic assignment model. TABLE 1 presents the design flows in the various movements within the intersection. The total flow is 3194 vph.

Control logic

The control logic is designed to minimize delay for BRT vehicles while limiting the delay to vehicles on the minor approaches. Phase A is served as a default. Phases B and C are activated only if vehicle presence is detected on the approaches served by these phases. When a BRT vehicle is detected, depending on the active phase, the control logic examines the need for early termination of phases B or C, or extension of phase A. BRT priority is overridden if queues are detected in Q1, if a maximum pedestrian waiting time is exceeded, or if phases B or C require compensation (i.e. they did not each get a minimum cumulative green time within a certain time period). The cycle length is fixed as this intersection is coordinated with other ones. The remaining green time after termination of all phases and until cycle completion is allocated to phase A. The control logic includes 14 parameters, which can be classified into several groups:

1. Cycle length and splits Parameters (7 parameters): These include the cycle length and the minimum and maximum green times for each of the three phases.
2. Pedestrians-related Parameters (1 parameters): The maximum length of time that a pedestrian indication may be red consecutively. This parameter defines the maximum waiting time for a pedestrian at a single crosswalk. Normally, the same value is used for all pedestrian crosswalks at the intersection.
3. Transit Priority Parameters (1 parameter): An early green parameter defines the time, in seconds, between the start of a transit phase and the estimated arrival time of a BRT vehicle to the stop line. It is designed to allow BRT vehicles not to stop or slow down in the approach to the intersection, even if they arrive earlier than expected.
4. Compensation parameters (2 parameters): These are used for ensure existence of sufficient green times to vehicles in the minor approaches. The parameters are the cumulative minimum green time that must be provided to a movement or phase, and the time period, expressed in number of cycles, within which it is measured (e.g. 10 seconds green time within two cycles). If the condition is not met, the green time for the phase is extended to meet the minimum value.
5. Queue detection parameters (2 parameters): This functionality uses information from detector Q1 to prevent creation of long queues of vehicles in the northbound approach. The relevant parameters are the location of the detector upstream of the stop line (expressed in terms of number of vehicles in the queue of to that point) and the length of a period of continuous vehicle detection in Q1 that activated the queue condition (leading to immediately providing green time to phase B).
6. Phase extension parameters (1 parameters): a maximum value on the gap time between activations of the detector by two consecutive vehicles that still triggers extension of the current phase. In this case, this parameter is relevant to the two minor approaches.

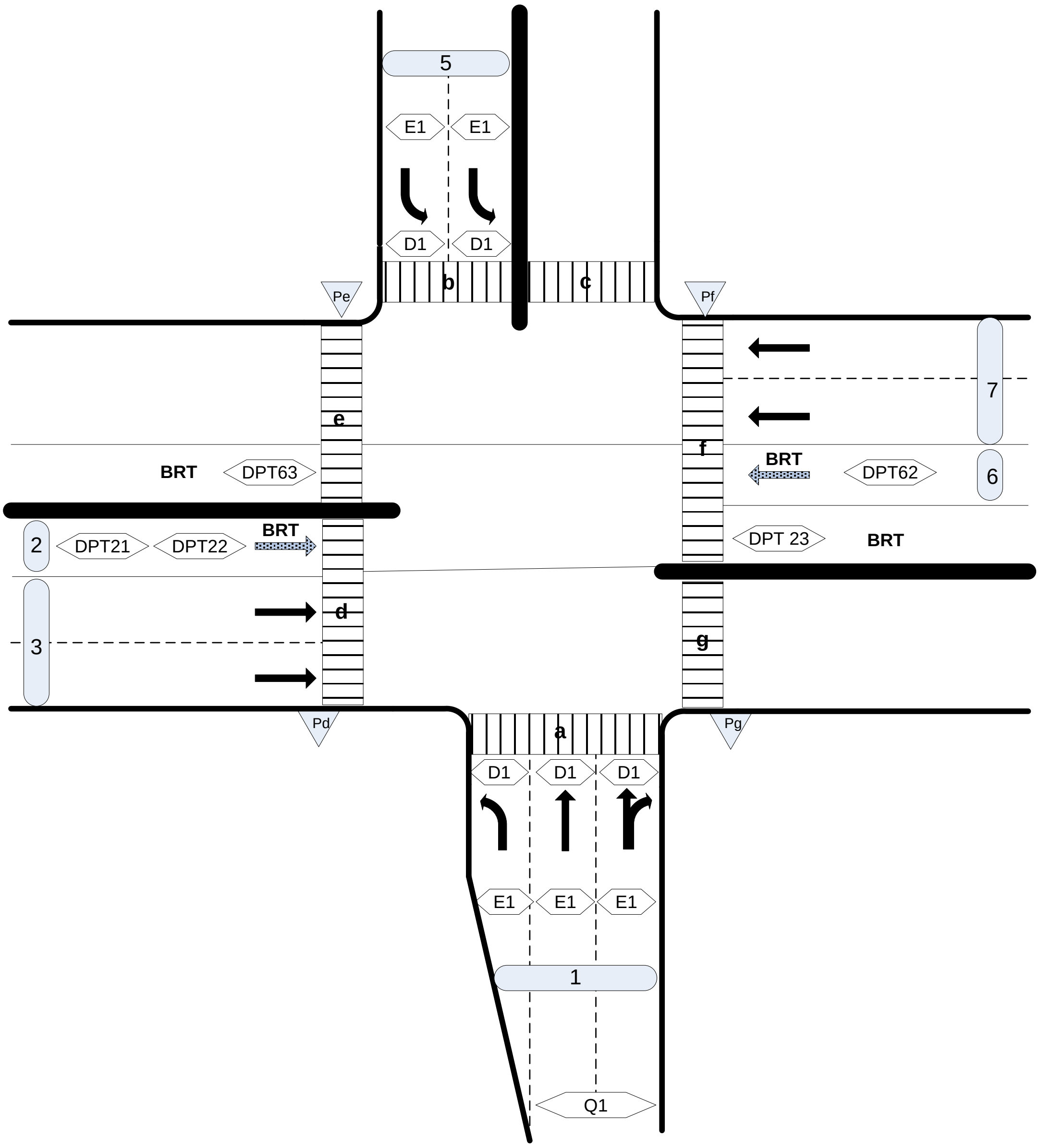

FIGURE 4: Case study intersection

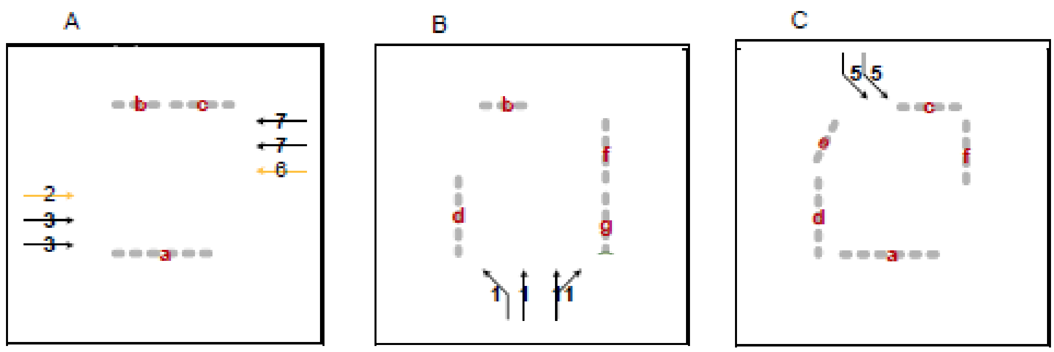

FIGURE 5: **Control phases at the intersection**

TABLE 1: Traffic flows at the intersection

movement	1L	1T	1R	2T	3T	5L	6T	7T
Traffic flow (vph)	115	144	45	30	1058	59	30	1713

Evaluation of Intersection Performance

FIGURE presents the average delay to road users at the intersection under different complexity of control plans. Four control plans with increasing levels of complexity were created: pre-timed, actuated without transit priority, unconstrained transit priority (eliminating the compensation and queue constraints) and constrained transit priority (the original design). The simpler control plans were created by eliminating functionalities and conditions from the original constrained transit priority plan. In the pre-timed plan, the green times for all phases were set to their respective maximum green times from the actuated plans. The remaining time within the cycle was allocated to phase A.

The results are averages of 100 replications in each case. The average person delay [Equation (3)], which accounts for the numbers of passengers in various vehicles and for pedestrians, slightly improved when transit priority functions were implemented. The transit priority functions significantly reduced BRT delays (by 76% from pre-timed to unconstrained transit priority plan), at the expense of small increases in the delays of other vehicles and pedestrians (by 4% and 3%, respectively). The elimination of compensation and queues constrains greatly simplifies the control logic, but does not have any noticeable effect on the intersection performance. In addition, there is only a small difference (4%) in the delays measured with the pre-timed and actuated (without priority) plans. In this intersection, vehicle actuation occurs only in the minor directions. Thus, the difference between the two plans is that with the actuated plan, less green time is allocated to the minor phases when their demand is low. However, since the time allocated to the major approach phase is very high in this intersection, the advantage to the vehicles using this phase from a further increase in their green time is minimal.

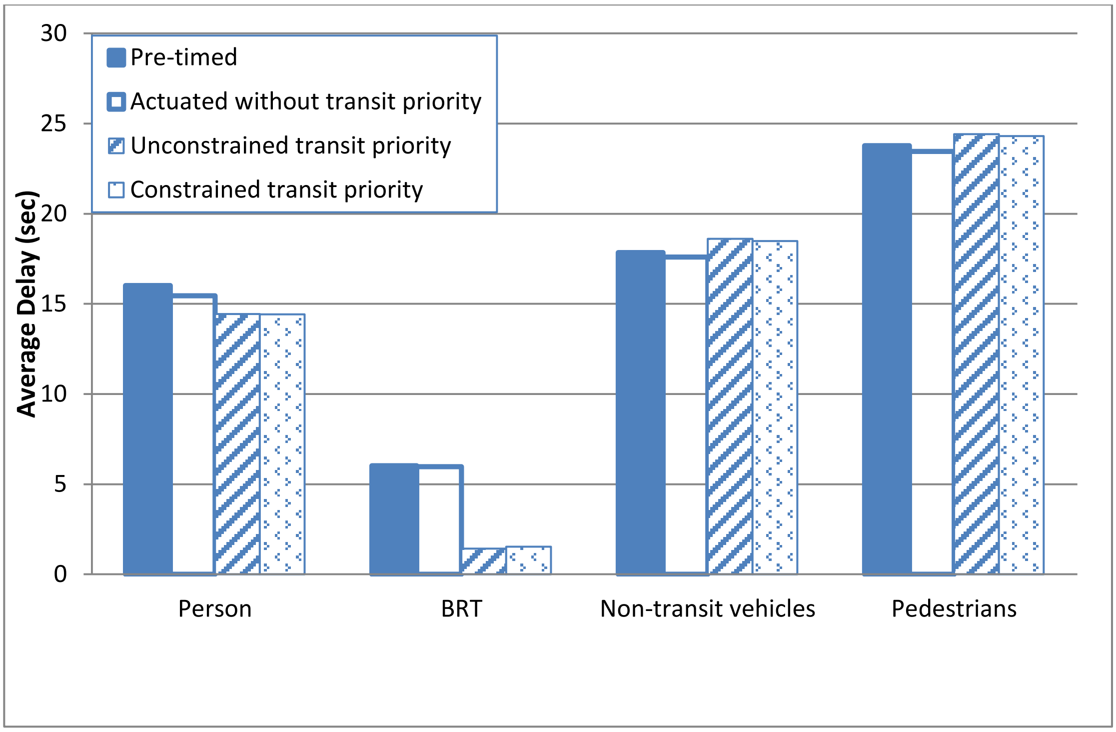

FIGURE 6: Average delay to road users under various control plans

Signal Plan Optimization

Optimization was performed for the intersection described above. The optimization uses an objective to minimize the delay per person assuming that the number of passengers in a BRT vehicle is 50 times that of cars. A genetic algorithm was used with a population of 100 individuals in each of 30 generations. Each individual was evaluated by 20 replications, thus requiring a total of 60,000 simulation runs. The optimization program took approximately five hours to run. A comparable optimization based on microscopic traffic simulation would not be computationally feasible, with an estimated running time of about a week.

TABLE compares the average delays for the various road users at the intersection under the original and optimal designs. The result shows a substantial improvement in the signal plan performance. The average passenger delay reduced by 28%, from 14.4 to 10.4 seconds. The reduction is attributed to significant reductions in the delays to non-transit vehicles and pedestrians without increasing the delays to BRT vehicles. Most of the design parameters significantly changed in a way that they allocate more green to pedestrians and non-transit vehicles as it described as following.

TABLE 2: Comparison between delays with initial and optimal designs

Road users	Average person delays (seconds)		Change (%)
	Initial parameters	Optimal parameters	
BRT	1.54	1.53	0
Non-transit vehicles	18.48	14.17	-23
Pedestrians	24.31	15.95	-34
All	14.41	10.40	-28

TABLE 3 compares between the original and the optimal parameter values. The parameter values for the original design are the ones proposed by the traffic engineers designing the control plan. The result highlights the importance of optimizing more design parameters and not only the basic ones of cycle time and green splits.

The cycle length significantly decreased from 110 seconds in the initial design to 72 seconds in the optimal design. This reduction is explained by the results of Webster [3] that show, in the context of fixed time plans, that the shortest practical cycle length is optimal. For the level of traffic demand at this intersection, longer cycles result in lower rates of vehicle discharges once the queue at the stop line has been fully dissolved.

The minimum green time to phase A increased in the optimal plan compared to the initial design. It should be noted that phase A may be extended, but only depending on arrivals of BRT vehicles. The demand of non-transit vehicles in the major approaches is high. The additional time for these vehicles improves the delays to these vehicles. However, the sensitivity of the delays to the value of this parameter is very low. This happens because the remaining green time, after all phases were completed and until the end of the cycle is all allocated to phase A. Therefore, the total green time allocated to phase A during a cycle depends on skipping or extending phases B and C. The minimum green parameter for phase A only affects the internal distribution of the green time within the cycle. In contrast to this, the maximum green to phase A has a substantial effect on delays. The optimal value of this parameter is lower compared to the initial design. A larger value of the maximum green time allows extension of phase A for long periods of time in order to provide transit priority. The extension decision does not take into account the flow of non-transit vehicles. Thus, large maximum green times for phase A may result in situations in which the flow through the intersection is low for an extended period of time between the BRT actuation in the far detector and until it arrives at the stop line and released from it. Lower maximum green times for phase A decrease delays to non-transit vehicles in the minor approaches and to pedestrians. Furthermore, lower maximum green values increase the flexibility of the control in that in some cases, it may be possible to complete transition to phases B and C and return to phase a before the BRT vehicle arrives at the stop line.

The optimal values of minimum green times to phases B and C decreased due to relatively low traffic demands. However, the optimal values of maximum green times increased in order to provide more flexibility to the plan to accommodate long queues on these approaches. Long queues on the minor approaches may result from early termination of phases B and C in order to provide transit priority to phase A. Allowing larger maximum green times to phases B and C reduced the need to activate the compensation mechanism and the occurrence of long queues in these approaches. These results are consistent with the results shown in FIGURE 6 that indicated that the signal plan could be simplified through elimination of compensation and queue constraints. This does not negatively affect the intersection delays, as these constraints can be implicitly satisfied by setting appropriate values for the cycle split parameters.

The maximum pedestrians waiting time parameter also decreased in the optimal design. This variable is the major factor causing the reduction in pedestrian delays presented in TABLE 2. It should be noted that there is a strong relation between the maximum pedestrian waiting time and the maximum green time to phase A, as they both limit the extension of phase A. A lower value of the maximum pedestrian waiting time parameter make it more likely that this parameter will be the one bounding the extension of phase A, and as a result decrease pedestrian delays.

The early green value increased in the optimal design. This improves the BRT delays by allowing it smooth dispatch from the intersection especially in cases that a BRT vehicle arrives at the stop line earlier than expected. This does not seem to have any noticeable effect on the delays of vehicles in the minor approaches, possibly due to the low demands on these approaches.

There was no change in the gap time values between the initial and optimal designs. Larger gap times would tend to lead to more frequent extension of green times in the minor approaches, even in times that the flow is lower. Lower gaps may result in not providing extension to phases B and c, even when these are useful.

TABLE 3: Initial and Optimal Signal Parameter Values

Design Parameter	Original Value	Optimal Value
Cycle Length [sec]	110	72
Minimum Green Phase A [sec]	11	19
Maximum Green Phase A [sec]	50	37
Minimum Green Phase B [sec]	10	5
Maximum Green Phase B [sec]	10	13
Minimum Green Phase C [sec]	10	5
Maximum Green Phase C [sec]	10	12
Maximum Pedestrians Waiting Time [sec]	138	110
Early Green Start [sec]	0	5
Cumulative Green Time [sec]	18	12
Compensation Period of Time [cycles]	2	4
Queue Length [vehicles]	30	27
Maximum Queue Time [sec]	5	15
Gap Time [sec]	3	3

Computational Performance

As noted above, computational efficiency of the simulation model is essential in order to support optimization of complex traffic signal plans with acceptable running times. The size of the intersection in terms of numbers of traffic lights, movements and phases, lanes, detectors and push buttons clearly affects the running time. For a given intersection, the model running time is affected by the following:

1. Level of traffic flow, both vehicular and pedestrians.
2. Complexity of the signal control logic.

It is expected that the model running time increases with an increase in each of these factors. FIGURE 7 presents the running times of MESCOP as a function of the total traffic flow in the intersection. The total flow was changed by scaling the base flows up or down. For comparison, these running times are compared to those obtained using TransModeler, a widely used commercial microscopic simulation model [22]. The reported running times are averages of 100 replications in each case. As expected, the mesoscopic MESCOP model produces running times that are lower by an order of magnitude compared to those of the microscopic TransModeler. For the base demand, MESCOP running time is 0.31 seconds, which is a reduction of 95% compared to the 6.5 seconds running time for TransModeler. Moreover, the running time for MESCOP are only slightly affected by the level of traffic flow. The running times increases by less than 0.1 seconds (20%) from the lowest (800 vph) to the highest (7000 vph) levels of demand that were tested. In comparison, for the comparable increase in flows, the running time of TransModeler increases by 180%. These results are not surprising given the level of detail in the models. Nevertheless, they highlight the computational efficiency of MESCOP as a model to evaluate signal control plans especially in the context of signal plan optimization, which may require a high number of simulation runs.

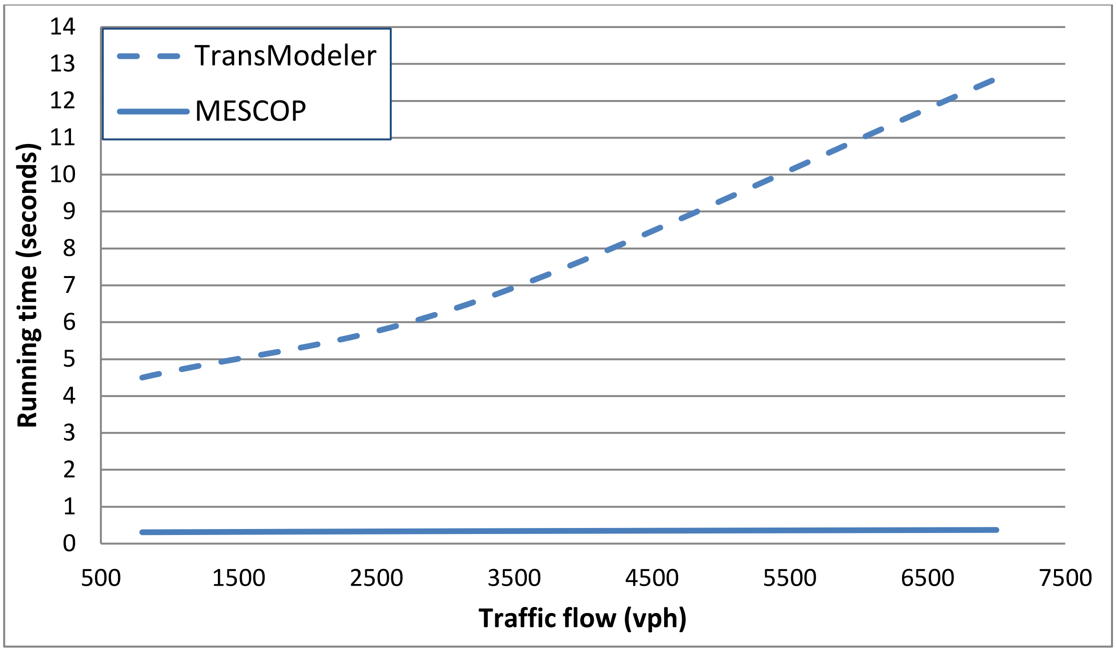

FIGURE 7: Running times for MESCOP and TransModeler

FIGURE 8 presents the effect of signal plan complexity on MESCOP running time. The running time results show that they are sensitive to the control logic complexity, but remain low in absolute values even with the most complex plan. The running time of the simulation with the constrained priority plan increases by 0.16 seconds (89%) compared to the pre-timed plan. The sensitivity of the running time to the control logic is explained by its internal breakdown among the model components, shown in TABLE 4. The reported results are for the case of constrained transit priority plan. The control logic execution accounts for most of the running time, while the traffic dynamics contribution is an order of magnitude smaller.

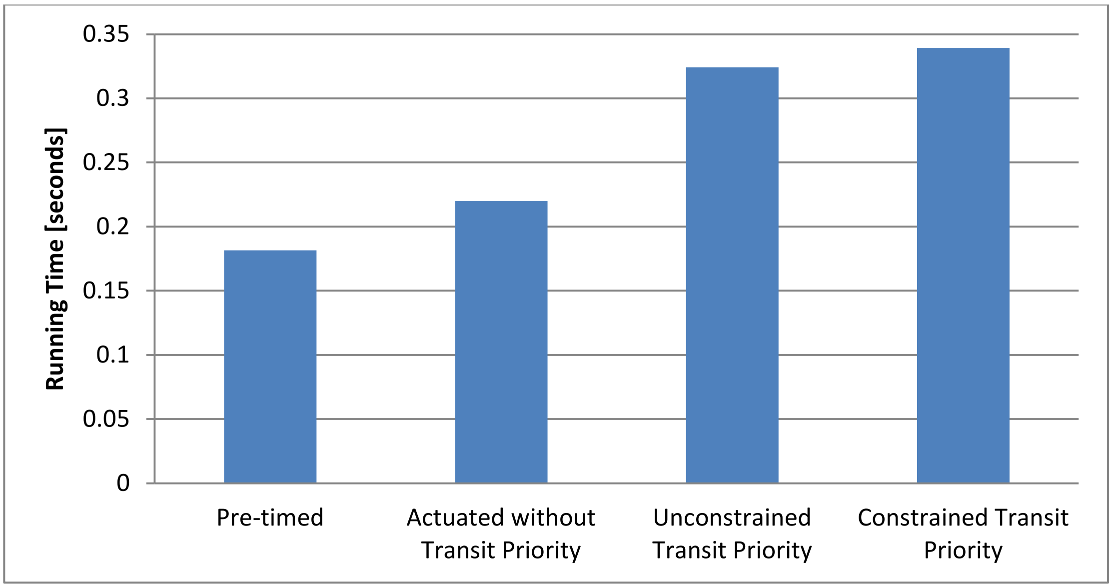

FIGURE 8: MESCOP running times for diffrent types of control logic

TABLE 4: Model Running Time for Each Basic Component

Model components	Running time [%]
Traffic dynamics	8.6
Control logic	81.1
Other (Input/Output)	10.3

The convergence properties of the simulation-base optimization system are plotted in FIGURE 9. The minimum person delay which is the "best" individual among 100 at each generation stabilized after about 10 generations. The optimal solution is the best plan achieved at generation 26. However, Convergence has been achieved approximately at generation 20 while the diffrence between the mean and the best person delay was stabilized.

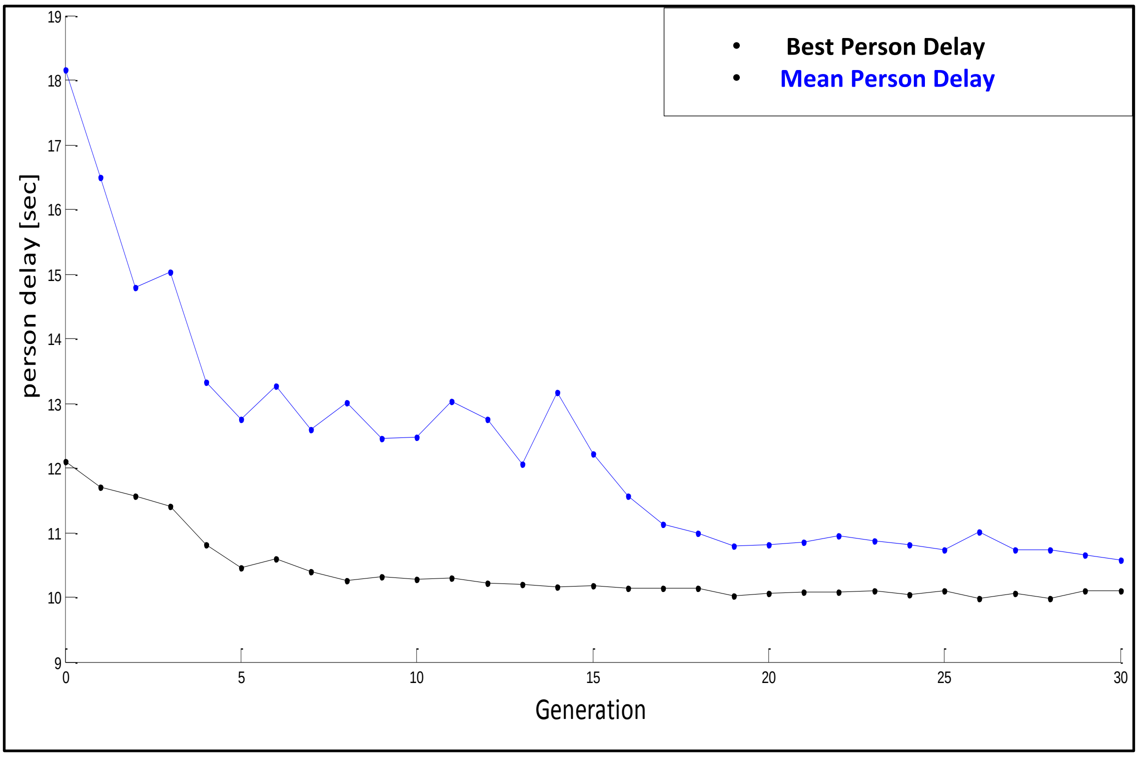

FIGURE 9: Convergence properties of GA

SUMMARY

This paper presented a simulation-based optimization system to optimize traffic signal plans. This system incorporates components of mesoscopic simulation model that represent vehicle movement, implementation of the detailed control logic and optimization algorithm. Application of the model to an intersection, which implements an actuated signal plan with transit priority, demonstrated substantial improvement in performance measure by optimizing large number of signal parameters. Using the mesoscopic traffic simulation, computational advantage has been achieved compared to microscopic traffic simulation models. The improved computational performance makes simulation-based optimization of traffic control plans feasible.

On-going research using this system focuses on methods to identify the most influential parameters within a control plan in order to reduce the dimensionality of the optimization problem. This may be achieved through various sensitivity analysis procedures. Other directions for work examine extension of the model to support application to multiple intersections, to support additional measures of performance and to allow optimization under uncertainty in traffic flows.

REFERENCES

1. Park B. B., Yun I., Evaluation of Stochastic Optimization Methods of Traffic Signal Control Settings for Coordinated Actuated Signal Systems, Research Report UVACTS-15-0-102. University of Virginia, Charlottesville, Jan. 2006.
2. Park B. B., Messer, C. J., Urbanik, T. II., Enhanced Genetic Algorithm for Signal Timings Optimization of Oversaturated intersections, Transportation Research Record 1727, TRB, 2000, pp. 32-41.
3. Webster F., Traffic Signal Settings, Road Research Technical Paper no. 39, Great Britain Road Research Laboratory, 1958, London.
4. HCM, Highway Capacity Manual, Washington DC: Transportation Research Board, National Research Council, 2010.
5. Husch D., J. Albeck., Trafficware SYNCHRO 6 User Guide, Trafficware, Albany, California, 2004.
6. Hale D. Traffic Network Study Tool – TRANSYT-7F, United States Version, McTrans Center in the University of Florida, Gainesville, January 2005.
7. Chang E. C., Messer C. J., Arterial Signal Timing Optimization Using PASSER II-90, Program User's Manual, Texas Transportation Institute, College Station, Texas, 1991.
8. Chaudhary et al., Passer V - Software for Timing Signalized Arterials, Texas Transportation Institute, College Station, Texas, 2002.
9. Little J. D. C., Kelson M. D., Gartner N. H., Maxband: A Program for Setting Signals on Arteries and Triangular Networks, Transportation Research Record, vol. 795, 1981.
10. Park B. B., Schneeberger J. D., Research Report UVACTS-5-0-4. University of Virginia, Charlottesville, Jan. 2003.
11. Foy M. D., Benekohal R. F., Goldberg D. E., Signal Timing Determination Using Genetic Algorithm, Transportation Research Record 1365, TRB, 1992, pp. 108-115.
12. Hadi M. A., Wallace C. E., Hybrid Genetic Algorithm to Optimize Signal Phasing and Timing. Transportation Research Record 1421, TRB, 1993, pp. 104-112.
13. Rouphail, N., B. Park, and J. Sacks. Direct Signal Timing Optimization: Strategy Development and Results, NISS technical report number 109, 2000.
14. Park B. B., Rouphail N. M., Sacks J., Assessment of stochastic Signal Optimization Method Using Microsimulation, Transportation Research Record 1748, TRB, 2001, pp. 40-45.
15. Stevanovic A., Martin P. T., Stevanovic J., VISSIM-Based Genetic Algorithm Optimization of Signal Timings, Transportation Research Record 2035, 2007, pp. 59-68.

16. Park B. B., Lee J., Stochastic Optimization of Coordinated Actuated Traffic Signal System Method Based on Shuffled Frog Leaping Algorithm, Transportation Research Record 2128, 2009, pp. 76-85.
17. Stevanovic J., Stevanovic A., Martin P. T., Bauer T., Stochastic Optimization of Traffic Control and Transit Priority Setting in VISSIM, Transportation Research record part C 16, 2008, pp. 332-349.
18. Stevanovic A., Stevanovic J., Kergaye C., Martin P., Traffic Control Optimization for Multi-Modal Operations in a Large-Scale Urban Network, IEEE Forum on integrated and Sustainable Transportation Systems, 2011, Vienna.
19. Yun I., Park B. B., Stochastic Optimization for Coordinated Actuated Traffic Signal Systems, Journal of Transportation Engineering, 2012, pp. 819-829.
20. Holland J. H., Genetic Algorithms, Scientific American, 1992, pp. 66-72.
21. Kovvali, V.G., Messer, C.J., Sensitivity analysis of genetic algorithm parameters in traffic signal optimization. Proceedings of the 81st Annual Meeting of the Transportation Research Board, Washington, DC, 2002.
22. Caliper, TransModeler Traffic Simulation, Software User's guide version 1.1, Caliper Co., 2007.