European Transport \ Trasporti Europei (2022) Issue 86, Paper n° 1, ISSN 1825-3997

https://doi.org/10.48295/ET.2022.86.1

Adaptive Signal Control of an Isolated Intersection Using Stop-Line Detection

Sadguna Nuli [1*], Tom V. Mathew[2]

[1]*Vardhaman College of Engineering, Shamshabad, Hyderabad-501218, India*
[2]*Indian Institute of Technology Bombay, Mumbai-400076, India*

Abstract

Adaptive traffic signal controllers offer better signal time management especially when the traffic flow pattern is not uniform on all approaches. Traditional adaptive traffic controller use upstream or advance vehicle detection which works well in situations where traffic follows good lane discipline. However, when the spacing between intersections increases or in the case of complex geometry these systems may not be efficient. This is primarily because of the inability of traffic flow models to accurately estimate the traffic demand from the upstream detectors. Using stop-line detector information is best suited in such traffic conditions as they do not require any explicit prediction models. Furthermore, there are many intersections which works using stop-line detectors with preset maximum green timings as vehicle actuated controllers. These controllers can be easily converted into truly adaptive by changing their maximum green timings continuously with respect to changing traffic flow pattern. Hence, this paper proposes an adaptive traffic control model which uses stop-line detector information instead of upstream detector. The model aims at real-time allocation of green time through reinforcement learning; an approach originated from the machine learning community. This approach has the ability to learn relationships between signal control actions and their effect on the queue while pursuing the goal of maximizing throughput which is a distinct improvement over the traditional vehicle actuated system. To demonstrate the performance of the proposed model a typical four-way intersection with four-phase scheme is evaluated for various flow conditions with the proposed model as well as with the traditional vehicle actuated system. The results show improvement over traditional system, especially when the flow is near the capacity.
Keywords: Signal control, Adaptive, Stop-line, Detector, and Reinforcement learning.

1. Introduction

Adaptive traffic signal control systems are relatively easy to implement if the traffic is more or less homogeneous vehicle types and drivers maintains good lane discipline. These systems assume accurate measurement of vehicle counts using vehicle detectors kept upstream of the intersection and availability of reliable traffic models to predict the demand at the downstream intersection. However, these systems are inefficient if the road or junction geometry is complex and drivers do not follow the lane discipline. Inaccurate estimation of turning proportion for a given movement from the upstream detector

* Corresponding author: Sadguna Nuli (findguna@gmail.com)

information is another issue. Further, when the spacing between the intersections and the number of lanes increases, the platoon of vehicles disperse fast, thereby reducing the benefit of coordination. This can be addressed to a great extent by placing the detector at the stop-line eliminating the need of traffic prediction from upstream detectors and using the stop-line information for the computation of signal timings. In addition, estimations from upstream detector information are prone to be highly erroneous. Since the adaptive signal controllers using stop-line detector has good prospect of addressing non-lane-based mixed traffic, this paper is an attempt to develop an adaptive control system for vehicle actuated controllers using stop-line detector. First, a brief review of the existing studies is presented.

Traffic signals typically operate in one of the three different control modes: pre-timed, actuated, and adaptive control. In pre-timed control, all of the control parameters such as cycle length, phase splits, and phase sequence are fixed based on historic data. The techniques such as equalizing degree of saturation or balancing level of service (LOS) for all critical lane groups, and minimizing total delay are generally used in determining the signal timings. These are obtained from the past traffic conditions for various time periods of the day (e.g., peak hours, off-peak hours). However, it lacks ability to respond to current traffic fluctuations.

To address this limitation, vehicle actuated controllers are used. In this approach, each phase has a pre-specified minimum and maximum green time, and the actual green time required depends on the time required to clear the queue obtained from vehicle sensors (advance detection) or gap between successive vehicles (stop-line detection) (Nuli and Mathew, 2015). Actuated control strategy can address the limitation of the pre-timed control strategy in a sense that it can respond to the real-time traffic arrivals of the current green phase. However, this strategy does not take into consideration of the demand on other conflicting phases. i.e., green time allocation or termination of active phase is independent of demand on the remaining phases. This results in inefficient control especially when the vehicle arrival rate is significantly different across various conflicting phases.

In order to address this limitation, adaptive control strategies are proposed which looks ahead in time using prediction models. Most of the operational adaptive control models such as SCATS, UTOPIA, PRODYN, SCOOT, RT-TRACS, and RHODES are designed for corridor or network level operations (Robertson and Bretherton, 1991; Gartner et al., 2002; Mirchandani and Head, 2001). Therefore, the adaptive controllers developed for corridor or network level may not yield optimum result for isolated intersections. Adaptive control models for an isolated intersection can be broadly classified into heuristic, fuzzy logic, optimization, and reinforcement learning methods. Various heuristic models for estimating minimum green, green extension, green termination, and special rules for congested conditions were proposed (Lin, 1988; Owen and Stallard, 1995; Nalge et al., 2011). Heuristic methods work well for stable flow conditions, but they are inefficient under dynamic or over saturated flow conditions.

Fuzzy logic has been extensively used for various traffic movements such as one-way, two-way, turning movements (Wei et al., 2001), geometry (Chou and Teng, 2002), number of phases (Murat and Gedizlioglu, 2002), phase sequence, phase choice, and signal control such as extend or terminate (Trabia et al., 1999; Zhang et al., 2005), and uneven traffic (Chou and Teng, 2002). An obvious advantage of using fuzzy logic for traffic signal control is that it handles uncertainty in the state of the system more efficiently. Apart from this, fuzzy logic-based controllers require relatively minimal computational resources.

Although using heuristic or fuzzy logic methods requires minimum computation time, it is difficult to obtain the optimal solution under dynamic flow conditions. Therefore, adaptive traffic control systems to perform optimally, deterministic optimization models such as DYPIC, OPAC, and MOVA were used (Robertson and Bretherton, 1974; Gartner, 1983; Vincent and Peirce, 1988). Although above models claim exact or global solution, they assume deterministic state transition which results in inaccurate cost estimates. Since the traffic flow is stochastic, traffic signal control can be formulated as Markov decision process and solved by dynamic programming algorithms (Yu and Recker, 2006).

While Markov decision process models accounts for stochasticity in traffic flow, obtaining real-time solution is difficult, especially for large scale problems. This is due to the exponential increase in the size of state transition probability matrix and computation load. Therefore, researchers opted for approximate dynamic programming to have real time feasibility (Cai et al., 2009). Apart from high computational demand Markov decision process-based modelling techniques additionally requires prediction models to estimate new state of the system. Further, when the traffic has no lane discipline these estimations are highly erroneous while using upstream detection information.

Few researchers have applied reinforcement learning (RL) techniques for solving small scale problems of isolated traffic signal control (Abdulhai et al., 2003; Adam et al., 2009). The advantage of RL technique is that they do not require any model of the traffic system to estimate state-transition probabilities and traffic arrival prediction models. Reinforcement learning models can learn the state-transition probability matrix interactively from the system operations. Further, after the model is trained, its computational requirement is comparable to non-optimal methods (Xie, 2007). Recently, the problems of scalability and generalization associated with the conventional reinforcement learning was addressed by various approximations such as modified Q-Learning (El-Tantawy and Abdulhai, 2010) and neuro-fuzzy actor-critic reinforcement learning technique (Nuli and Subbarao, 2018). A comprehensive study on various aspects of RL such as learning methods and parameters representation by El-Tantawy et al., 2014 supports RL is probably the most suitable modeling framework for adaptive traffic control.

From the above discussion, following observations about adaptive traffic control system for an isolated intersection can be summarised as: (i) adaptive control of traffic signal for non-lane-based traffic can be addressed to a great extent by placing the vehicle detectors at the stop-line; (ii) however, suitable algorithms for adaptive control of traffic using stop-line detector information needs further study; (iii) since the traffic flow is stochastic in nature, assumption that the state transition is deterministic is not valid and will results in sub-optimal solution; and (iv) conventional reinforcement learning algorithms are not able to deal with large scale problems such as multi-phase signal control on real time.

To address these limitations, an adaptive traffic control using actor-critic reinforcement learning is proposed for an isolated intersection control using stop-line detection information. The advantage of this model is that it does not require any explicit traffic prediction model and it can easily incorporate adaptive feature to the traditional vehicle actuated controllers. The proposed model is henceforth referred as TRASCR-J (TRaffic Adaptive Signal Control using Reinforcement learning - Junction).

2. Methodology

The objective of the present study is to develop an adaptive traffic control system using stop-line vehicle detection information. Traditional vehicle actuated (VA) controllers are also adaptive in some restricted sense, where individual phase termination or extension is based on real-time field data. However, green extension is subjected to some pre-determined maximum green time. Further, the green extension and termination of a given phase is independent of the traffic state on other phases. Hence, the traditional vehicle actuated controllers are not fully adaptive. In order to develop a fully adaptive system, the strategy adopted in the present study is to determine the maximum green time based on the traffic condition experienced by all the phases and override the pre-set maximum green time for each subsequent phase. Hence, the proposed system works similar to the traditional vehicle actuated system except that the maximum green time for each phase is determined by a reinforcement learning model considering the traffic state from all the phases. It may be noted that, placing the detectors on the stop-line eliminate any prediction errors arising out of non-lane-based mixed traffic movements. Further, since the proposed system rely on the relative changes in the traffic pattern to modify the signal timing, it is reasonable to expect the detection errors, if any, will be distributed uniformly and hence the detection errors will not affect the performance of the system. The proposed architecture is shown in Figure 1. Note that the box with dashed line represents proposed reinforcement learning model (TRASCR-J) which receives the current utilized green time and throughputs, and compute the maximum green time for each phase of the subsequent cycle.

The proposed control model begins like traditional VA controllers by sup plying and initialization of various controller and model parameters (Figure 1). The controller scans at each time step whether the phase is active (GREEN) or not (RED). If the phase is active, it calculates gap between vehicles based on detectors information, or else it sets the next phase active. The controller terminates active phase (set signal to RED) if the threshold gap or maximum green criteria satisfies, or else it continues same phase. After completing first cycle, the controller estimates the state of the system with the help of latest greens (utilized), thereby switching from vehicle actuated control to adaptive control. The controller supplies latest greens of all the phases and corresponding throughputs to the model. The model then determines maximum green for each phase of the subsequent cycle using reinforcement learning. A brief description of the reinforcement learning is first presented and then how the adaptive signal control problem is modeled using reinforcement learning.

2.1. Reinforcement Learning

Reinforcement learning is an area of machine learning concerned with how an agent ought to take actions in an environment so as to maximize some notion of cumulative reward. The goal of a reinforcement learning agent is to collect as much reward as possible. At the beginning of learning, agent can choose any action a_t for a given state s_t, but as the learning progresses it minimizes exploration and opt for exploitation of the knowledge gained.

It is important to mention at this point, how the adaptive traffic signal control problem is mapped to various aspects of the reinforcement learning such as agent, state, critic value, reward, action, and action value. The agent is the proposed reinforcement learning model (TRASCR-J). The state of the system is represented by the actual or utilized green times (g_i) associated with each phase of the signal controller. Since the green termination is based on threshold gap detection at stop-line, the actual green can act as a proxy to the queue length. The critic value quantifies goodness of the current policy. In other words, it is the expected return from a given state by following the current policy. Hence, the throughput is considered as the critic value. The reward is defined as the increase in the total throughput, i.e., the difference between total throughput between two successive cycles. A positive value of the reward indicates that the total throughput was increased by this value after executing the action and vice versa. The control action of the agent is assigning maximum green time for each phase of the subsequent cycle. The action value represents the expected return from a given state by taking a specific action and following a current policy thereafter. The action value in this study is the expected throughput by adopting certain set of maximum green times from the subsequent cycles. The concept of this is given in Figure 2.

It may be noted that the conventional reinforcement learning representation of multi-phase signal control problem may result in large state space. For instance, assume that a typical four-phase traffic signal is operating with a cycle time range of 60 to 180 seconds and a minimum green of 10 seconds, then each phase will get a maximum green of 150 seconds. Thus, a single phase can take any green time between 10 and 150 seconds. Hence, a four-phase signal controller would experience as many as 140^4 possible states which show multiphase signal control is a large-scale problem. It may be noted that the controller may not experience all these possible states in an actual operation. Nonetheless, which all states will be used is not known a-priori and therefore the model should be able to handle any possible state. Such large-scale problems prohibit application of conventional reinforcement learning such as Q-learning and SARSA (Xie, 2007). In order to address such scalability and generalization issues associated with above techniques, the proposed model uses actor-critic reinforcement learning with gradient descent-based function approximation to learn the relation between various states and corresponding control actions (Grondman et al., 2012).

2.2. Working Principle

The reinforcement learning model begins by taking the current utilized greens associated with each phase and corresponding throughputs as input (Figure 2). The model then computes the green utilization defined as the ratio of sum of green time for all the phases to the maximum cycle length, as given below:

$$s^k = \frac{\sum_{i=1}^{F} g_i^k}{C_{max}}. \tag{1}$$

The model then estimates an expected throughput V_s^k of the k^{th} cycle from the green utilization s^k and the reward weight λ^k shown as below:

$$V_s^k = s^k \times \lambda^k. \tag{2}$$

Similarly, the expected throughput A_p^k corresponding to possible maximum green times is estimated from the current green utilization s^k and the associated weights w_p^k with each possible maximum green as below:

$$A_p^k = s^k \times w_p^k. \quad p \in Q \tag{3}$$

Where Q is the set of possible maximum green values. The maximum green values are the neighbourhood of the latest actual greens measured i.e., it is an increment to latest greens. The tuning of the reward weights λ^k and action weights w_p^k is done using a well-known gradient descent update rule, by first estimating temporal difference (Sutton and Barto, 1998) error δ as:

$$\delta = (V^{k+1} - V^k) + \gamma \times V_{s\prime}^k - V_s^k. \tag{4}$$

Where, V^{k+1} and V^k are the actual throughputs measured in the cycles $k + 1$ and k; $V_{s\prime}^k$ and V_s^k are the expected throughputs from the subsequent cycles estimated for the corresponding states of the green times using the weights updated in cycle k; and γ is discount factor for the expected throughput in the subsequent cycles. If γ is zero, then it takes only immediate reward and discard all subsequent additional rewards. On the other hand, if γ equal to one, then the model gives equal weightage to immediate as well as all other subsequent future rewards. The optimum value of this parameter needs to be found out by trial and error. The new weights are updated according to the action p as shown below:

$$\lambda^{k+1} = \lambda^k + \propto \times \delta \times s^k. \tag{5}$$

$$w_p^{k+1} = w_p^k + \propto \times \delta \times s^k. \tag{6}$$

In order to explore all possible green-maximum green time combinations (state-action space) and select the best value of maximum green time, a well-known ϵ-greedy policy is used (Abdulhai et al., 2003). A set of maximum green times (a_t) having highest expected throughput is selected from the various possible set of maximum green values for most of the time except for ϵ amount of time, when it selects a random action uniformly. The value of ϵ will be very high at the beginning (typically 0.9) to facilitate better exploration and is decreased gradually to very low value (typically 0.1) to facilitate exploitation as the system converges to optimal weights (El-Tantawy et al., 2014).

2.3. Algorithm

Various steps of the algorithm is shown in Figure 3. The algorithm begins with the initialization of various controller and model parameters. At each time step it calculates elapsed green time from the start of the phase (line 4). If the state of current signal phase $\emptyset_i$ is green, then the status of the respective detector d_i is obtained (line 6). If vehicle is present on the detector, then the vehicle count n is updated (line 8) and stores time of departure t_n (stop-line detector) of the vehicle (line 9). If the elapsed green time g_i is more than or equal to initial green time g_{ini}, then the green time is extended by unit-extension e_0 since the presence of vehicles on the detector indicate that the queue is not cleared (line 10). Further, it computes the gap h between vehicles using their departure information (lines 11-15). Algorithm terminates active phase if the gap is more than the threshold gap and current green is greater than the end of current extension or the current green is greater than the current maximum green time (lines 16-17) and sets next phase (lines 18, 30-32); otherwise, the current phase continues (line 33). If all the phases in the current cycle is completed then the cycle count k is updated, the current phase is set to the first phase, and cycle throughput V^k is computed (line 19). Once, the first cycle is completed, the reinforcement learning to estimate new maximum green begins (line 20). Algorithm computes expected throughput for all possible maximum green times A_p^k (line 21) and the average expected throughput V_s^k (line 22). The learning weights are updated in the next cycle according to gradient descent update rule (lines 23-27). Further, the maximum green for each phase of the subsequent cycle is calculated according to the action selected (lines 28 and 29). The algorithm terminates when the end of the control period (line 34).

3. Model Testing

The performance of the proposed traffic control system is evaluated using various measures of effectiveness such as average intersection delay, total throughput, and average queue length. The performance of the system without the proposed adaptive maximum green time (which is same as the traditional vehicle actuated control) is also evaluated and compared with the proposed adaptive control.

VISSIM, a scalable, high-performance microscopic simulator, is used for the evaluation. Because of its high-fidelity microscopic behaviour models, it has been widely employed in traffic engineering applications such as traffic flow, signal control, and evaluation of various Intelligent Transportation Systems (ITS) initiatives. In order to demonstrate the working of the proposed model (TRASCR-J), a typical four-phase isolated intersection as shown in Figure 4 is considered. The intersection has three lanes each in East-West directions, and two lanes each in North-South directions. To have better safety under mixed traffic conditions each approach is assigned to an individual phase. This intersection model is created in VISSIM environment by considering mixed vehicle type and non-lane-based vehicle movements after an earlier study (Mathew and Radhakrishnan, 2010). The traffic is represented by multiple vehicle types, their composition, and their special movements. The vehicle composition consists of buses (2.5%), trucks (3.0%), LCV (1.7%), cars/jeeps (25%), three-wheelers (17%), and two-wheelers (52%). The static and dynamic characteristics of nonconventional vehicles such as three wheelers (auto-rikshaw) are modelled by taking nearest standard vehicle such as car as the base model. Further, non-lane-based movements of vehicles are represented by setting the driving behavior of the VISSIM to place the vehicle anywhere on the lane, and permitting vehicle to overtake along left or right of a slower vehicle.

To evaluate robustness of the model, three traffic flow cases are considered, namely, under-saturated (v/c ~= 0.7), saturated (v/c ~= 1.0), and over-saturated (v/c ~= 1.2). Each case is characterised by two peaks separated by an off-peak which is analogues to morning peak, off-peak, and evening peak traffic flow pattern normally observed in urban areas (Figure 5). Thus, the total simulation period of 16 hours consisting of two peaks of about 4 hours each, one off-peak period of about 2.5 hours and the rest is transition or normal traffic (Figure 5).

The base line parameters of vehicle actuated controller such as minimum green time (g^{min}), maximum green time (g^{max}), unit extension (e_0), and threshold gap (h_{th}) are determined as 10, 45, 3, and 2 seconds respectively by minimizing total delay of the intersection using trial and error. Similarly, the model parameters such as γ and α are set to 0.4 and 0.1.

In order to examine whether the proposed model is able to adapt to changes in the traffic pattern, the allotted green time and the corresponding demand is measured. The demand is the number of vehicles approaching the intersection and obtained from the simulator by defining data collection point at the upstream. A truly adaptive control will propose green time in response to the current demand. Hence, the measured demand and the allotted green times on a minor (North) and a major (West) approach using the proposed model and the traditional vehicle actuated control for the saturated flow case is shown respectively in Figure 6 and Figure 7. The comparison indicates that though there is high demand, VA control is saturated to a pre-determined maximum green on both approaches, whereas the proposed TRASCR-J control is sensitive to the flow pattern. i.e., model is flexible to change its maximum green according to the flow pattern on each approach (phase).

The evaluation results from 16-hour simulation for each of the three cases of under-saturated, saturated, and over-saturated flow conditions are summarized in Table 1. The system performance measures such as delay, queue, and throughput characteristics are computed separately for total period, peak, and off-peak period for each case. These measures obtained from the proposed model is compared with VA control for each approach (phase) and also for the whole intersection. It can be observed from the above result that almost all the approaches and the whole intersection experienced reduction in delay for all the cases including total, peak, and off-peak periods. The maximum decrease of 26% is found on the East approach for over-saturated condition during peak period. However, delay is increased on certain occasions; for instance, an 85% increase is found on the South approach for under-saturated condition during its peak period. Nevertheless, such increase is observed only in 12 out of 45 cases (Table 1) and that too in the minor approaches (South and North). This is an expected situation, since the objective of the proposed model is to increase total intersection throughput and not individual approaches.

Consistent with delay results, queue is also reduced in most cases. A maximum of 38% reduction is observed on the North approach for saturated flow condition during its off-peak period. However, queue is increased on certain occasions; 10 out of 45 cases. For instance, a maximum of 129% increase is found on South approach for under-saturated case during its peak period.

Similar results are found with respect to throughput. Highest increase in throughput is noticed under over-saturated condition. For instance, a maximum of 19% increase is observed on the East approach during its peak period. However, there is no increase on many occasions during under-saturated conditions, possibly due to lack of demand. Overall, the throughput improved in all the cases except 10 out of 45 cases and is consistent with delay and queue length results. Therefore, it can be concluded from the above results that the proposed control model (TRASCR-J) is truly adaptive to the changes in flow pattern. This has resulted in an increased intersection throughput, decreased delay and queue lengths compared to the traditional vehicle actuated control.

4. Conclusion

This study is an attempt to provide fully adaptive feature to the stop-line detector-based vehicle actuated controller, especially for the traffic characterized by non-lane-based movement and the presence of mixed vehicle type. Placing detectors at the stop-line addresses issues such as inaccurate estimation of vehicle counts and unreliable traffic predictions associated with upstream detection techniques. Hence, in this study, traditional vehicle actuated controller is modified by proposing maximum values of green time by considering the traffic state from all other phases. These values are determined by an actor-critic reinforcement learning model which can handle multiple phases. Thus, the clear contribution of this study is the development of an adaptive traffic control model to the stop-line based vehicle actuated controller, especially to deal with non-lane-based vehicle movements. The performance of the proposed model is evaluated by obtaining various measures-of-effectiveness from a typical four-phase signalized intersection for under-saturated, saturated, and over-saturated traffic flow using a robust traffic simulator. The results indicate that the proposed model is able to change maximum green timings responding to the varying traffic conditions. Results are also obtained using traditional vehicle actuated controllers and comparison indicates that model performed much better for saturated traffic where the intersection delay is reduced by 5%, queue length is decreased by 8% and the throughput is increased by 6% for the overall period of 16 hours. Having shown the advantage of the proposed model in dealing with multiphase signal control and its feasibility to real time operation, future directions of research include extension to a corridor involving multiple intersections.

Notation

The notations used in this paper are described below.

$\emptyset_i$ = signal state of the i^{th} phase {GREEN or RED}{i=1,.....,F};
α = learning rate;
γ = discount rate;
δ = temporal difference error;
λ^k = weight associated with the critic value V_s^k in the k^{th} cycle;
A_p^k = p^{th} action value in the k^{th} cycle where p ∈ {1, ..., Q};
A_s = action selected where s ∈ {1, ..., Q};
C_{max} = maximum cycle time;
F = total number of phases;
Q = total number of actions;
V_s^k = critic value for a given state s in the k^{th} cycle;
V^k = cycle throughput in the k^{th} cycle;
V_i = throughput in the i^{th} phase;
d_i = detector state of the i^{th} phase {PRESENT or ABSENT};
e_0 = unit extension;
g_i = actual green for the i^{th} phase;
g_{ini} = initial green;
g_i^{max} = maximum green for the i^{th} phase;
g_i^{min} = minimum green for the i^{th} phase;
h = gap between vehicles;
h_{th} = threshold gap;
i = phase;
k = cycle count;
n = vehicle count;
t = current simulation time;
t_g = start of green;
t_n = arrival time of n^{th} vehicle;
w_p^k = weight associated with the p^{th} action value in the k^{th} cycle

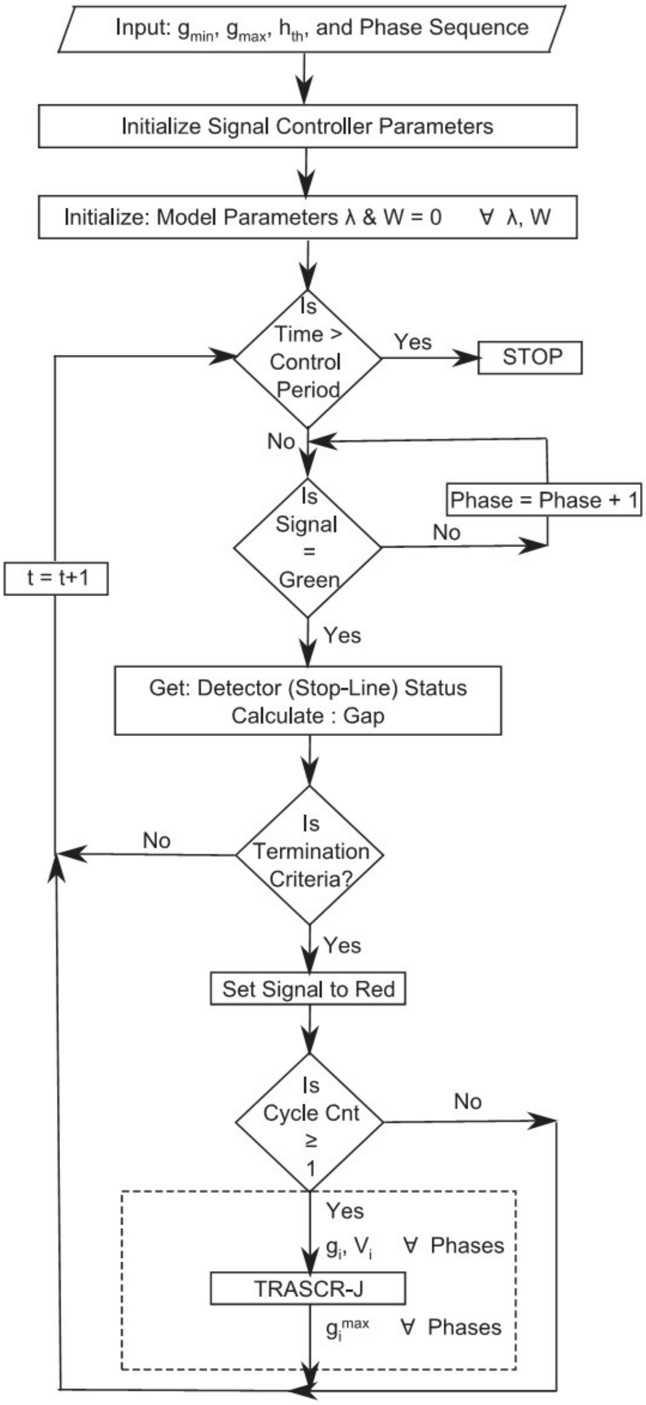

Figure 1: Modified VA controller with TRASCR-J model

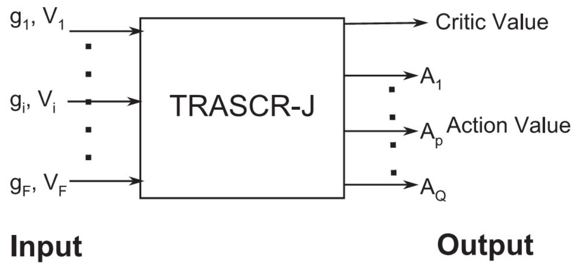

Figure 2: TRASCR-J model

```
Input: g_i^{min}, g_i^{max}, h_{th}, and phase sequence
Set Simulation period, t = 0, n = 0, i = 1, t_g = t, g_{ini} = g_i^{min}, λ^k = 0, and w_p^k = 0
repeat (for each step of the control period)
    g_i = t − t_g
    get status of φ_i
    if (φ_i = GREEN)  get status of d_i
        if (d_i = PRESENT)
            n = n+1
            t_n = t
            if (g_i ≥ g_{ini})  g_{ini} = g_{ini} + e_0
            if (n > 1)  h = t − t_{n−1}
            else  h = t − t_g
        else
            if (n ≥ 1)  h = t − t_n
            else  h = t − t_g
        if ((h > h_{th}  and  g_i > g_{ini})  or  (g_i > g_i^{max}))
            set φ_i = RED
            i = i + 1, n = 0
            if (i > F)  k = k+ 1, i = 1, and calculate V^k
                if (k ≥ 1)
                    compute action values A_p^k (Eq. 3)
                    compute critic value V_s^k (Eq. 2)
                    if (k ≥ 2)
                        compute learning error δ (Eq. 4)
                        update
                        reward weight λ^k (Eq. 5)
                        action weights w_p^k (Eq. 6)
                    select action A_s
                    calculate g_i^{max} = g_i + A_s
            g_{ini} = g_i^{min}
            set φ_i = GREEN
            t_g = t
        t = t + 1
until 't' is equal to control period
```

Figure 3: Modified VA control algorithm with TRASCR-J model

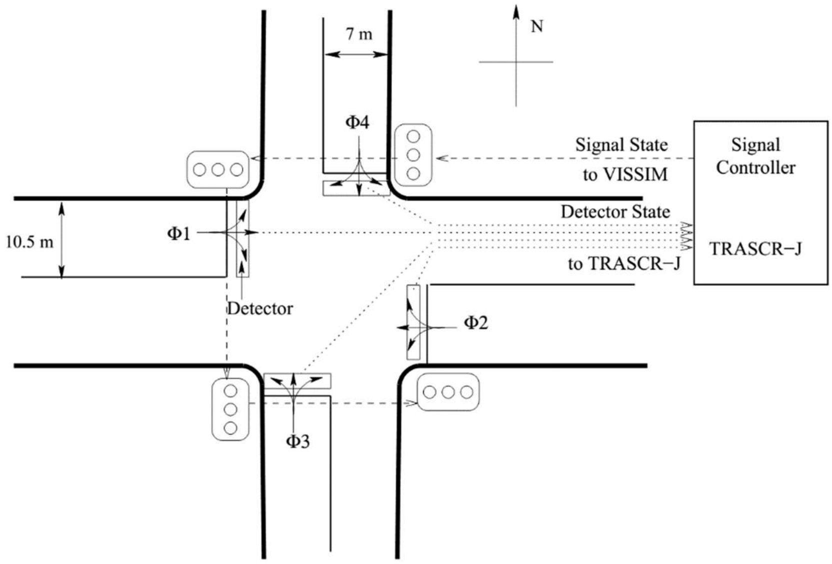

Figure 4: Illustrates geometry, phasing, and left bound traffic of a typical intersection

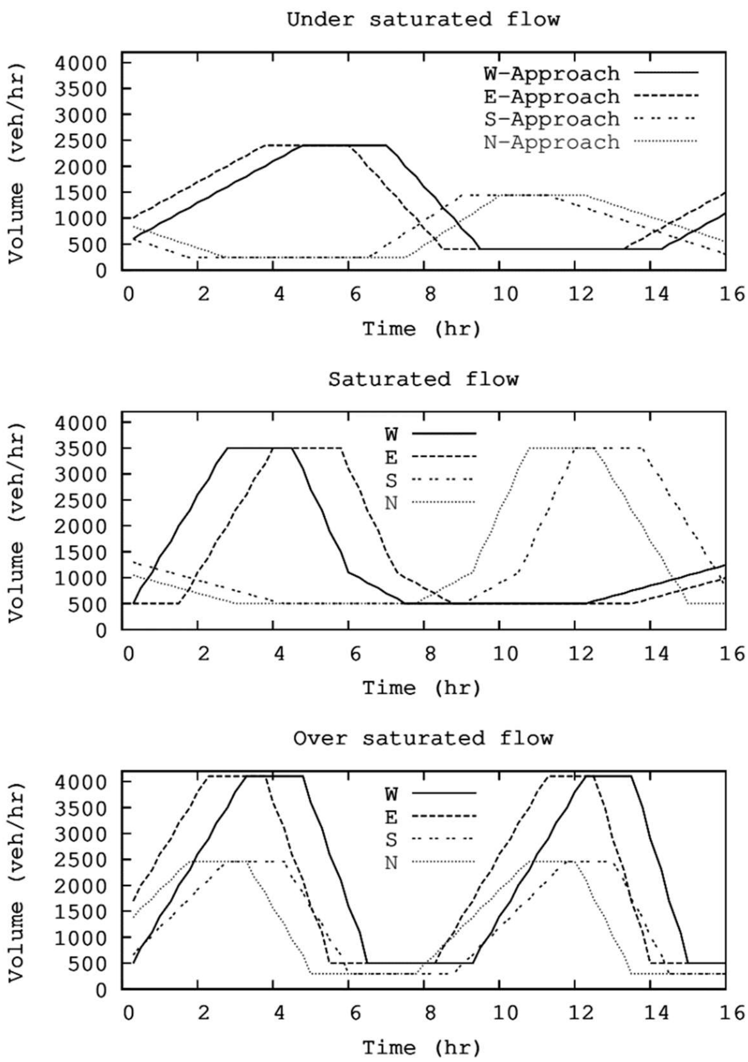

Figure 5: Flow pattern on various approaches of the study intersection

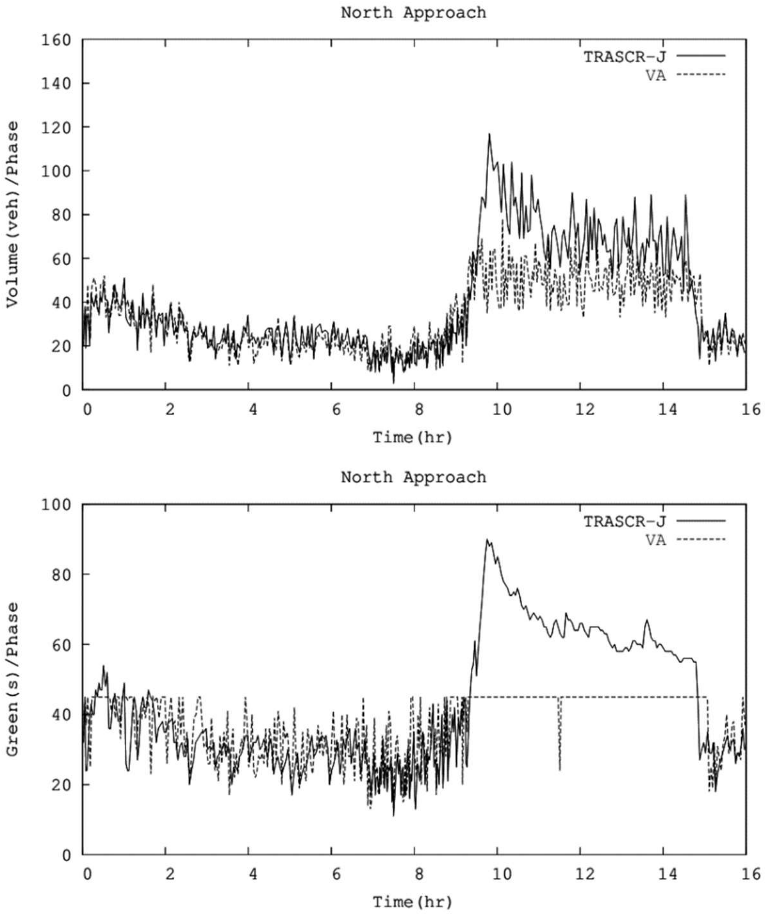

Figure 6: Comparison of volumes and green times between TRASCR-J and VA control on the North to South approach

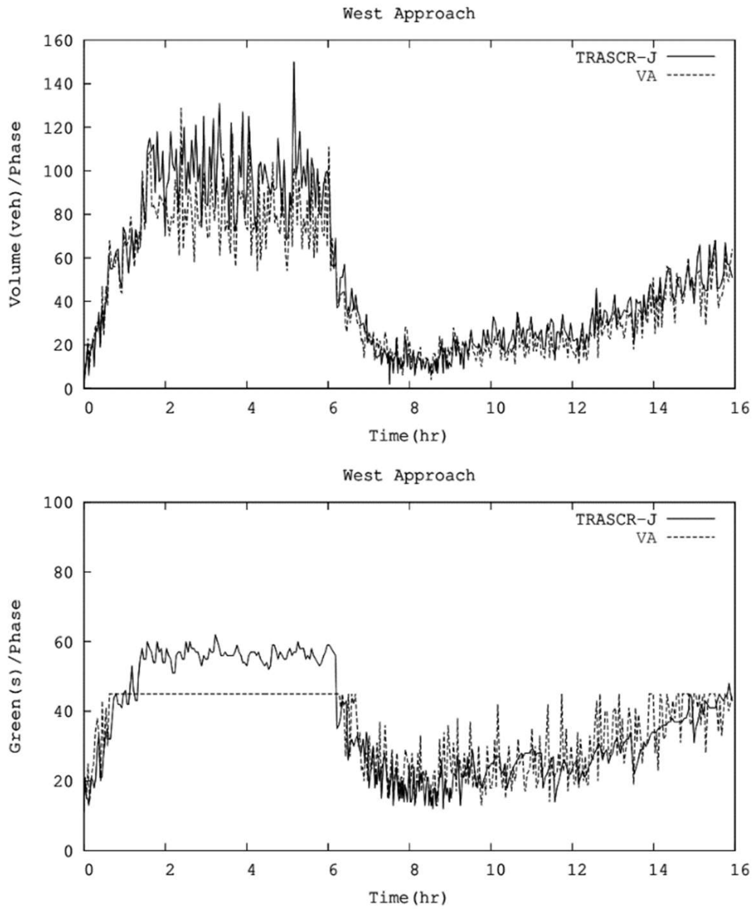

Figure 7: Comparison of volumes and green times between TRASCR-J and VA control on the West to East approach

Table 1: Comparison of delay, queue, and throughput with VA control for all the three flow conditions

	Under-Saturated		***Saturated***		***Over-Saturated***	
	Delay (s/veh)* ***total[1] (peak[2], off-peak[3])***	***% Change wrt VA***	***Delay (s/veh)**** ***total[1] (peak[4], off-peak[5])***	***% Change wrt VA***	***Delay (s/veh)**** ***total[1] (peak[6], off-peak[7])***	***% Change wrt VA***
West	51 (75, 31)	-9 (-13, -18)	84 (132, 48)	-3 (-11, -10)	106 (151, 15)	-18 (-15, -21)
East	47 (65, 30)	-11 (-22, -9)	89 (146, 57)	7 (-13, -18)	112 (155, 16)	-19 (-26, -6)
South	62 (89, 32)	19 (85, -11)	79 (66, 43)	-13 (27, 5)	203 (266, 16)	44 (37, -20)
North	52 (72, 28)	11 (53, -10)	77 (67, 39)	-13 (26, -23)	178 (288, 16)	28 (35, -11)
Junction	52 (71, 30)	-1 (-11, -12)	83 (124, 47)	-5 (-8, -15)	135 (191, 16)	-1 (-3, -14)
	Queue (m)* ***total[1] (peak[2], off-peak[3])***	**% change** ***wrt VA***	**Queue (m)*** ***total[1] (peak[4], off-peak[5])***	**% change** ***wrt VA***	**Queue (m)*** ***total[1] (peak[6], off-peak[7])***	**% change** ***wrt VA***
West	41 (90, 12)	-15 (-17, -14)	64 (133, 26)	-5 (-9, -17)	98 (139, 7)	-10 (-6, -36)
East	36 (77, 16)	-23 (-29, -6)	64 (155, 49)	-1 (-10, -18)	104 (154, 7)	-9 (-11, 0)
South	40 (32, 18)	8 (129, -10)	68 (29, 20)	-11 (29, 9)	126 (178, 5)	15 (14, -17)
North	27 (17, 18)	-10 (42, -10)	63 (26, 23)	-15 (23, -38)	113 (182, 6)	10 (12, -14)
Junction	37 (77, 16)	-13 (-21, -9)	65 (120, 32)	-8 (-7, -21)	107 (157, 6)	-2 (-2, -20)
	Throughput (veh)* ***total[1] (peak[2], off-peak[3])***	**% change** ***wrt VA***	**Throughput (veh)*** ***total[1] (peak[4], off-peak[5])***	**% change** ***wrt VA***	**Throughput (veh)*** ***total[1] (peak[6], off-peak[7])***	**% change** ***wrt VA***
West	18733 (11004, 1504)	0.0 (1.5, 0.1)	17910 (7048, 2680)	5 (3.4, -2.5)	23433 (7772, 1429)	10 (15, 4)
East	19747 (10722, 2416)	-0.1 (0.2, 0.0)	16716 (8494, 4340)	3 (8.4, 2.7)	22655 (8551, 1394)	12 (19, 6)
South	11349 (1617, 1690)	-0.1 (-1.8, 1.4)	14873 (2124, 2231)	6 (0.0, -0.7)	10662 (3628, 765)	-15 (-18, 4)
North	11718 (1237, 2214)	0.0 (0.2, 1.3)	14728 (1967, 3571)	9 (-0.2, 10.5)	11321 (3665, 1145)	-7 (-13, 8)
Junction	61547 (24580, 7824)	0.0 (0.6, 0.7)	64227 (19633, 12822)	6 (4.7, 3.0)	68071 (23616, 4733)	3 (5, 5)

Note: * indicate values obtained from TRASCR-J model
[1]total indicate 16 hrs period
[2]peak (3.0 to 8.0 hrs) [4]peak (3.0 to 7.0 hrs) [6]peak (9.3 to 13.7 hrs)
[3]off-peak (13.5 to 16.0 hrs) [5]off-peak (6.0 to 10.0 hrs) [7]off-peak (6.2 to 8.6 hrs)

References

Abdulhai, Baher, Rob Pringle, and G. J. Karakoulas (2003) "Reinforcement Learning for True Adaptive Traffic Signal Control." *Journal of Transportation Engineering (ASCE)* 129 (3): 278– 285.

Adam, Zain. M., Montasir. M. Abbas, and Pengfei. Li (2009) "Evaluating Green-Extension Policies with Reinforcement Learning and Markovian Traffic State Estimation." *TRR: Journal of the Transportation Research Board* 2128 (1): 217–225.

Cai, Chen, Chi. Kwong. Wong, and Benjamin. G. Heydecker (2009) "Adaptive traffic signal control using approximate dynamic programming." *Transportation Research Part C* 17: 456–474.

Chou, Chih-Hsun, and Jen-Chao Teng (2002) "A fuzzy logic controller for traffic junction signals." *Information Sciences* 143: 73–97.

El-Tantawy, S., Abdulhai, B., (2010) "An agent-based learning towards decentralized and coordinated traffic signal control." in: *13th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, 665–670, IEEE, Funchal.

El-Tantawy, Samah, Baher Abdulhai, and Hossam Abdelgawad (2014) "Design of Reinforcement Learning Parameters for Seamless Application of Adaptive Traffic Signal Control." *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations* 18 (3): 227–245.

Gartner, N. H. (1983) "OPAC: a demand-responsive strategy for traffic signal control." *TRR: Journal of the Transportation Research Board* 906 (906): 75–81.

Gartner, Nathan H, Farhad J Pooran, and Christina M Andrews (2002) "Optimized policies for adaptive control strategy in real-time traffic adaptive control systems: Implementation and field testing." *TRR: Journal of the Transportation Research Board* 1811 (1): 148–156.

Grondman, Ivo, Lucian Busoniu, Gabriel AD Lopes, and Robert Babuska (2012) "A Survey of Actor-Critic Reinforcement Learning: Standard and Natural Policy Gradients." *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews* 42 (6): 1291–1307.

Lin, F. B. (1988) "Comparative Analysis of Two Logics for Adaptive Control of Isolated Intersections." *TRR: Journal of the Transportation Research Board* 1194 (1194): 6–14.

Mathew, Tom V, and Padmakumar Radhakrishna (2010) "Calibration of microsimulation models for nonlane-based heterogeneous traffic at signalized intersections." *Journal of Urban Planning and Development* 136 (1): 59–66.

Mirchandani, Pitu, and Larry Head (2001) "A real-time traffic signal control system: architecture, algorithms, and analysis." *Transportation Research Part-C* 9 (6): 415–432.

Murat, Yetis. Sazi, and Ergun Gedizlioglu (2002) "A new approach for fuzzy traffic signal control." In *Proc., 13th Mini-EURO conference on Artificial Intelligence in Transportation Systems and Science, Bari, Italy,* .

Nalge, Megraj, Revathy Nair, Chandan kancharla, Tom. V Mathew, and Narayan Rangaraj (2011) "Real-Time Adaptive Control Policy for Intersection with

Downstream Detection.” In *18th World Congress on Intelligent Transport Systems,* No. 3254. Orlando.

Nuli, S., Mathew, T.V., (2015) “Vehicle actuated control for heterogeneous traffic using stop line detection.” *International Journal of Transport Economics, Engineering and Law* 57.

Nuli, S., Subbarao, S.S.V., (2018) “Adaptive signal control for non-lane based vehicle movements.” *International Journal for Traffic & Transport Engineering* 8.

Owen, L. E., and C. M. Stallard (1995) “Rule-Based Approach to Real-Time Distributed Adaptive Signal Control.” *TRR: Journal of the Transportation Research Board* 1683 (1683): 95–101.

Robertson, D. I., and R. D. Bretherton (1974) “Optimal Control of an Intersection for Any Known Sequence of Vehicle Arrivals.” In *Proceedings of the 2nd IFAC-IFIP-IFORS Symposium of Traffic Control and Transportation Systems,* 3–17. Amsterdam, Netherlands.

Robertson, Dennis I., and R. David Bretherton (1991) “Optimizing Networks of Traffic Signals in Real Time-The SCOOT Method.” *IEEE Transactions on Vehicular Technology* 40.no.1 (1).

Sutton, Richard S, and Andrew G Barto (1998) *Reinforcement learning: An introduction*. Vol. 1. MIT press Cambridge.

Trabia, Mohamed. B., Mohamed. S. Kaseko, and Murali Ande (1999) “A two-stage fuzzy logic controller for traffic signals.” *Transportation Research Part C* 7: 353–367.

Vincent, RA, and JR Peirce (1988) *’MOVA’: Traffic Responsive, Self-optimising Signal Control for Isolated Intersections*. Technical Report No.70. Transport Research Laboratory.

Wei, Wu, Yi Zhang, Jean Bosco Mbede, Zuo Zhang, and Jingyan Song (2001) “Traffic signal control using fuzzy logic and MOGA.” In *IEEE International Conference on Systems, Man, and Cybernetics,* Vol. 21335–1340. Tucson, Arizona: IEEE.

Xie, Yuanchang (2007) “Development and evaluation of an arterial adaptive traffic signal control system using reinforcement learning.” Ph.D. thesis. Texas A&M University.

Yu, X. H., and W. W. Recker (2006) “Stochastic adaptive control model for traffic signal systems.” *Transportation Research Part-C* 14: 263–282.

Zhang, L., H. Li, and P. D. Prevedouros (2005) “Signal Control for Oversaturated Intersections Using Fuzzy Logic.” In *84th Annual Meeting of the Transportation Research Board,* Washington DC, USA.