

Performance Comparison Between Queueing Theoretical Optimality and Q-learning Approach for Intersection Traffic Signal Control

Pitipong Chanloha*, Wipawee Usaha[†], Jatuporn Chinrungrueng[‡] and Chaodit Aswakul[§]

**Department of Electrical Engineering, Faculty of Engineering, Chulalongkorn University Patumwan, Bangkok, Thailand, 10330. Email: chanloha@ieee.org*

[†]*School of Telecommunication Engineering, Institute of Engineering, Suranaree University of Technology Muang district, Nakhon Ratchasima, Thailand, 30000. Email: wusaha@ieee.org*

[‡]*National Electronics and Computer Technology Center (NECTEC), Thailand Science Park, Klong Luang, Pathumthani, 12120 Thailand. Email: jatuporn.chinrungrueng@nectec.or.th*

[§]*Department of Electrical Engineering, Faculty of Engineering, Chulalongkorn University Patumwan, Bangkok, Thailand, 10330. Email: chaodit.a@chula.ac.th*

Abstract—This paper proposes the performance comparison for optimal traffic signal controls based on the following two frameworks: M/M/1 and D/D/1 queueing models, and Q-learning approach. Firstly, using the M/M/1 and D/D/1 models, the optimal split derivation has been obtained to minimise the mean waiting time of an intersection. Additionally, the Q-learning framework has been proposed in conjunction with the use of the macroscopic cell transmission model (CTM) to update the vehicle state dynamics upon Q-learning actions. The two approaches have been compared in terms of the network throughput and the average vehicle delay per completed trip in nine scenarios. The simulation results from the microscopic AIMSUN traffic simulator show that the Q-learning approach can greatly improve the intersection throughput and can significantly reduce the average vehicle delay per completed trip with the respective M/M/1 and D/D/1 approaches.

Keywords—Q-learning, queueing theory, cell transmission model (CTM).

I. INTRODUCTION

Due to the increase of traffic demands, the burden on the traffic control systems becomes a major concern. From the past history, the first sophisticated traffic control strategy has been manually operated by a policeman. The evolution of the control methodology for traffic signal grows rapidly. Fortunately, the growing emphasis on information systems and communication technologies is able to handle the traffic problem by using advanced traffic information and control systems. One of the most common goals of the researchers is to improve the efficiency of the traffic signal control by maximising the traffic throughput at an intersection.

Methods have been reported in the literature for controlling traffic signals at an intersection. Due to the capacity limitation of urban area, it is therefore critical to improve the performance of traffic network. A classical method is to analyse an isolated traffic intersection in the steady-state by adopting the queueing theory for traffic signal control. Yu and Stubberud [1] are the first to address the traffic signal control problem as a Markov decision process. Newell [2]

proposes an adaptive control strategy for traffic signal control by modifying the green time until a queue vanishes within a finite time horizon. Mirchandani and Ning [3] develop and evaluate an adaptive signal control method based on queueing theory. Their proposed method is based on the First-in First-out (FIFO) queueing systems. The method trying to minimise the average vehicle delay by using minimal weight matching has been proposed by Wunderlich et. al [4]. The shortcoming of these analytical methods is that they cannot deal with abrupt changes of traffic patterns.

To cope with the dynamic changes, a flexible approach has been proposed to learn good traffic signal controls from experiences gained gradually by interacting directly with the environment. This approach, referred to as a reinforcement learning (RL), is a class of machine learning related to the artificial intelligence [5]. RL is a class of unsupervised learning that has potentials to deal with traffic engineering problems [6]. Jacob and Abdulhai [7] addresses Q-learning which is an RL tool to deal with the highway traffic problems. For an isolated intersection control, [8], [9], [10], [11] consider Q-learning with different objective functions, whereas [12] investigates the green splits weighted by employing RL in order to minimise the number of vehicles in the system. Hong et. al [13] and Choy et. al [14] investigate the traffic signal control using a neural network which yields a high computational complexity and results in the impracticality in realistic scenarios. The literature above seek for an optimal traffic signal control for an isolated intersection. However, these RL approaches have been considered the individual movement of the vehicles in the microscopic level. Therefore, the computational burden becomes demanding.

To alleviate the computation burden caused by the microscopic behavior, a simple macroscopic model cell transmission model (CTM) is employed in this paper. Similar to our approach, Sadek and Basha [15] propose Q-learning for a traffic route guidance problem and uses also CTM. The CTM represents traffic flow phenomena for updating the vehicle

state of the system. However, this paper differs from [15] in that we are interested in the traffic signal control instead.

In addition, this paper compares the performance of Q-learning with the optimal split which has been derived for an isolated intersection with two conflicting flows whose steady state dynamics are captured by two queueing models, i.e. M/M/1 and D/D/1. The comparative results have been reported from our AIMSUN platform.

The rest of this paper is organised as follows. Section II presents the optimal split formula for the traffic queueing models. Section III formulates the Q-learning approach explainable in two parts. Firstly, the state of the system uses CTM to update flow dynamics. Secondly, the Q-learning algorithm is presented in Section IV. The simulation results are given in Section V and the conclusion is given in Section VI.

II. QUEUEING TRAFFIC MODEL

This section introduces a simplified queueing model with two buffers and single server which can be mapped into two conflicting flows in an isolated intersection.

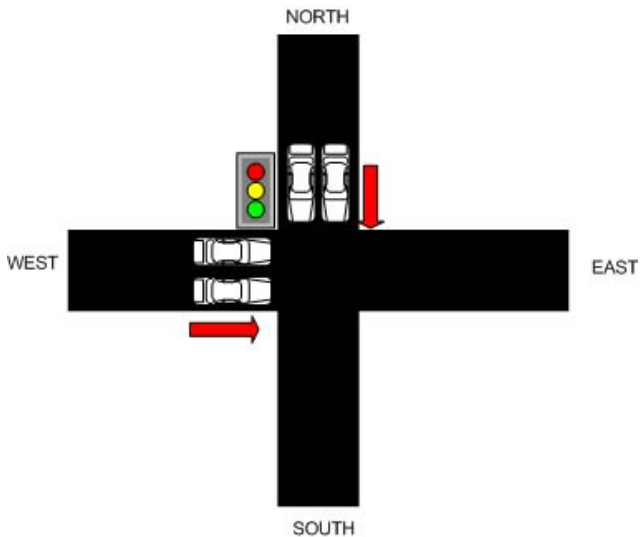


Figure 1. Model for two conflicting flows in an isolated intersection.

Fig. 1 illustrates an isolated intersection which serves two flows from west to east and north to south. Fig. 1 can be converted into a basic queueing model with two buffers and a single server as shown in Fig. 2, where λ_p denotes the traffic arrival rate of the system for direction $p = 1, 2$. Let μ be the intersection service rate of the system. Let w_p be the ratio of green time allocated to direction p (or its split) in a signal cycle. The objective here is to find the optimal split w_p^* that minimises the mean waiting time of the considered intersection system.

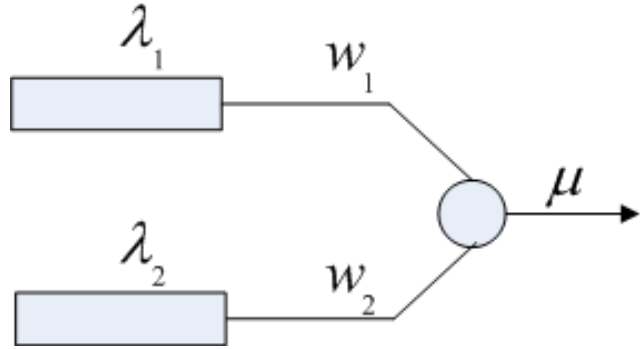


Figure 2. Queueing model with two incoming requests.

A. Steady State Analysis

The steady-state derivation is based on an M/M/1 queueing model where the vehicle arrivals in each direction are assumed to be an independent Poisson process and, during their green time period, each vehicle is assumed to spend exponentially distributed travel time through the intersection. As illustrated in Fig. 1, an intersection has two individual conflicting flows with mean arrival rates λ_1 and λ_2 , respectively. Let μ be the saturation flow rate, the flow rate at which vehicles can pass through a signalised intersection in a stable moving queue [16]. Let ρ_p be the offered load in direction p so $\rho_p = \frac{\lambda_p}{w_p \mu}$ for $p = 1, 2$. To guarantee the stability condition of the system, it is assumed that the intersection's saturation flow rate is greater than the total input flow rate from all approaching directions. Let L denote the total loss time value per signal cycle being normalised by the cycle period. Thus,

$$\sum_{\forall p} w_p + L = 1. \quad (1)$$

In the queueing steady state, the mean waiting time T_p in the system for direction p can then be obtained as follows [17]

$$\begin{aligned} T_p &= \frac{\rho_p}{1 - \rho_p} \\ &= \frac{\lambda_p}{w_p \mu - \lambda_p}. \end{aligned} \quad (2)$$

The total network delay T is given by

$$T = \sum_{\forall p} T_p \quad (3)$$

To minimise the total network delay, differentiating T in (3) with respect to w_1 and equating it to zero finally give:

$$0 = \frac{\partial}{\partial w_1} \left[\frac{\lambda_1}{w_1 \mu - \lambda_1} \right] + \frac{\partial}{\partial w_2} \left[\frac{\lambda_2}{w_2 \mu - \lambda_2} \right] \frac{\partial w_2}{\partial w_1}. \quad (4)$$

Therefore, the equation becomes

$$\frac{\lambda_1 \mu}{(w_1 \mu - \lambda_1)^2} = \frac{\lambda_2 \mu}{(w_2 \mu - \lambda_2)^2}, \quad (5)$$

where

$$\frac{\partial w_1}{\partial w_1} + \frac{\partial w_2}{\partial w_1} + \frac{\partial L}{\partial w_1} = \frac{\partial 1}{\partial w_1} \quad (6)$$

$$\frac{\partial w_2}{\partial w_1} = -1. \quad (7)$$

The optimal split from M/M/1 model $w_{1,MM1}^*$ and $w_{2,MM1}^*$ can be expressed finally as

$$w_{1,MM1}^* = \frac{\left[\Lambda (1 - L) + \left(\frac{\lambda_1 - \lambda_2 \Lambda}{\mu} \right) \right]}{(1 + \Lambda)}$$

$$w_{2,MM1}^* = \frac{\left[(1 - L) + \left(\frac{\lambda_2 \Lambda - \lambda_1}{\mu} \right) \right]}{(1 + \Lambda)}, \quad (8)$$

where $\Lambda = \sqrt{\lambda_1/\lambda_2}$. This result from equation (8) represents an optimal split weighted to each individual flow.

In a realistic scenario, the arrival process and queueing service time may not be Poisson and exponential. Thus, in this paper, another model, D/D/1, has also been used, where the incoming stream of vehicles arrive at a fixed deterministic rate and their service time through the intersection is assumed constant for every vehicle. The optimal split of D/D/1 model can be obtained similarly to the case of M/M/1 model and the final result becomes

$$w_{1,DD1}^* = \frac{\Lambda(1 - L)}{(1 + \Lambda)}$$

$$w_{2,DD1}^* = \frac{1 - L}{(1 + \Lambda)}. \quad (9)$$

III. Q-LEARNING MODEL

A. State Space Definition

Define \mathcal{S} as the state space of the intersection system with two conflicting flows. Let $\mathbf{s} \in \mathcal{S} \subset \mathbb{Z}_+^2$ be the state vector which represents the total number of vehicles waiting for the green light at the intersection. Let $s_p(t)$ be the state variable which represents the number of vehicles in direction p at time instance t where $p = 1, 2$. Therefore, the state space \mathcal{S} of all vehicle profiles in the system is given by

$$\mathcal{S} := \{\mathbf{s} = [s_1(t), s_2(t)]\}. \quad (10)$$

B. Cell Transmission Model

CTM [18] is here employed to update the Q-learning state dynamics. CTM captures the effect of control actions decided by Q-learning on the flow of vehicles in the system. The updating state depends on the green time allocated to each of approaching directions. The updating process of CTM can be summarised as follows.

1) *Sending Capability*: Let $y_p(t)$ be the number of vehicles that can pass through the intersection in direction p at time step t :

$$y_p(t) = \min \{s_p(t), q_p(t)\}, \quad (11)$$

where $q_p(t)$ represents the maximum flow rate at which vehicles can flow from their intersection upstream to downstream road segments along each direction p at time step t .

2) *Receiving Capability*: The receiving capability in CTM normally depends on the maximum flow rate $q_p(t)$ as

$$r_p(t) = \min \{q_p(t), \varepsilon_p(t)\}, \quad (12)$$

where $\varepsilon_p(t)$ denotes the residual capacity in direction p at time step t .

C. Action Space Definition

In each interval, the agent must select whether it would remain in the current signal indication or change it. The decision is referred to as an *action*. The action space, denoted by \mathcal{A} , is the set of all possible actions which the traffic signal controller of the considered intersection can take. Action $a \in \mathcal{A}(\mathbf{s})$ refers to the action which the agent can take at state \mathbf{s} .

D. Vehicle Delay

Vehicle delay is defined as the number of vehicles that cannot pass through the intersection. The vehicle delay accumulated at time step t can be expressed as

$$d_p(t) = s_p(t) - y_p(t). \quad (13)$$

Note that if the allocated green time can serve all traffic in $s_p(t)$, i.e., $s_p(t) = y_p(t)$, then there is no delay happening.

E. Reward Function

The aim of Q-learning here is to find the optimal policy that minimises the total network delay, which can be expressed in terms of the delay $d_p(t)$ at each time step t as:

$$\Upsilon(t) = \sum_{\forall p} d_p(t)$$

$$= \sum_{\forall p} (s_p(t) - y_p(t)). \quad (14)$$

Note that $q_p(t)$ is affected by the action a , which specifies the direction that receives the green light as follows

$$q_p(t) = \begin{cases} \mu & , a = p \\ 0 & , a \neq p. \end{cases} \quad (15)$$

Equation (15) represents an action which allows the vehicles to pass through the intersection in direction p at time step t . The state dynamics of CTM can then be updated in according to the chosen action in each time step as

$$s_p(t+1) = s_p(t) + x_p(t) - y_p(t), \quad (16)$$

where $x_p(t)$ represents the newly incoming demands in direction p at time step t .

Table I
PSUEDO-CODE OF Q-LEARNING ALGORITHM [5]

IV. Q-LEARNING ALGORITHM

Table I depicts the standard Q-learning algorithm which is applied to solve the problem formulated as an MDP.

In Table I, $Q(s, a)$ represents the *action value* function representing the average future reward expected to be incurred given that the action a has been taken at the state s [5]. According to the epsilon greedy policy, on the best apparent action will be selected with high probability of $1 - \epsilon$, and the other actions will be tried out randomly with a small probability of ϵ . Therefore, the best apparent action or *greedy* action is exploited most of the time. And with probability ϵ , the concept of exploration is to ensure that all of states are adequately visited. The parameter α is a small positive fraction, namely, the step-size parameter which influences the learning rate. Step-size parameter determines how much the new state action value tends towards the newly obtained reward and value of the next state-action pair. The parameter γ represents the discount rate which is used to determine the present value of future reward.

V. RESULTS AND DISCUSSIONS

In this section, the research finding from our results will be reported. The reported results are obtained from the MATLAB® and the AIMSUN. Firstly, the optimal split obtained from the CTM-based Q-learning, the queueing model M/M/1 and the queueing model D/D/1 have been calculated from MATLAB®. Secondly, the obtained optimal split is set to the allocation of the green signal in 1 cycle time to each direction where 1 cycle time is 120 seconds. The reported results from the AIMSUN are the network throughput, the link delay, the average vehicle delay per completed trip and the mean queue length, respectively.

For the system environments, suppose the length of each road from the entry of the road to the stop line is 800 metres. The maximum flow rate has been measured from AIMSUN under the condition that the vehicles are unaffected by the red signal. From the measurement, the maximum flow rate is 2.61 pcu/s (peak car unit per second). The results from AIMSUN have been reported from 1 hour of the simulation time. For the Q-learning environment, an action decision has

been chosen every 60 seconds. By using the CTM-based Q-learning approach, the algorithm will repeat the learning process as illustrated in Table I for 50 episodes to reach the desired accuracy.

Table II illustrates the nine different sets of traffic arrival where each arrival process is Poisson. The results have been considered into two operation regions, which are the under-saturated and jamming regions, respectively. Note that all nine cases are identical, except for the approaching demand to an intersection and the allocated green time. In fact, the undersaturated traffic conditions occur when the vehicle arrival rate is less than the maximum flow rate. However, if the vehicle arrival rate is greater than the maximum flow rate, then the mathematical solution cannot be solved analytically. The vehicle arrival rates have been varied to produce the offered load ratio varying from 0.2 to 1.2. Although the stability condition is not held, the jamming conditions have been investigated for reporting the applicable range.

Load type	λ_1 pcu/s	λ_2 pcu/s	Offered load ratio
1	0.435	0.087	0.2 μ
2	0.87	0.174	0.4 μ
3	1.305	0.261	0.6 μ
4	1.74	0.348	0.8 μ
5	2.175	0.435	1.0 μ
6	2.61	0.522	1.2 μ
7	3.045	0.609	1.4 μ
8	3.48	0.696	1.6 μ
9	3.915	0.783	1.8 μ

Table II
TYPES OF LOAD

As illustrated in Fig. 3, the results show the allocated green time to each direction for each scenario. In D/D/1 queueing model, the optimal split from (9) is unaffected by the service rate. Therefore, the optimal split from the D/D/1 depends on the proportion of vehicle arrival rates only.

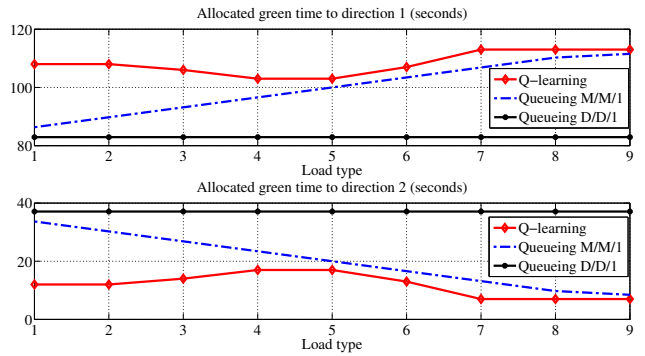


Figure 3. Allocated green time to each direction

Fig. 4 reveals that the improvement of the network throughput in the jamming conditions can be greatly improved by up to 1.7-8.3% from the M/M/1 and can be

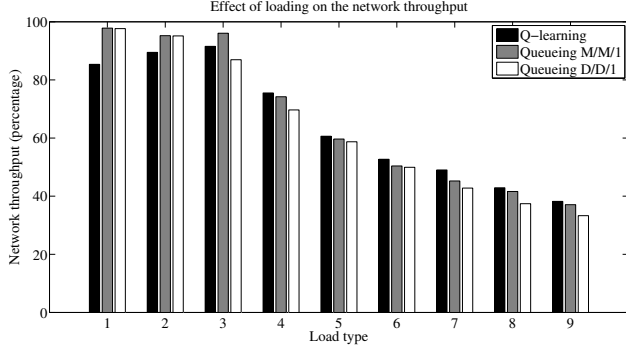


Figure 4. Network throughput comparison among Q-learning, Queueing M/M/1 and Queueing D/D/1

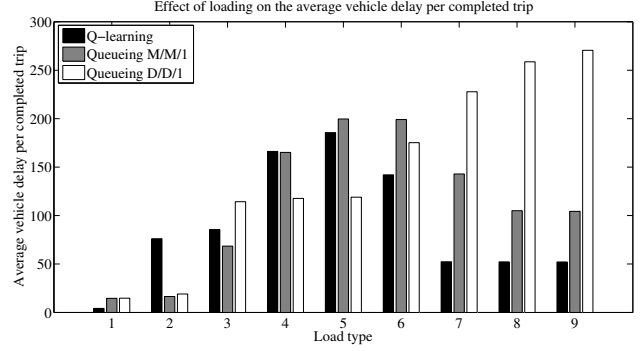


Figure 7. Average vehicle delay per completed trip comparison among Q-learning, Queueing M/M/1 and Queueing D/D/1

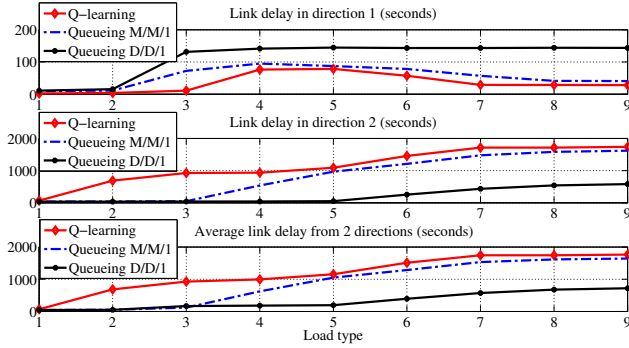


Figure 5. Link delay obtained from AIMSUN

significantly improved up to 3.2-14.8% from the D/D/1. Note that in the undersaturated conditions, the network throughput by both the M/M/1 and the D/D/1 also outperform the proposed CTM-based Q-learning algorithm. Fig. 5 explains why Q-learning performs well and badly in different traffic conditions. The link delay is generally known as the difference between the time spent to travel along a particular road and the free flow travel time along the road. Fig. 5 illustrates the individual link delay for each direction and

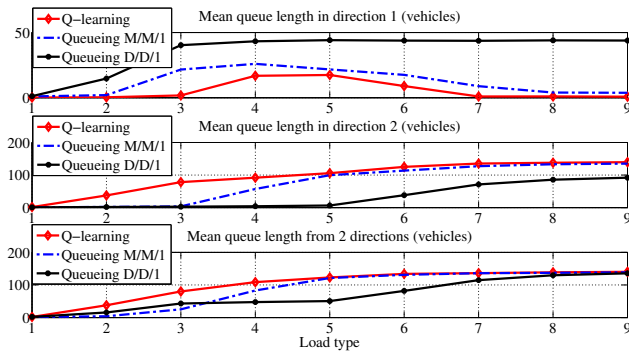


Figure 6. Mean queue length obtained from AIMSUN

the average link delay from two directions. In each cycle time, the Q-learning has been allocated the green time more often to the direction that has higher vehicle arrival rates. However, in both queueing models, the allocated green time in each direction is directly proportional to the incoming traffic demand of its direction. Therefore, by using Q-learning in the undersaturated conditions, the obtained optimal split leads the system to the wasted green scenario. However, the link delay of Q-learning performs well in the jamming conditions because the Q-learning can reduce the link delay from the higher vehicle arrival rates that dominate the overall link delay of the systems. As illustrated in Fig 6, the results for the mean queue length can be used explained with the same discussions as the link delay.

These three approaches share the common goal of minimizing the total network delay. Generally, the total network delay has been calculated from the difference of the time spent to complete a network trip and the free flow travel time along the network path. For each vehicle, the average vehicle delay per completed trip \widetilde{AD} can be calculated by

$$\widetilde{AD} = \frac{\sum_p (ALD_p \times CPT_p)}{\sum_p CPT_p}, \quad (17)$$

where ALD_p is the average link delay in direction p and CPT_p is the number of completed trips in direction p . In Fig 7, for the jamming condition, the reduction of the average vehicle delay per completed trip can be greatly reduced by up to 7.0-63.4% from the M/M/1 and can be significantly reduced up to 18.9-80.7% from the D/D/1.

VI. CONCLUSION AND FUTURE RESEARCH DIRECTIONS

This paper evaluates an optimality analysis based on queueing models and compares with Q-learning to control the traffic signal at an isolated two-phase intersection. The Q-learning approach can improve the intersection throughput by up to 1.7-8.3% and by up to 3.2-14.8% in jamming

condition in comparison with the respective M/M/1 and D/D/1 approaches. Moreover, the average vehicle delay per completed trip can also reduce by up to 7.0-63.4% and by up to 18.9-80.7% in comparison with the respective M/M/1 and D/D/1 approaches.

In this paper, the basic assumptions are based on the M/M/1 model which is rather restricted, thus we are currently investigating other distributions. Furthermore, we are also investigating the scenario when the actual state (i.e. the number of vehicles) is concealed from the agent. In such case, the agent does not have a complete knowledge of the state of the system and must select a traffic signal under such a circumstance. In addition, the existing works related to Q-learning have not considered the scalability issues due to the limitation in terms state space explosion. However, we attempt to alleviate the explosion by employing state space quantisation and control traffic signal in such network scenarios. Methods to find the best possible traffic signal for the road traffic problems in a jamming condition become crucial. Therefore, the extension of the CTM-based Q-learning algorithm and its ability to deal with the jamming conditions will be reported in the forthcoming paper.

ACKNOWLEDGMENT

The authors would like to thank for support received from the Honours Program Scholarship from Electrical Engineering Department of Chulalongkorn University and Thailand Graduate Institute of Science and Technology (TGIST), associated with National Science and Technology Development Agency (NSTDA).

REFERENCES

- [1] Xiao-Hua Yu and Allen R. Stubberud. Markovian decision control for traffic signal systems. *Proceedings of the 36th IEEE Conference on Decision and Control.*, 1997.
- [2] G. F. Newell. The rolling horizon scheme of traffic signal control. *Transportation Research Part A: Policy and Practice.*, 32(1):39-44, 1998.
- [3] P. B. Mirchandani and Z. Ning. Queuing models for analysis of traffic adaptive signal control. *IEEE Transactions on Intelligent Transportation Systems.*, 8(1):50-59, 2007.
- [4] R. Wunderlich, C. Lui, I. Elhanany, and T. Urbanik. A novel signal-scheduling algorithm with quality-of-service provisioning for an isolated intersection. *IEEE Transactions on Intelligent Transportation Systems.*, 9(3):536-547, 2008.
- [5] Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning: An Introduction*. The MIT Press, 1998.
- [6] B. Abdulhai and L. Kattan. Reinforcement learning: Introduction to theory and potential for transport applications. *Canadian Journal of Civil Engineering.*, 30(6):981-991, 2003.
- [7] C. Jacob and B. Abdulhai. Integrated traffic corridor control using machine learning. *IEEE International Conference on Systems, Man and Cybernetics.*, 2005.
- [8] S. Richter. Traffic light scheduling using policy-gradient reinforcement learning. *The International Conference on Automated Planning and Scheduling.*, ICAPS, 2007.
- [9] M. A. Wiering, J. Vreeken, J. Veenen, and A. Koopman. Simulation and optimization of traffic in a city. *IEEE Intelligent Vehicles Symposium.*, 2004.
- [10] Y. Li, R. Wu, and W. Li. The coordination between traffic signal control agents based on q-learning *The 5th World Congress on Intelligent Control and Automation.*, 2004
- [11] S. Lu, X. Liu, and S. Dai. Incremental multistep q-learning for adaptive traffic signal control based on delay minimization strategy. *Proceedings of the World Congress on Intelligent Control and Automation.*, WCICA, 2008.
- [12] W. Kaige, Q. Shiru, and Z. Yumei. A stochastic adaptive control model for isolated intersections. *IEEE International Conference on Robotics and Biomimetics.*, 2008.
- [13] Y.S. Hong, J.S. Kim, J.K. Son, and C.K. Park. Estimation of optimal green time simulation using fuzzy neural network. *IEEE International Fuzzy Systems Conference Proceedings.*, 1999.
- [14] M. C. Choy, D. Srinivasan, and R. L. Cheu. Cooperative, hybrid agent architecture for real-time traffic signal control. *IEEE Transactions on Systems, Man, and Cybernetics Part A: Systems and Humans.*, 33(5):597-607, 2003.
- [15] A. Sadek and N. Basha. Self-learning intelligent agents for dynamic traffic routing on transportation networks. *International Conference on Complex Systems.*, 2006.
- [16] C.H. Shin and K. Choi. Saturation flow rate estimation under rainy weather conditions for on-line traffic control purpose. *ASCE Journal of Civil Engineering.*, 2(3):211-222, 2008.
- [17] L. Kleinrock. *Queuing Systems Volume 1: Theory*, 1975.
- [18] C. F. Daganzo. The cell transmission model part II: Network traffic. *Transportation Research Part B: Methodological.*, 29b(2):79-93, 1995.
- [19] H. K. Lo, E. Chang, and Y. C. Chan. Dynamic network traffic control. *Transportation Research Part A: Policy and Practice.*, 35(8):721-744, 2001.