Agent-Based Optimization for Multiple Signalized Intersections using Q-Learning

Kenneth Tze Kin Teo, Kiam Beng Yeo, Yit Kwong Chin, Helen Sin Ee Chuo, and Min Keng Tan
Modelling, Simulation & Computing Laboratory, Material & Mineral Research Unit
Faculty of Engineering, Universiti Malaysia Sabah
Kota Kinabalu, Malaysia
msclab@ums.edu.my, ktkteo@ieee.org

***Abstract*—Relieving urban traffic congestion has always been an urgent call in a dynamic traffic network. The objective of this research is to control the traffic flow within a traffic network consists of multiple signalized intersections with traffic ramp. The massive traffic network problem is dealt through Q-learning actuated traffic signalization (QLTS), where the traffic phases will be monitored as immediate actions can be taken during congestion to minimize the number of vehicles in queue. QLTS is tested under two cases and has better performance than common fixed-time traffic signalization (FTS). When dealing with the ramp flow, QLTS has flexibility to change the traffic signals according to the traffic conditions and necessity.**

Keywords-Disturbance, Multi-agent, Q-learning, Traffic Signalization, Traffic Flow Optimization.

I. Introduction

Urban area is a large population coverage area with complex traffic network. The traffic network plays a major role in civilization development, as it serves as the communication or travel linkage between places. Relieving traffic predicament from the deteriorating congestion has always been an urgent call to the urban society.

One of the major traffic network dynamicity comes from the progressive inflow from any possible branches, recognizable as a traffic flow disturbance which might hit the traffic bottleneck and cause exceeding road capacity. Failure in coping with the sudden vehicles increment usually ends up in traffic immobility. The congestion effect at a single traffic will be carried forward and affecting the nearby intersections. It is therefore important to look into the congestion in a network rather than a single intersection problem. Flexible traffic signals are required at multiple intersections to react upon the traffic needs when the congestion strikes [1]. The cost of traffic congestion is not limited to the time lost during the travelling, but the possible consequences and the chain effect caused by the congestions.

The basic traffic modeling provides an insight for traffic estimation. On a contrary, precise and robust dynamic traffic modeling is costly and restricted to particular geographical areas and varying parameters. This work suggested agent-based traffic management to control the traffic signals at double intersections. Suiting the frequent changing traffic environment, the agent plays an important role in autonomous data collection and dynamics studies of traffic intersections for future self-judgment of green time duration. The situation whereby disturbance occurs at the intersection is included in this work to depict the network dynamicity and to test the robustness of the developed system towards the changing network conditions. The agent is able to change the traffic signals according to the traffic needs.

II. Overview of Traffic Control Systems

Under the condition of limited land and the difficulty to rebuild the road infrastructure, one of the best solutions to the current traffic scenario is to study and design a heuristic traffic flow controller to enhance traffic smoothness in the congested urban area. Several important parameters in the traffic flow control study are extracted based on the official reports and manuals from Malaysian Public Works Department (JKR) [2], as well as Transportation Research Board's Highway Capacity Manual, as defined below [3]:

- Vehicles in queue: Total number of vehicles that line up in front of the intersection waiting to be given the permission for passing the intersection.
- Traffic phase: Group of traffic lanes given the same traffic signal to avoid conflicts with other traffic lanes.
- Cycle time: Time for the traffic light signals to be circulated once after all the phases have their turn of green signals.
- Lost time: Time lost between the interchange of the signals or the time lost due to the driver's behavior at the intersection, which may increase the number of interchanges between the signals.

The most common traffic management system is the fixed-time traffic signalization (FTS) where duration of each traffic signal is fixed. Due to the increasing vehicles on road, the traffic signalization is improved with predetermined strategic management according to the collected traffic data [4]. The Webster's traffic model [4] has a large influence in the designs of traffic signalization system, as many researchers and department of transportation are still using the model of Webster as the basis of their practical design procedure. JKR published its guidelines in traffic signal designs using the implementation of Webster's study [2]. However, predetermined FTS does not have the ability to react towards the dynamic environment such as unforeseen disturbances and delays in the traffic network. Therefore, heuristic control and adaptation to the dynamic changes of the traffic flow is important to help prevent congestion.

Various artificial intelligence (AI) methods have been attempted in the traffic system due to limiting performance of the existing traffic controllers. In recent years, the common algorithms implemented include fuzzy logic (FL), genetic algorithm (GA) and reinforcement learning (RL).

FL based traffic control and optimization had been researched to slow down the accumulation rate of waiting vehicles at the intersections [5-7]. Azimirad *et al.* (2010) worked out the queue length and the vehicles' waiting time at the intersection as the input parameters of traffic signal timing to optimize the traffic flow using FL embedded control system [5]. However, the simulation was performed at an isolated intersection. The effects of other intersections towards the simulated traffic intersection are not being considered. The FL based traffic management strategy requires precise definition of the membership functions for mapping the input and the output, which causes the shortcoming of not being able to act when the data input is not within the range.

GA based traffic optimizer has been proposed specifically for the purpose of optimizing the oversaturated traffic conditions [8-10]. Sánchez *et al*. (2008) developed microscopic traffic model using cellular automata (CA) where it is optimized by a developed GA [8]. CA served as the fitness evaluation functions, with the chromosomes of GA as the input. GA can learn and evolve from the environment. However, its chromosome representation will become more complicated when the traffic problem is expanded. The number of chromosomes within GA is highly depending on the problem size.

RL is introduced for its ability to learn from the traffic problem environment and build up memory for future decision and traffic optimization [11-13]. Arel *et al*. (2010) developed a RL based traffic optimization control, implemented as a multi-agent system in the traffic environment [11]. The traffic flow is the input and state of the Q-learning (QL) agent in every intersection. QL is one of the variance in RL. Liu and Ma (2007) suggested Dyna-Q algorithm but is not implemented in multi-agent system for the optimization purpose [12]. This work is interested in exploring the potential of QL in a traffic network of non-singular intersections including the situations whereby traffic disturbance exists. The development of Q-learning actuated traffic signalization (QLTS) algorithm will be discussed in the next section.

III. Q-Learning Actuated Traffic Signalization

Intelligent traffic management is able to extend the existing FTS to cope with the overwhelming dynamic traffic flows in the urban traffic network. QL values the effects of actions taken and the resulting state in a traffic environment, which is very much necessary for a learning agent to react appropriately in a fast changing traffic environment [14-15]. QL tends to seek for the best solutions through action that returns the highest reward. For the process of managing the traffic signal, QL has to be defined precisely, in order for the algorithm to be able to interact with the traffic environment and decide the best actions that can counteract the severe traffic congestions. The state-action pairs of the QL are the main part that determines the robustness of the algorithm. Besides the state-action pairs, the reward function is also a crucial function that will decide the policies of the algorithm in its learning process.

In this study, the level of vehicles in queue at each intersection is chosen to be the state of the designed QL. The vehicles in queue are produced from the difference between the traffic inflow and the traffic outflow. Equation (1) is the calculation for the vehicles in queue Q_i , where i is the current intersection and j is the source of the departure traffic. $z^{j,i}$ is the traffic flow rate from intersection j into i while y_i is the average outflow at i . In (2), C is the cycle length in the developed model, $\sum z$ is the total of traffic flow rate into the intersection i , and lastly g_i is the effective green signal duration.

$$Q_i = z^{j,i}(C) - y_i(g_i) \tag{1}$$

$$g_i = \frac{z^{j,i}}{\sum z}(C) \tag{2}$$

To categorize the level of vehicles in queue, the maximum vehicles in queue of the traffic model can be determined using the maximum cycle length as shown in (3).

$$Q_{\max_i} = z^{j,i}(C_{\max}) - y_i\left(\frac{z^{j,i}}{\sum z}C_{\max}\right) \tag{3}$$

Equation (3) is a derivation from (1) where $Q_{\max}$ is the maximum vehicles in queue, and $C_{\max}$ is the maximum cycle length allowed. After $Q_{\max}$ is determined, the level of the vehicles queue in different links can be identified. The definitions of each level are described in Table I, where s_1 represents "no vehicle in queue", s_2 is "less vehicles in queue". Level of "moderate vehicles in queue" is s_3 and s_4 is "maximum vehicles in queue". With each intersection having 4 traffic phases, the total possible states are 256 states combination from 4 phases and 4 levels of vehicles in queue. The illustration of the states' combinations is revealed in Figure 1.

The learning attributes of QL is based on the reward function. Besides rewarding each proper action, penalties can be given to those unproductive actions. The goal of the proposed QL algorithm is to yield the minimum possible vehicles in queue for each intersection. Thus, reward function is computed carefully for each appropriate green signal distribution. Meanwhile, the action that yields excessive idle green signal will be penalized. Excessive idle green signal is a period when there are no more waiting vehicles at the intersection, but the green signal is still being allocated and triggered for that period. The cycle length available is always limited and it constrains the distribution of green signal within a cycle length. Thus, when certain traffic phase is getting longer green signal than it should be, other traffic phases has to share the common loss from the decisions. Table II shows the pseudo codes for the algorithm of reward function in the QL.

Simulation tests are computed to verify the function and to test the performance of the QLTS. Various traffic flows have been set to simulate different traffic scenarios. Those scenarios include under-saturated traffic condition, a high traffic flow phase in the traffic intersections, and oversaturated traffic intersections. The saturation flow, S in this research is measured in vehicles per hour (veh/h). The width of the traffic lane in this study is designed to be 5 m, hence the saturation flow for the traffic model is 2560 veh/h. In order to identify or classify the level of traffic flow condition, (4) is used [3].

TABLE I. FOUR STATE DEFINITIONS AND CONDITIONS

State	*Condition*
s_1	$Q = 0$
s_2	$0 < Q < 0.5Q_{\max}$
s_3	$0.5Q_{\max} < Q < 0.9Q_{\max}$
s_4	$Q > 0.9Q_{\max}$

$$
\text{States list}\left\{
\begin{array}{lcllll}
 & & \multicolumn{4}{c}{\overbrace{\hspace{6em}}^{\text{Traffic Phase}}} \\
 & & i & j & k & l \\
 & & \downarrow & \downarrow & \downarrow & \downarrow \\
S1 & \leftarrow & s_1 & s_1 & s_1 & s_1 \\
S2 & \leftarrow & s_1 & s_1 & s_1 & s_2 \\
S3 & \leftarrow & s_1 & s_1 & s_1 & s_3 \\
S4 & \leftarrow & s_1 & s_1 & s_1 & s_4 \\
 & & & \vdots & & \\
Sn & \leftarrow & s_a & s_b & s_c & s_d \\
 & & & \vdots & & \\
S256 & \leftarrow & s_4 & s_4 & s_4 & s_4
\end{array}\right.
$$

Figure 1. Q-learning states' combinations.

TABLE II. PROPOSED REWARD FUNCTION ALGORITHM

Reward Function

1. Identify current green signal by actions
2. Identify current level of vehicles in queue
3. **If** (*vehicles in queue = 0 & green signal > 0*)
4. # Excessive green signal condition
5. *Reward* = −1 (penalty)
6. **Else** (vehicles in queue > 0 & green signal > 0)
7. # Green signal for current vehicles in queue
8. *Reward* = 1
9. **End**
10. **Return** *Reward*

$$x_i = \frac{f_i}{S\left(1 - \dfrac{L}{C}\right)} \tag{4}$$

The term x_i is the flow rate to capacity ratio which can assist to determine the level of traffic condition. f_i is the actual flow rate into the intersection i, L is the total lost time and C is the cycle length of the intersection. S is the saturation flow. Following the guidelines of JKR's manual, L is set to be 16 s and C is 120 s. Thus, by substituting the known variable into (4), the equation can be further simplified into (5).

$$x_i = \frac{f_i}{2218.67} \tag{5}$$

Table III shows the relationship between the x_i ratio with the level of traffic flow capacity at the intersection. Traffic intersections will experience traffic congestion or oversaturated traffic situation when the traffic flow to capacity ratio, X, is over 1, where the traffic light system is unable to handle the overwhelming traffic flow. Table III is used to determine the traffic condition of the traffic flow input into the intersection during the simulations [3].

IV. SIMULATION

In a traffic network of multiple intersections, the environment for each traffic intersection is not the same. Each intersection consists of major and minor traffic phases. The network topology in Figure 2 is built based on the two-intersection traffic model. An extra ramp is added in between the intersection A (INT A) and intersection B (INT B) for the purpose of introducing extra traffic flow into INT A as a disturbance. The distance of INT A and INT B is 5 km whereas the ramp between the two intersections is distanced 2.5 km from INT A and INT B respectively.

TABLE III. RATIO (X) WITH RESPECT TO INTERSECTION CAPACITY

Ratio (X)	*Capacity*
$X \le 0.85$	Under capacity
$0.85 < X \le 0.95$	Near capacity
$0.95 < X \le 1.00$	At capacity
$X \ge 1.00$	Over capacity

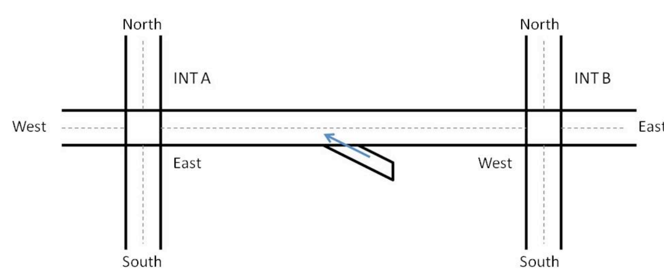

Figure 2. Traffic network topology for two intersections with a ramp.

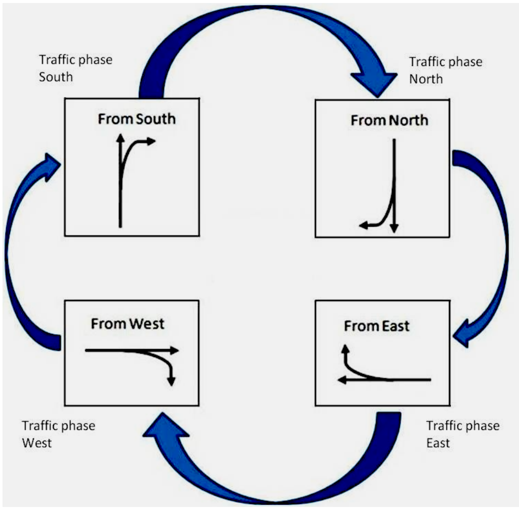

Figure 3. Traffic phase transitions in a cycle length.

TABLE IV. TRAFFIC PARAMETERS FOR NETWORK TOPOLOGY

Parameters	***CS1***	***CS2***
Distant between INT A and B	5 km	
Distant between INT A and ramp	2.5 km	
Traffic Flow Input	***(veh/h)***	
INT A Phase South	72	72
INT A Phase North	72	72
INT A Phase East	-	-
INT A Phase West	720	720
INT B Phase South	72	72
INT B Phase North	72	72
INT B Phase East	1152	1080
INT B Phase West	-	-
From ramp into INT A (after 900s)	-	900

This setup is for simulating the situations of fast arrival traffic flows. The traffic phase assignments to both intersections are the same. The first traffic phase is the south phase, followed by the north phase as the second phase, west phase as the third phase and the east phase as the last. The transitions of the traffic phases in the traffic network are illustrated in Figure 3.

There are two conditions at intersection tested in this work. Condition 1 (CS1) is tested with under-saturated INT A and saturated INT B. In the saturated situation, the aim is to test the reaction of QLTS towards the effects of the high traffic condition of INT B. The details of the simulation setup for each simulation case are shown in Table IV. In Condition 2 (CS2), the ramp will release its traffic flow into INT A. This will simulate the situation where a sudden traffic flow entering the traffic network between two intersections, which is the introduction of a traffic flow disturbance by the ramp from 900 s until 1500 s of simulation time. The ramp input is set as 900 veh/h for 600 s in the simulation to test the ability of both systems.

V. RESULTS AND DISCUSSION

The simulation results of the topology in CS1 conducted with FTS and QLTS are shown in Figure 4. For CS1, FTS has equal period of repeating pattern at INT A. The traffic phase East of INT B has increasing accumulation of vehicles in queue throughout the simulation, which affecting the consecutive INT A in the traffic network. Due to the high traffic inflow at INT B, most of the traffic flows pass through INT B will start to accumulate at the East traffic phase of INT A, referring to Figure 4. The traffic flow is irreversible for the FTS management as the traffic signals for the intersection is fixed. For CS1 QLTS, the vehicles in queue of the traffic phase East at INT B has been successfully reduced throughout the simulation time. The outgoing traffic flow from the traffic phase East at INT B is directed to its following traffic phase at the next INT A. The increasing traffic flow does not have much effect on the INT A and INT B because the green signal duration is flexible and wisely distributed over the intersections based on the traffic environmental learning of QLTS at both INT A and INT B. Suitable green signal durations is allocated based on the traffic conditions and needs. When the QLTS agents sense the heavy traffic flow accumulating at phase(s), they react towards the situation by compromising the other traffic phases which has comparatively less vehicles in queue. The compromising effect of the QLTS are reflected at the South and North phases of both intersections, giving approximately even distribution of vehicle at each traffic phase.

The simulation results of topology in CS2 conducted with FTS and QLTS are depicted in Figure 5. At the East traffic phase of INT A, extra traffic flows from the ramp is flowing into the traffic phases, resulting in congestion for INT A Phase East with the total 318 vehicles in queue at the end of the simulation, as shown in Figure 5. Vehicles in queue are not relievable and exceeding over 100 vehicles as the traffic flow from the ramp accumulates at the intersection at around 1200 s. The FTS management system is unable to react to the congestion with constant green signal duration allocations. INT A suffered from severe traffic jam and caused failure in traffic management. On the contrary, the performance of QLTS for the traffic disturbance shows no extreme level of vehicles in queue at each phase throughout the complete run time. The vehicles in queue for INT A and INT B are stable and well distributed under traffic disturbance. QLTS agents at both intersections are giving the appropriate green signals duration to their intersections and manage to keep the level of vehicles in queue at a minimum level. This also reflecting in relatively considerable increasing queuing vehicles (longer red signal duration) in INT A traffic phase South, North and West to compensate for the extra time needed to relieve the traffic congestion at INT A Phase East. The phenomenon of compromising the INT A traffic phases except Phase East does not last long, as the vehicles in queue of all the INT A traffic phases start to reduce after the traffic flow from the ramp has been relieved. The simulation data of average vehicles passing through each intersection during the green signals for CS1 and CS2 are tabulated in Table V and VI.

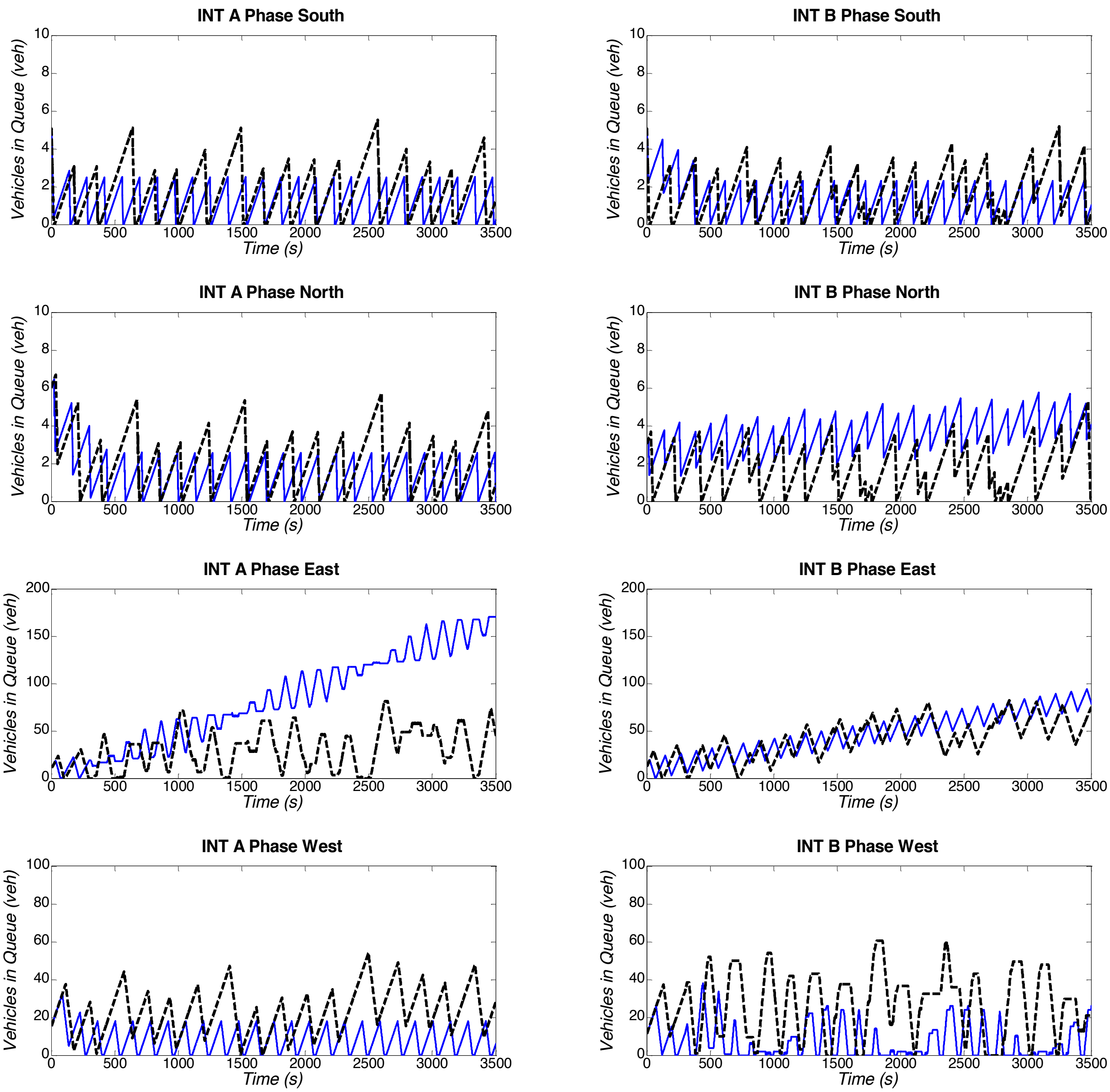

Figure 4. Simulation results of FTS and QLTS agent system at different intersection phases for CS1. [—— FTS; ---- QLTSTP]

TABLE V. SIMULATION DATA FOR TRAFFIC CONDITION 1

CS1		FTS		QLTS	
		INT A	*INT B*	*INT A*	*INT B*
Average vehicles relieved during green signal (veh)	Phase South	55	64	62	62
	Phase North	58	71	66	62
	Phase East	960	1070	1085	1120
	Phase West	643	632	707	736
	TOTAL	1716	1837	1920	1980
No. of cycles		26	30	18	26

TABLE VI. SIMULATION DATA FOR TRAFFIC CONDITION 2

CS2		FTS		QLTS	
		INT A	*INT B*	*INT A*	*INT B*
Average vehicles relieved during green signal (veh)	Phase South	55	64	62	55
	Phase North	58	71	66	57
	Phase East	960	1070	1210	1035
	Phase West	643	632	704	744
	TOTAL	1716	1837	2042	1891
No. of cycles		26	30	18	27

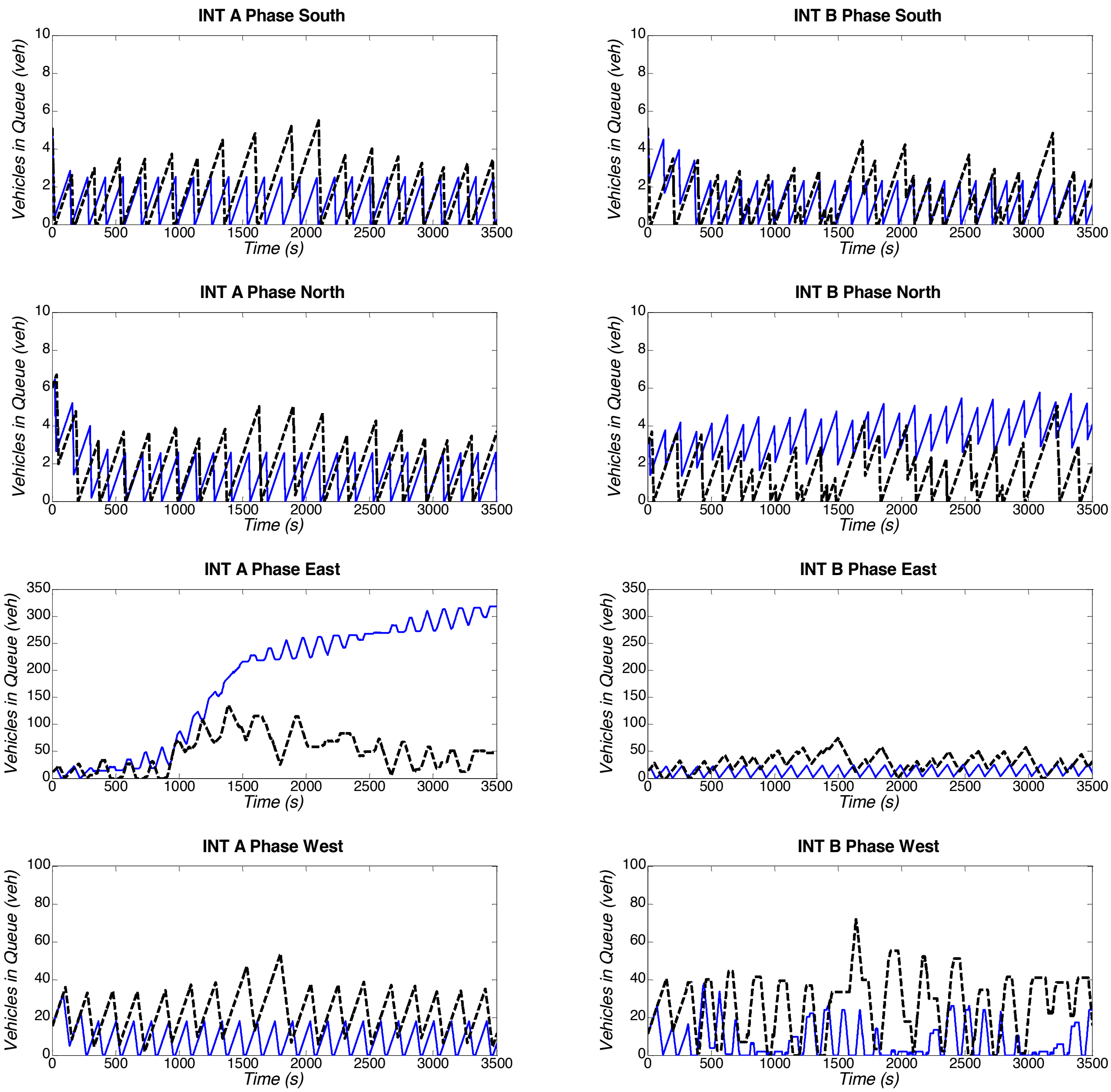

Figure 5. Simulation results of FTS and QLTS agent system at different intersection phases for CS2. [—— FTS; ---- QLTSTP]

Comparing FTS and QLTS, the number of traffic signal cycles is lesser using QLTS for both CS1 and CS2, indicating lost time is reduced at the intersections. The duration of the effective green time of QLTS has increased and more vehicles are relieved even under ramp disturbance. As a result, QLTS system has 11.9 % and 7.7 % more vehicles passing through INT A and INT B respectively in CS1, tabulated in Table VII. As for CS2, 19.0 % more vehicles passed through INT A while INT B showed 2.9 % improvement.

TABLE VII. PERFORMANCE COMPARISON FOR FTS AND QLTS

Cases	Improvement of Vehicles Pass Through (QLTS against FTS)	
	INT A	*INT B*
CS1	11.89 %	7.70 %
CS2	18.99 %	2.90 %

VI. Conclusion

The vehicles queuing condition at traffic intersections can be improved by QLTS. Due to the learning behavior, QLTS changes the green signal duration according to the traffic flow environment at intersections particularly during congestion and traffic flow disturbance. Summarizing both cases studies, the overall improvement of vehicles passed through the intersections using QLTS lies between 2.9 % – 19.0 % as compared to the FTS management.

The investigation of weight distribution between exploration and exploitation could be explored in future study. Under normal traffic condition, QL acts more dominantly by exploitation where the past experience and historical data would be referred. Under conditions such as non-recurrent traffic congestion which affects the normal smooth traffic flow, the agent will act more depending on exploration to refine the workable range of parameters and decide suitable actions for coping the new situations.

In a nutshell, the agent-based traffic management is potentially contributable to the integrated solutions of heterogeneous intelligent traffic management system as it is adaptable, autonomous and decisive in dynamic changing environment and able to work together with the emerged communication technology in informational transportation.

Acknowledgment

The authors would like to acknowledge the Ministry of Education Malaysia (KPM) for supporting this research under Exploratory Research Grant Scheme (ERGS) No. UMS/ERG0021-TK-1/2012 and UMS/ERG0047-ICT-1/2013 respectively.

References

[1] Y.K. Chin, K.C. Yong, N. Bolong, S.S. Yang, K.T.K. Teo. “Multiple intersections traffic signal timing optimization with genetic algorithm,” Proc. International Conference on Control System, Computing and Engineering (ICCSCE 2011), 2011. pp. 454-459. doi: 10.1109/ICCSCE.2011.6190569

[2] Jabatan kerja Raya, A Guide to the Design of Traffic Signal, 1987.

[3] Transportation Research Board, Highway Capacity Manual. Washington DC 113: National Reseach Council, 2000.

[4] F.V. Webster, “Traffic signal settings,” Road Research Technical Paper No. 39. UK: Road Research Laboratory, 1958.

[5] E. Azimirad, N. Pariz, and M.B.N. Sistani, “A novel fuzzy model and control of single intersection at urban traffic network,” IEEE Systems Journal, vol. 4, issue 1, 2010, pp. 107-111. doi: 10.1109/JSYST.2010.2043159.

[6] P. Jain and M. Sethi, “Fuzzy based real time traffic signal controller to oprimize congestion delays,” Proc. International Conference of Advanced Computing & Communication Technologies (ACCT), 2012, pp. 204-207. doi: 10.1109/ACCT.2012.55.

[7] H.M. Kammoun, I. Kallel, J. Casillas, A. Abraham, and A.M. Alimi, “*Adapt-Traf*. An adaptive multiagent road traffic management system based on hybrid ant-hierarchical fuzzy model,” Transportation Research Part C: Emerging Technologies, vol. 42, 2014, pp. 147-167. doi: 10.1016/j.trc.2014.03.003.

[8] J. Sánchez, M. Galan, and R. Enrique, “Applying a traffic lights evolutionary optimization technique to a real case: ‘Las Ramblas’ area in Santa Cruz de Tenerife,” IEEE Transactions on Evolutionary Computation, vol. 12, issue 1, pp. 25-40, 2008. doi: 10.1109/TEVC.2007.892765.

[9] K.T.K. Teo, W.Y. Kow, Y.K. Chin, “Optimization of traffic flow within an urban traffic light intersection with genetic algorithm,” Proc. 2nd International Conference on Computational Intelligence, Modelling and Simulation (CIMSiM 2010), 2010, pp. 172-177. doi: 10.1109/CIMSiM.2010.95.

[10] H. Dezani, R.D.S. Bassi, N. Marrangehello, L. Gomes, F. Damiani, and I.N.D. Silva, “Optimizing urban traffic flow using genetic algorithm with petri net analysis as fitness function,” vol. 124, 2014, pp. 162-167. doi: 10.1016/j.neucom.2013.07.015.

[11] I. Arel, C. Liu, T. Urbanik, and A.G. Kohls, “Reinforcement learning-based multi-agent system for network traffic signal control,” IET Intelligent Transport Systems, vol 4, issue 2, pp. 128-135, 2010. doi: 10.1049/iet-its.2009.0070.

[12] Z.Y. Liu and F.W. Ma, “On-line reinforcement learning control for urban traffic signals,” Control Conference (CCS), pp. 34-37, 2007. doi: 10.1109/CHICC.2006.4347023.

[13] M.A. Khamis and W.Gomaa, “Adaptive multi-objective reinforcement learning with hybrid exploration for traffic signal control based on cooperative multi-agent framework,” Engineering Applications of Artificial Intelligence, vol. 29, 2014, pp. 134-151. doi: 10.1016/j.engappai.2014.01.007.

[14] Y.K. Chin, N. Bolong, A. Kiring, S.S. Yang, and K.T.K. Teo, “Q-learning based traffic optimization in management of signal timing plan,” International Journal of Simulation, Systems, Science & Technology (IJSSST), vol. 12, issue 3, pp. 29-35, 2011. doi: 10.5013/IJSSST.a.12.03.05.

[15] J.C. Medina, MASTraf: A Decentralized Multi-Agent System for Network-Wide Traffic Signal Control with Dynamic Coordination. PhD Dissertation. 2013. University of Illinois.