ISSN: 2229-6956(ONLINE)
DOI: 10.21917/ijsc.2013.0073

OPTIMIZATION OF URBAN MULTI-INTERSECTION TRAFFIC FLOW VIA Q-LEARNING

Yit Kwong Chin[1], Heng Jin Tham[1], N.S.V. Kameswara Rao[1], Nurmin Bolong[1] and Kenneth Tze Kin Teo[2]
Modeling, Simulation & Computing Laboratory, Material & Mineral Research Unit
School of Engineering and Information Technology, Universiti Malaysia Sabah, Malaysia
E-mail: [1]msclab@ums.edu.my and [2]ktkteo@ieee.org

Abstract
Congestions of the traffic flow within the urban traffic network have been a challenging task for all the urban developers. Many approaches have been introduced into the current system to solve the traffic congestion problems. Reconfiguration of the traffic signal timing plan has been carried out through implementation of different techniques. However, dynamic characteristics of the traffic flow increase the difficulties towards the ultimate solutions. Thus, traffic congestions still remain as unsolvable problems to the current traffic control system. In this study, artificial intelligence method has been introduced in the traffic light system to alter the traffic signal timing plan to optimize the traffic flows. Q-learning algorithm in this study has enhanced the traffic light system with learning ability. The learning mechanism of Q-learning enables traffic light intersections to release itself from traffic congestions situation. Adjacent traffic light intersections will work independently and yet cooperate with each others to a common goal of ensuring the fluency of the traffic flows within the traffic network. The simulated results show that the Q-Learning algorithm is able to learn from the dynamic traffic flow and optimize the traffic flow accordingly.

Keywords:
Reinforcement Learning, Q-Learning, Traffic Networks, Traffic Signal Timing Plan Management, Multi-Agents Systems

1. INTRODUCTION

Traffic network with multiple intersections are developed to support the traffic demands within the fully developed urban areas. Transportations within the urban areas are done by using the traffic network. Different kinds of traffic infrastructures have been designed and constructed to ensure fluency of all the travelling pedestrians and on-road vehicles within the traffic networks. Unfortunately, within the highly populated urban area, traffic demands towards the traffic networks are extremely high, as the daily operations of the community require high mobility. Due to the traffic demands rise, the existing traffic networks start to encounter problems like traffic congestions when the traffic infrastructures failed to restrain the saturated traffic flows. The limited landscapes of the urban area have constrained the possibilities of rebuilding the existing traffic networks with infrastructures like fly-over ramps and roundabouts. Thus, improvement of the traffic flow management within the traffic network depends on the traffic light systems.

Traffic light systems unable to maintain their performance in traffic network optimization when the traffic demands from the on-road vehicles increased. The reason of the underperformance of traffic light system is the dynamic characteristic of the traffic flows within the networks. Statistic traffic information which is used by the conventional traffic signal timing plan management lacks the ability to rapidly adapt into the dynamic traffic flows. Thus, continuous learning from the dynamic traffic environment for adaptability is necessary in the development of intelligent traffic signal timing plan management.

Q-Learning algorithm gathers information from its learning process as its experience and learns from the environments. This learning ability is emphasized in this study of traffic flow optimization system to act as multi-agent systems. The Q-learning algorithm implemented at traffic light intersections will be able to learn from the traffic environment for increasing its adaptability and making a better decision in the future.

2. REVIEWS OF TRAFFIC CONTROL

In the studies of traffic signal timing plan management, various approaches had been done by researchers to increase the intelligent of the traffic management system. The basics of the traffic lights system has to be studied thoroughly before the implementation of advance approaches.

In the traffic signal timing plan, traffic phases has to be coordinated by the traffic light system, for the prevention of crashes between traffic phases and ensure the smoothness of the traffic flow within the traffic network. There will always be only one traffic phase which is activated at a single time. Traffic signals which are accepted by the world are red, amber and green signals. Red signal represents the restrictions to pass through the intersections as red lights have the lowest frequencies in the light spectrum among the three signals. This enables it to travel the furthest to warn the arriving vehicles. Amber signal is a warning for vehicles to slow down for stoppage at the intersections, and it always acts as a cushion period between red and green signals to prevent sudden stop for the vehicles. Green signalized a “go” permission for vehicles to pass through the intersections. Every traffic light systems have to complete a sequence of green, amber and red signals for each traffic phase. And red signals will always be the one having the longest period, follows by green signals and amber will be the signal with the shortest period. When the traffic light systems circulate through all the traffic phases, it is called as a complete cycle for the traffic light intersections [1].

The most conventional traffic signal timing plan is the fixed-time traffic signal plan. In the fixed-time traffic signal plan management, all the duration of the traffic signals are preset in the database and being executed repeatedly. The configurations of the traffic signals durations are done based on the historical traffic statistic gathered for a long period. The setting of the fixed time traffic signal plan will not be changed until the next review on the statistic traffic information. This method is found to be flawed after the increasing traffic demands within the traffic networks start emerge as the method is unable to adapt itself into the dynamic traffic environment. Therefore, researches

of developing traffic flow management systems with artificial intelligence has been carried out.

Artificial intelligence techniques with the ability to learn have been proposed in the development of the intelligent traffic light systems history. Artificial neural networks technique is one of the methods that being introduced by researchers into the traffic flow management systems. Extension of green signal duration in a traffic phase is decided by the neural networks algorithm in various traffic situations and the algorithm will continue to learn itself from the environment through reinforcement learning [2]. However, the neural network method for the traffic signal plan has to be trained for a large amount of data in order for the algorithm to perform in the dynamic traffic environment. Other researchers also interested in establishing communication elements within the traffic network. Communications between vehicles and vehicles to infrastructure are also being proposed as a solution to the traffic congestion [3]. In such research, problems met by the researchers are the secure connections between vehicles, as the connection established is considered as a mobile wireless network. The moving vehicles will cause difficulties to ensure the reliability of the traffic data transferred. Another research is also conducted in implementing intelligence approach, but using another advance control methods which is fuzzy logic controller. Fuzzy logic is able to interpret the linguistic values into numerical values, thus being used to interpret the traffic congestion situations which are hard to be clearly classified. With the aid of fuzzy logic, the traffic management system is able to extend the green signal duration for the continuous incoming traffic flow [4]. There is also research in microscopic traffic control involving only single intersection using fuzzy logic algorithm to comprehend the traffic situation [5]. Fuzzy logic has the advantage of visualizing situation with indistinguishable boundary. In the environment of traffic networks, traffic conditions need to be defined before further control or process can be done, but there are no clear cut boundaries between high traffic flows and low traffic flows, so by using fuzzy logic, traffic conditions input can be classified for better optimization process. However, fuzzy logic still lacks the ability to learn from the situations.

Introduction of evolvable techniques such as genetic algorithm are also being done by some other researchers in the traffic management system. Genetic algorithm or evolutionary algorithm is an algorithm that mimics the characteristics of survival in the nature law. The basic concept is the fittest individual among the populations will survive until the end through the evolutions of the populations throughout the generations. The fittest individual will be treated as the most optimum solution for the designed problem. In traffic management system, populations of genetic algorithm are defined with different configurations of chromosomes. The most surviving chromosomes in the populations are the most optimum green signal durations [6]-[7]. In another research, genetic algorithm is use to re-routing the traffic flow within the traffic network to avoid traffic congestions [8]. The drawback of genetic algorithm is the computation power needed by the algorithm, as the nature of the algorithm needs to run through all the evolution generations to search for the optimum or the best solutions. So the algorithm might need a strong computation device to keep up with the dynamic changes of the traffic network flow.

Advantages of reinforcement learning have been shown with the ability of exploring the environment to exploit the most suitable actions in the dynamic situations in various studies and researches. Researches involving reinforcement learning in the traffic signal management systems have shown significant results and thus have drawn more attentions to traffic management systems' researchers. Reinforcement learning is used to manage the traffic flow within the networks with the longest-queue-first algorithm [9]. Q-Learning as one of the reinforcement learning algorithm is applied in the optimization of the traffic flow at single traffic intersection [10]-[11]. Besides that, Q-learning algorithm has also been highly valued in the researches of traffic control system as multi-agents systems [12]-[13]. In this study, Q-learning algorithm has been proposed to be studied in the traffic signal timing plan management of traffic networks.

3. Q-LEARNING ALGORITHM

Reinforcement learning is an algorithm that can improve and evolve itself from the past experiences. Assessment of the decisions made from the past is done and stored as experiences data which act as the experience references in the future. Reinforcement learning is usually presented as Q-Learning algorithm which values not only the actions taken but also the states caused by the actions. This feature has made this study to believe in Q-Learning to have the prospective ability to be implemented into the traffic flow control within the traffic networks.

Q-Learning's concepts can be analogues by many phenomena in the real life of nature. One of the most common metaphor used to demonstrate the algorithm is the relationship between a trainer and trainee, for example, a teacher and student, a dog trainer and his dog, or parents and their child. By taking the process of training session between a dog and its trainer as an example; the trainer will evaluate each actions of the dog after the commands are given. The dog will respond to the command when it is given to the dog, and then the trainer will observe and evaluate the performance of the dog. If the dog's action is within the expectations, a reward in the form of food or snacks will be given to the dog; and nothing will be given if the dog did not behave accordingly. In this way, the dog will make a connection between the commands and the actions; it will realize that only the correct command and action pairs will be rewarded. All these experiences encourage the dog to respond according to the command of the trainer, as the dog desires rewards from its trainer. The dog will able to learn every trick that is taught by the trainer after the processes are repeated in times, as it know which command and action pairs will lead it to a reward and which is not. In simple words, Q-Learning algorithm is a process of deciding actions towards the environment and receiving rewards according to the action taken.

3.1 STRUCTURE OF Q-LEARNING

The flow chart of Q-learning is being illustrated in Fig.1. The process of Q-learning starts with the initialization of the states and actions in the Q-table. After the initialization, Q-learning will identify its current state in the environment. An action will be chosen from the action lists available by searching for the

maximum possible rewards returned by the action. Then, the actions chosen will be executed or evaluated. The rewards gained from the actions chosen will be updated in the Q-table. After the actions have been executed, Q-learning will identify the next states in the environment model. Finally, Q-learning will check for the goal accomplishment, the process will restart from the beginning if the goal did not accomplish.

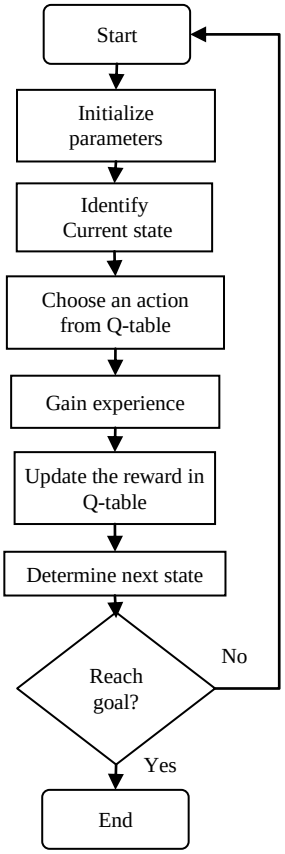

Fig.1. Q-Learning algorithm flow chart

As stated in the previous section, there are many ways to describe about the proposed algorithm's learning process. In Q-learning algorithm, the role of the trainer is played by the environment model, while the Q-learning algorithm itself learns from the environment. Thus, in the development of a Q-learning algorithm, each operation involves in the process must be determined carefully. Q-Learning algorithm is composed with several steps or operations. The implementation of Q-learning operation is applied through the evaluation of Eq.(1).

$$Q(s,a)_i = (1-\alpha)Q(s,a)_{i-1} + \alpha[R(s,a)_i + \gamma \max_{a'} Q(s',a')] \quad (1)$$

where,

S = current state
α = action taken in current state
s' = next state
a' = action taken in next state
i = iteration
α = learning rate
γ = discounting factor

Q-Learning is evaluated from Eq.(1), each of the evaluated Q-value is the rewards gained from the experiences in the exploration process. Q-table is the memory of the Q-learning algorithm, storing every single state and action pairs along with their rewards.

α, learning rate is an important variable in evaluation of the Q-value. α is the factor that will influence the learning rate of the Q-learning algorithm. Learning rate of Q-learning is ranged from 0 to 1, and responsible for the weight of the newly learnt experience. When the learning rate is equal to 1, the Q-learning algorithm will toss away its old experience and treat the newly learnt action as its only experience. This will set the Q-learning to be opportunist which only cares for the immediate rewards. If the learning rate is set to be too low or near to 0, then the Q-learning will suffer from the slow learning rate. For zero learning rates, the experience newly gained is not relevant, and Q-learning will be in the exploitation mode which only acts based on its past experience, and caused the Q-learning algorithm to stop exploring in the environment.

Discounting factor γ is the variable that decides the importance of the future states. High discounting factor will make Q-learning algorithm to be too speculative, where it will focus more on the possible future rewards and neglect the importance of the current experience. The advantage of having a high discounting factor is enabling the Q-learning algorithm converges in a faster rate. Optimum value of the discounting factor is important to let Q-learning algorithm having its speculative characteristic towards the future for long term rewards, and still focus on the short term rewards from the current experience.

In the processing of choosing the actions from the actions list, Q-learning will search for the actions with the maximum rewards benefits. But, there will always be cases where the actions with current maximum rewards are not producing the real highest rewards of return. There is a mechanism of Q-learning acting as a support protocol to minimize the possibility of actions being trapped in a local maximum. It is called the ε-greedy selection which is triggered by a greedy probability.

Greedy probability, ε allows the Q-learning to have a chance of choosing an action from the actions lists at random which does not return the highest rewards [14]. Greedy probability, ε provides Q-learning to be able to continuously explore itself in the new environment for other possibilities of actions despite of the current highest rewards. However, if the greedy probability is too high, Q-learning will face the difficulty of converging, as the greedy probability will prompt the Q-learning to continue exploring in the environment.

3.2 STATE-ACTION PAIRS

Q-Learning algorithm gains its experience through the exploration in the modelled environment. An accurately defined environment will ease the Q-learning's exploration process. The environment of the Q-learning is formed by the states and actions [15]. Each state which is available in the state space represents the boundary of the environment's map. If the definition of the states is wrongly done, then the Q-learning algorithm might lose itself in the process of exploration or learning.

In this study of traffic signal timing plan management, the states of the designed Q-learning are the level of vehicles in queue at each intersection. There are 4 levels of vehicles in queue defined for this study, where they are categorized from no vehicles in queue to high vehicles in queue. The levels of vehicle in queue are defined through the studies of the capacity ratio of the traffic roads with the traffic flows [16]-[17]. With each intersection having 4 traffic phases, the total possible states are 256 states combination from the permutation of 4 phases and 4 levels of vehicles in queue. Fig.2 illustrates the combination of each state of the Q-learning based traffic signal optimizer in this study. It can be seen that there are 4 traffic phases involved in the definitions of states, where *i*, *j*, *k*, and *l* are the 4 different traffic phases available respectively.

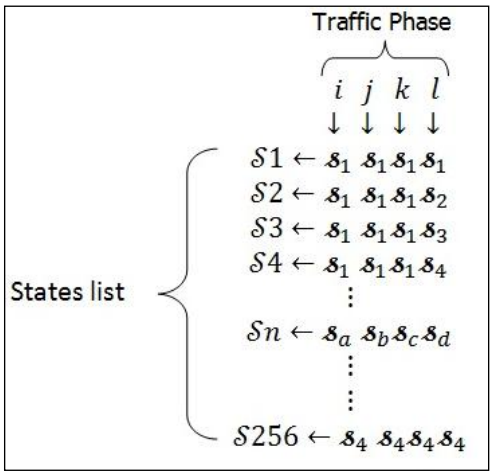

Fig.2. States Definition of Q-learning

Action list of Q-Learning consists of every possible actions or execution command of the agent in different states of Q-learning. It is a very important parameter, as the actions act as the navigators for the Q-Learning algorithm in the environment. Each action done by the Q-Learning will lead to another state. If the state and action pairs are not set correctly, then the whole Q-learning will not be able to achieve optimum solution within the process.

The levels of vehicles in queue at the intersections will be increased by the incoming traffic flow and decreased when they pass through the intersection during the green signal periods. Therefore, green signals are defined as the actions of the proposed Q-Learning algorithm. The actions available in the actions lists are 1 second and 5 seconds of green signals distribution. In the process of exploration, the chosen action will be stored in the memory of traffic signal timing plan management system until the goal of the Q-Learning is achieved. After that, the green signals will be allocated to each traffic phase in the intersection.

3.3 REWARDS AND PENALTIES FUNCTIONS

Although states and actions pairs of the Q-learning algorithm have been decided, rewards and penalties for each chosen actions have to be determined to ensure the Q-Learning algorithm is performing well. The basic rule of Q-Learning is the highest rewards will be given towards the best actions while the worst actions will be assigned with the least rewards. Besides rewarding each proper action, penalties can be given to those unproductive actions. The purpose of introducing penalties in Q-learning is to prevent the algorithm from bias during the rewarding of the actions.

The aim of the proposed Q-learning algorithm is to yield the minimum possible vehicle in queue for each intersection. Thus, rewards functions are computed carefully for each appropriate green signal distributions and the actions that yield excessive idle green signals will be penalized. Excessive idle green signals is a period when there are no more waiting vehicles at the intersection, but the green signal is still being allocated and being triggered for that period. Actions that allocate excessive idle green time will be penalized. The proposed Q-learning algorithm will stop after it reaches the goal of the system, which is every single traffic link has been allocated with their corresponding traffic signal durations. All of the rewards and penalties returned by the reward and penalties function are stored in the memory of Q-learning as their own experience for their future references.

4. SIMULATIONS

The study of this paper focused on the implementation of Q-learning algorithm in the traffic signal timing management within the traffic network. A traffic network consists of two traffic light intersections has been developed for the simulation of this paper study. The illustration of the traffic network is shown in Fig.3; intersection A (INTA) is located at the west and intersection B (INTB) at east. The relationship between INTA and INTB is the main focus of this study.

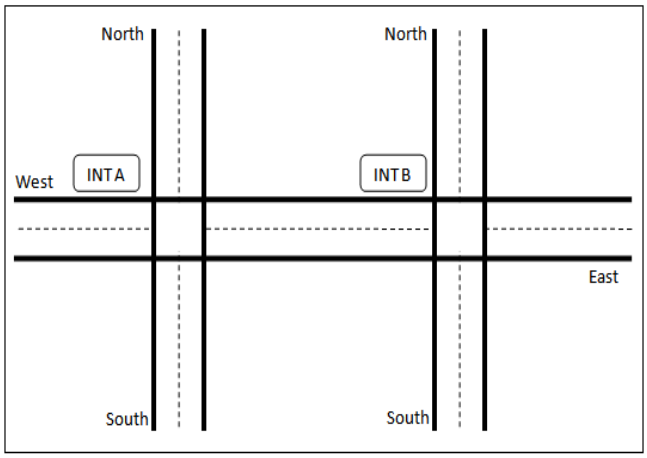

Fig.3. Traffic Network of 2 Intersections

The two intersections in this traffic network consists the configurations of 4 traffic phases as shown in Fig.4. Traffic phases are the sequences of the traffic signals activation and only one traffic phase maybe undergo the green signal at a particular time [1]. This is to avoid traffic crashes from different directions at the intersections.

Graphs of simulation results in Fig.5 are the simulation results of the Q-learning based traffic signal optimizer (QLTSO) at Intersection A for 250 seconds. The 4 graphs in Fig.5 represented the 4 different traffic phases at Intersection A. The bar graphs of each graph shows the activation duration of the green signals for each traffic phase. The graphs show that no two

traffic phases' green signals are activated simultaneously throughout the whole simulation period, but they are triggered one after another. There are gaps between the active green signals showing the red and amber signals for the traffic phases. The increasing part of the line graphs indicate the accumulation of vehicles in queue during the simulation, while the decline slopes are representing the releasing of the vehicles in queue.

Different situations within the traffic network are simulated to test the ability of QLTSO in multiple intersections. Each intersection in the network has its own QLTSO and work individually but sharing the traffic information together for the global optimization of the traffic flow.

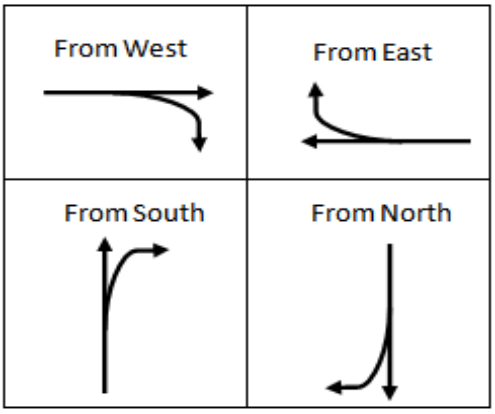

Fig.4. Traffic Phases of Traffic Intersections

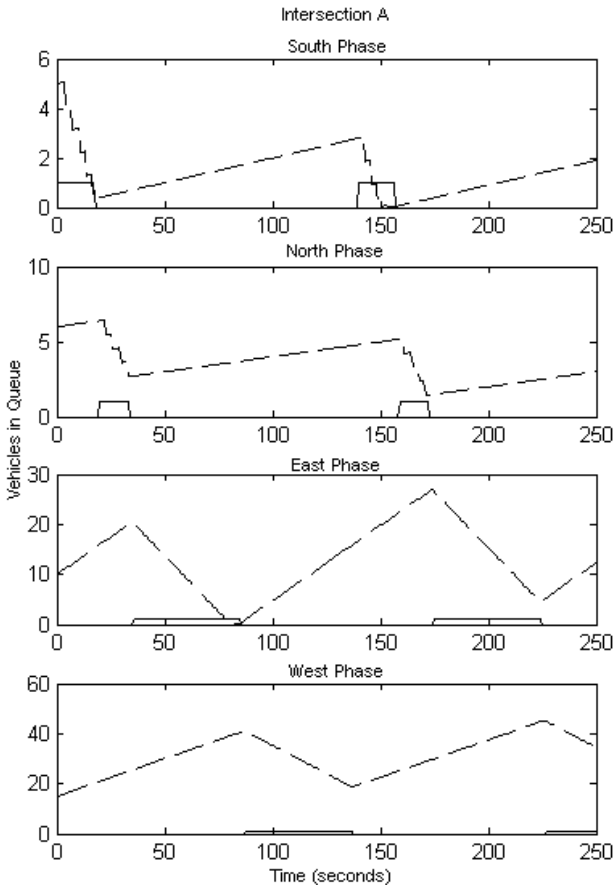

Fig.5. Green Signals Distribution of Traffic Phase

5. RESULTS AND DISCUSSION

In this paper, there are 2 types of traffic environment being simulated and analyzed. Those traffic environments are the simulations of single traffic intersection and the multi-intersection. The assessment of the QLTSO should be carried out on the single traffic intersection before it can be implemented on the multi-intersection traffic network. The results and discussion of simulations from both environments will be described in details in the following sections.

5.1 SINGLE TRAFFIC INTERSECTION

Simulations on QLTSO on a single traffic intersection are considered first before the simulations on the multi-intersection traffic network. The simulation period are set to be 3600 seconds which is equivalent to 1 hour. This allows observation on the performance of the QLTSO to be done in details. The simulation of the single intersection is carried out for 2 traffic signal management systems which are fixed-time traffic signal timing plan and the developed QLTSO. The traffic condition simulated in this test is oversaturated traffic system, where there are heavy traffic flows from East and West phase.

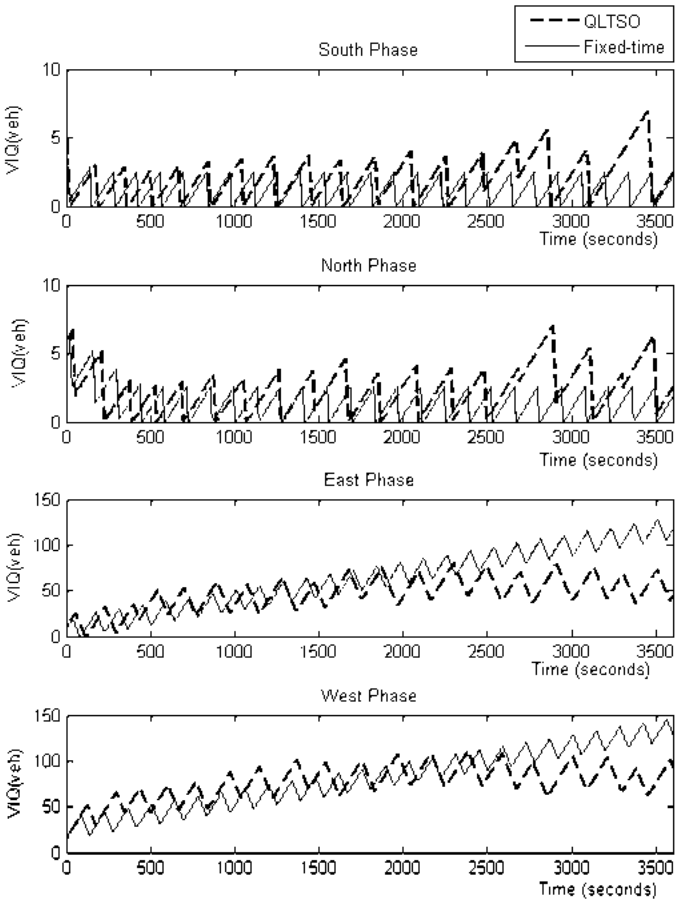

Fig.6. Green Signals Distribution of Traffic Phase

The graphs in Fig.6 show the simulation results of the single intersection. All the 4 traffic phases are illustrated to show the performance comparisons between the 2 traffic signals timing plan management. As shown in Fig.6, the dotted line represents

the developed QLTSO, while the solid lines show the results of the fixed-time traffic signal plan.

Overall, the developed QLTSO has performed better than the fixed-time traffic signal plan. From the graphs, observations show that fixed-time traffic plan is unable to maintain the vehicles in queue for traffic phase East and West by allowing the vehicles in queue at both traffic phases to increase continuously throughout the simulation period. This situation indicates the occurrence of traffic congestions. For QLTSO, the levels of vehicles in queue are increased due to the heavy incoming traffic flows, but QLTSO has successfully reduced the vehicles in queue. This shows that the QLTSO learnt from the traffic situation and counteract towards the situations.

Besides these observations, there are also slight increasing of the levels of vehicles in queue at traffic phase South and North for the QLTSO. This is because Q-learning looks at the traffic situations from all the traffic phases and decides to compromise the level of vehicles in queue at South and North for the benefits of the overall traffic conditions. This decision allows traffic phase East and West to have longer green signal period to pass more vehicles to prevent the traffic congestions. These results encourage the implementation of QLTSO in multi-intersections traffic network.

5.2 MULTI-INTERSECTIONS TRAFFIC NETWORK

After the simulation study of the single traffic intersection, the simulation of the developed QLTSO in traffic network with multiple intersections has been carried out for a period of 3600 seconds. Besides QLTSO, fixed-time traffic signal timing plan has also being simulated in the same environment setting for the analysis purposes. In this simulation, the traffic flows are those input through intersection A for north phase, south phase and west phase respectively. As for intersection B, north phase, south phase and east phase are being used as inputs for the traffic flow.

The east phase of the intersection A and west phase of the intersection B are linked together and being considered as a closed traffic environment, as the input for both phases are generated from the traffic intersections. The traffic that passes through east link of intersection A will flow into the west link of intersection B and vice versa. The results of the simulation are tabulated in Table.1 for analysis because the massive traffic information is not able to be presented in the graphs form. In Table.1, the number of vehicles that successfully passed through the traffic lights intersections is shown. From Table.1, QLTSO has the better performance as compared to the fixed-time traffic signal timing plan.

The difference in the number of vehicles passing through the intersection between the two traffic signal timing management systems can be considered as significant as both of the systems are simulated under the same environment and same kind of data. Besides from north phase and east phase of intersection B, other traffic phases are releasing more vehicles with the developed QLTSO. QLTSO successfully optimized every traffic phases during the simulation. The results from the table already indicated that QLTSO allows more vehicles passing through the intersections compared with the conventional fixed-time traffic signal timing plan. Under the same period of simulation time, QLTSO has released a total of 3921 vehicles within the traffic network compared with the fixed-time traffic management which allowing total of 3548 vehicles to travel through the traffic network.

Table.1. Number of Vehicles Pass during Simulation

Traffic Phase	Fixed Time Signal	QLTSO
INT_A_south	59	60
INT_A_north	97	68
INT_A_east	998	1106
INT_A_west	607	714
INT_B_south	68	61
INT_B_north	68	59
INT_B_east	1045	1113
INT_B_west	636	740
Total Pass	3548	3921

From the results, QLTSO has shown the ability of adapting into the traffic environment. Instead of having a fixed duration of green signals during the simulation, QLTSO adapted itself towards the different traffic demands, and calculated the optimum green signal duration for each traffic phase. In south and north phase of intersection A, QLTSO did not neglect the demands from the waiting vehicles and still able to let more vehicles from those two phases to pass through. As a result, QLTSO has the ability to optimize the traffic flows within the traffic network with the different levels of traffic demands.

Another part of the results from the simulation is shown in Fig.7 where east phase of intersection A is chosen as the traffic phase of interests. This figure consists of two graphs, the dotted line represents the fixed-time traffic signal timing plan management and the other solid line is showing the results of QLTSO. In the graphs, the results of east phase at intersection A shown that QLTSO has outperformed the fixed-time traffic signal timing plan. Both of the graphs start to show their difference in the performance at simulation time about 1500 seconds, when the fixed-time traffic signal timing plan still continue to accumulate the vehicle in queue.

Observation of the fixed-time traffic signal timing plan during 1500 seconds of simulation time and 2500 seconds of simulation time has revealed the weakness of the fixed-time traffic signal timing plan. The graph shows that the fixed-time traffic signal unable to reduce the vehicles in queue at the traffic phase. In this situation, the traffic phase does not have the enough green signal duration to release the accumulated vehicles in queue. But the fixed-time traffic signal plan already predetermined its green time durations and unable to deal with the incoming traffic flow. As for QLTSO, the ability of Q-learning learnt from the traffic environment and hence allocating longer green signal for that situation to let more vehicles to pass through has been observed.

As a result, the graphs of Fig.7 differ greatly at the end of the simulation. The fixed-time traffic signal plan which is unable to handle more incoming vehicles ends up by accumulating high vehicles in queue at the traffic phase. QLTSO is able to alter the duration of the green signal distribution according the traffic conditions, and successfully maintain the vehicles in queue at

the traffic phase to avoid over accumulated vehicles and traffic congestions.

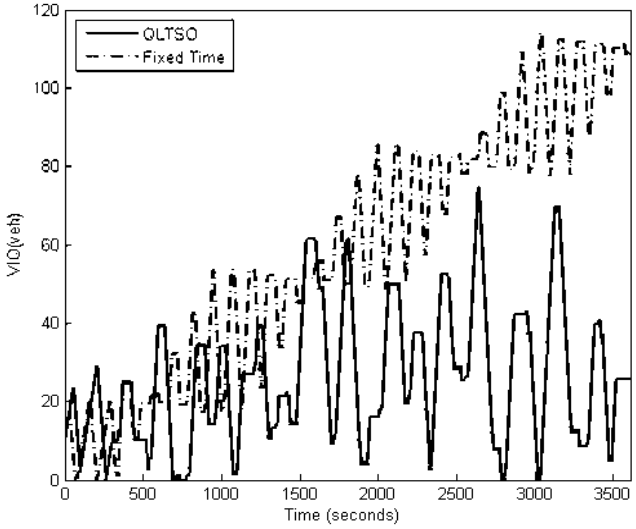

Fig.7. Results of Intersection A - East Phase

6. CONCLUSION

The developed Q-learning traffic signal optimization system is performing well throughout the simulations. This shows the potentials and the abilities of Q-learning in the traffic signal timing plan management system. Even without centralized control, the QTSO of each intersections able to work independently and having traffic information sharing with other adjacent traffic intersections. Offset between two adjacent traffic phases should be part of the future research of QLTSO in traffic network. Q-learning algorithm has shown its strength in exploration in the dynamic traffic environment as well as the adaptability towards the rapid changes of the environment by successfully managing the traffic signals distribution within the traffic networks.

ACKNOWLEDGEMENT

The authors would like to acknowledge the financial assistance from Ministry of Higher Education of Malaysia (MoHE) under Exploratory Research Grant Scheme (ERGS) No. ERGS0021-TK-1/2012, Universiti Malaysia Sabah (UMS) under UMS Research Grant Scheme (SGPUMS) No. SBK0026-TK-1/2012, and the University Postgraduate Research Scholarship Scheme (PGD) by Ministry of Science, Technology and Innovation of Malaysia (MOSTI).

REFERENCES

[1] Nicholas J. Garber and Lester A. Hoel, "*Traffic and Highway Engineering*", 3rd Edition, Thomson Learning, 2001.

[2] Y. Dai, J. Hu, D. Zhao and F. Zhu, "Neural network based online traffic signal controller design with reinforcement training", *14th International IEEE Conference on Intelligent Transportation Systems*, pp.1045-1050, 2011.

[3] V. Gradinescu, C. Gorgorin, R. Diaconescu and V. Cristea, "Adaptive Traffic Light using Car-to-Car Communication", *Proceeding of the 65th IEEE Conference on Vehicular Technology Conference*, pp.21-25, 2007.

[4] K. Khiang Tan, M. Khalid and R. Yusof, "Intelligent Traffic Lights Control by Fuzzy Logic", *Malaysian Journal of Computer Science*, Vol. 9, No. 2, pp. 29-35, 1996.

[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, No. 1, pp. 107-111, 2010.

[6] B. Park, C.J. Messer and T. Urbanik, "Traffic signal optimization program for oversaturated conditions: Genetic algorithm approach", *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1683, No. 1, pp.133-142, 1999.

[7] Y.K. Chin, K.C. Yong, N. Bolong, S.S. Yang and K.T.K. Teo, "Multiple intersections traffic signal timing optimization with genetic algorithm", *Proceeding of the IEEE International Conference on Control System, Computing and Engineering*, pp.454- 459, 2011.

[8] F. Teklu, A. Sumalee and D. Watling, "A Genetic Algorithm Approach for Optimizing Traffic Control Signals Considering Routing", *Computer-Aided Civil and Infrastructure Engineering*, Vol. 22, No. 1, pp. 31-43, 2007.

[9] 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, No. 2, pp. 128-135, 2010.

[10] 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 and Technology*, Vol. 12, No. 3, pp. 29-35, 2011.

[11] Z.Y. Liu and F.W. Ma, "On-line Reinforcement Learning Control for Urban Traffic Signals", *Proceedings of the 26th Conference on Chinese Control*, pp. 34 - 37, 2007.

[12] P.G. Balaji, X. German and D. Srinivasan, "Urban Traffic Signal Control using Reinforcement Learning Agents", *IET Intelligent Transport Systems*, Vol. 4, No. 3, pp. 177-188, 2010.

[13] B. Abdulhai, R. Pringle and G.J. Karakoulas, "Reinforcement Learning for True Adaptive Traffic Signal Control", *Journal of Transportation Engineering*, Vol. 129, No. 3, pp. 278-285, 2003.

[14] Michel Tokic, and G. Palm, "Value-Difference based Exploration: Adaptive Control between Epsilon-Greedy and Softmax", *KI 2011: Advances in Artificial Intelligence, Springer*, Vol. 7006, pp. 335-346, 2011.

[15] C.J.C.H. Watkins and P. Dayan, "Technical Note: Q-learning", *Journal on Machine Learning*, Vol. 8, No.3, pp. 279-292, 1992.

[16] Special Report, "*Highway Capacity Manual*", Transportation Research Board, National Research Council, Washington DC 113, 2000.

[17] Jabatan Kerja Raya, "A Guide to the Design of Traffic Signal", Technical Report, Arahan Teknik (Jalan), 1987.