ResearchGate

Multi-Objective Traffic Light Control System based on Bayesian Probability Interpretation

Conference Paper *in* Conference Record - IEEE Conference on Intelligent Transportation Systems · September 2012

DOI: 10.1109/ITSC.2012.6338853

3 authors:

Mohamed AbdElAziz Khamis Omar
Ejada Systems Ltd.

Walid Gomaa
Egypt-Japan University of Science and Technology

Hisham El-Shishiny
IBM

Multi-Objective Traffic Light Control System based on Bayesian Probability Interpretation

Mohamed A. Khamis
Dept. of Comp. Sc. and Eng.
Egypt-Japan Univ. of Sc. and
Tech. (E-JUST), Alexandria, Egypt
Email: mohamed.khamis@ejust.edu.eg

Walid Gomaa*
Dept. of Comp. Sc. and Eng.
Egypt-Japan Univ. of Sc. and
Tech. (E-JUST), Alexandria, Egypt
Email: walid.gomaa@ejust.edu.eg

Hisham El-Shishiny
IBM Center for Advanced Studies
in Cairo, Giza, Egypt
Email: shishiny@eg.ibm.com

***Abstract*— Traffic light control is a challenging problem in modern societies. This is due to the huge number of vehicles and the high dynamics of the traffic system. Poor traffic management causes a high rate of accidents, time losses, and negative impact on the economy as well as the environment. In this paper, we develop a multiagent traffic light control system based on a multi-objective sequential decision making framework. In order to respond effectively to the changing environment and the non-stationarity of the road network, our traffic light controller is based on the Bayesian interpretation of probability. We use the open source Green Light District (GLD) vehicle traffic simulator as a testbed framework. The change in road conditions is modeled by varying the vehicles generation probability distributions and adapting the Intelligent Driver Model (IDM) parameters to the adverse weather conditions. We have added a set of new performance indices in GLD based on collaborative learning to better evaluate the performance of our multi-objective traffic controller.**

***Keywords*— Multi-objective traffic light controller, Multiagent reinforcement learning system, Intelligent Driver Model, Bayesian probability interpretation, Adverse weather conditions**

I. INTRODUCTION

Traffic light control is a critical problem that daily affects vehicle drivers in urban areas. In this paper, we present a realistic traffic light control and simulation system that facilitates its application in controlling real traffic junctions.

The contributions of this paper are: (1) applying a reinforcement learning traffic light control system on the IDM time-continuous acceleration model (current contributions are applied on a time-discrete model, e.g., [1], [2], [3]), (2) extending the single objective traffic light controller to a multi-objective controller adaptive to the different road conditions, (3) using the Bayesian probability interpretation that allows the multi-objective controller to make real time (online) adaptation in the sense that it responds effectively to the changing environment and the non-stationarity of the road network, (4) checking the traffic light controller against varying driver behavior and traffic demand caused by adverse weather conditions, and (5) adding a set of new performance indices based on collaborative learning to the GLD traffic simulator [4], [5] for better performance evaluation. GLD is an open source traffic simulator that facilitates the traffic light control research by using customizable road networks.

*Currently on-leave from Faculty of Engineering, Alexandria University.
This work is partially funded by IBM PhD Fellowship and Faculty Award.

Traffic light control can be viewed as a multi-objective optimization problem. The multi-objective function can have a global objective for the entire road network or there may be different objectives for different parts of the road network (e.g., maximize safety especially in residential and school areas), or even different times of the day for the same part of the road network. Our multi-objective function currently includes the following components: (1) maximize the flow rate, (2) minimize the Average Trip Waiting Time (ATWT), (3) minimize the Average Trip Time (ATT), (4) minimize the Average Junction Waiting Time (AJWT), and (5) maximize the vehicles safety by taking decisions that minimize the vehicles speed. Up to our knowledge, almost all traditional traffic control methods are single objective. Duan *et al.* [6] present a multi-objective reinforcement learning traffic control. The traffic adaptation is done offline by activating one objective function at a time according to the current number of vehicles entering the network per minute.

The remaining of this paper is organized as follows, the related work is discussed in section II. A background on the GLD traffic simulation/control and a background on the IDM acceleration model is presented in section III. Our multi-objective Bayesian traffic light control system is described in section IV. System performance evaluation is presented in section V. Finally, section VI concludes the paper.

II. RELATED WORK

There exist several approaches to implement an adaptive traffic light control. These approaches are based on Reinforcement Learning (RL), Fuzzy Logic, Evolutionary Algorithm, and Artificial Neural Networks.

RL is the most promising approach for traffic light control, although traffic light-based reinforcement learning methods, e.g., [7],[8],[9] suffer from the growth in the number of states when scaling to larger networks, thus they are only applied to small scale traffic networks such as a single junction or using the same policy of a single trained junction to control a number of junctions. Vehicle-based RL methods are much applicable, e.g., [1], [2], [3], [6]. We adopt a vehicle-based RL controller [1] in which a predictor for each vehicle is used to estimate the waiting time of the vehicle alone when the traffic light is red or green. All vehicles predictions are then combined for the traffic light controller decision.

III. BACKGROUND

A. Traffic Simulation Model

In order to test our traffic light control algorithm, we need some experimentation platform, that is a traffic simulator. The widely used microscopic traffic flow simulation programs (e.g., VISSIM, Paramics, ... ,etc.) do not give the researcher an easy free way to create new research traffic light controllers. There are several research traffic simulators in the literature, e.g., [4], [10]. However, they are still primitive with a lot of oversimplifications, thus in this paper we extend and significantly enhance one of these simulators, namely MoreVTS [10] that is based on GLD [4].

In GLD, each traffic light controller can take some decisions representing the traffic light configurations that do not cause any possible accidents between vehicles moving in the green light lanes when crossing the controlled junction.

B. RL for Traffic Light Control

We adopt Wiering RL approach [1] for traffic light control. RL is used to make the agent learn traffic light control through the interaction with the environment and the feedback obtained in the form of a reward signal. Using a trial-and-error process, the agent will learn a policy that optimizes the cumulative reward gained over time. In this approach, the vehicle state means that the vehicle is at a specific lane that is controlled by a specific traffic light tl and currently at a specific position in the lane p and has some destination edge node des. Thus, the current state s is denoted by $[tl, p, des]$ and the next state s' is denoted by $[tl', p']$ where the vehicle final destination des does not change by the state transition.

The state transition probability is represented by a lookup table consisting of the probabilities $\Pr(s, a, s')$ where a denotes the action of the traffic light tl is red or green. $\Pr(a|s)$ is the probability that the traffic light tl is red or green given that a vehicle is at state s. For computing $\Pr(a|s)$, some counters are updated every time step. $C(s, a)$ counts the number of times the traffic light tl is red or green when a vehicle is at position p within the lane controlled by tl with a specific destination des. $C(s)$ counts the number of times vehicles are at state s. Thus, the probability $\Pr(a|s)$ is given by: $\Pr(a|s) = C(s, a)/C(s)$. The state transition probability $\Pr(s, a, s')$ is given by: $\Pr(s, a, s') = C(s, a, s')/C(s, a)$.

We use the Bayesian probability interpretation for traffic light control [11] that outperforms the aforementioned frequentist probability interpretation.

$Q(s, a)$ denotes the total estimated waiting time for a vehicle currently at state s until it reaches its destination des given that the action of the traffic light tl is a. The Q-function for the original controller TC-1 [1] is given by: $Q(s, a) = \sum_{s'} \Pr(s, a, s')(R(s, a, s') + \gamma V(s'))$ where γ is the future discount factor ($0 < \gamma < 1$) that is used to ensure that the Q-values are bounded. The reward function is given by $R(s, a, s') = 1$, if a vehicle stays at the same position, otherwise, $R = 0$ (vehicle moves ahead). In this paper, we define a set of new reward functions each depends on the optimized objective that is mentioned later in details.

$V(s)$ denotes the average waiting time for a vehicle at s until it reaches its destination des without knowing the traffic light action. The V-function is given by: $V(s) = \sum_a \Pr(a|s)Q(s, a)$.

The traffic light controller at every junction sums up the individual gain $Q(s, red) - Q(s, green)$ for all vehicles waiting at the traffic lights that can be set to green by the traffic light controller decision (while all other traffic lights are set to red) and chooses the traffic light configuration that minimizes the expected waiting time of all vehicles at all traffic lights met before exiting the city. Our multi-objective controller takes its decision according to the approaching vehicles as well (i.e., not only the waiting vehicles).

C. IDM Acceleration Model

We have added the IDM [12] acceleration model in GLD that is used to simulate the longitudinal dynamics, i.e., accelerations and braking decelerations of the drivers [11]. The acceleration dv/dt of a vehicle depends on its velocity v, the distance s to the front vehicle, and the velocity difference Δv (positive when approaching the front vehicle),

$$\frac{\mathrm{d}v}{\mathrm{d}t} = a\Big[1 - \Big(\frac{v}{v_0}\Big)^{\delta} - \Big(\frac{s^*}{s}\Big)^2\Big],$$

$$s^* = s_0 + min\Big[0, \Big(vT + \frac{v\Delta v}{2\sqrt{ab}}\Big)\Big].$$

The acceleration consists of two terms: (1) the desired acceleration $a[1 - (\frac{v}{v_0})^{\delta}]$ on a free road, (2) braking decelerations due to the front vehicle $-a[(\frac{s^*}{s})^2]$. The parameters of the IDM acceleration model are [12]: (1) the desired velocity on a free road, v_0, (2) the desired safety time headway, T, (3) the acceleration in usual traffic, a, (4) the braking deceleration in usual traffic, b, (5) the minimum bumper-to-bumper distance, s_0, (6) acceleration exponent, δ.

In the rest of this paper, we refer to the absolute velocity of the vehicle by its *speed* which is always positive.

IV. PROPOSED TRAFFIC LIGHT CONTROL

A. Time-Continuous Model

One of the important extensions to GLD is changing the precision of the vehicles movement in time/space domains from discrete to continuous. This update is mainly made to better model the reward values in the speed related traffic objectives, e.g., the safety objective reward that depends on the vehicle position. In the time domain, the precision of the vehicle position update time granularity δt is changed from a discrete value ($\delta t = 1$ second in the GLD model) to a real value ($\delta t = 0.25$ second as in the IDM model) that affects the speed and position calculations:

$$speed_{new} = speed_{old} + acceleration_{IDM} * \delta t,$$

$$position_{new} = position_{old} - speed_{new} * \delta t.$$

Note that in GLD the values of vehicles positions are decreasing as vehicles move from its source towards the junction. In the space domain, the positions become more precisely represented (real values instead of integer values).

B. Multi-Objective Traffic Light Control

In order to consider a multi-objective function with possibly conflicting objectives, we develop a notion of optimality by converting the multi-objective function to a single-objective with assigned weights, these weights may be different for different parts of the road network and need also to be learnt.

M. Steingröver *et al*. [2] present two traffic controllers called State Bit for Congestion (SBC) and Gain Adapted by Congestion (GAC). Adding a new bit to indicate the degree of congestion in the next lane increases the state space and slows the learning process. GAC does not learn anything permanent about congestion, in addition this method can not be easily generalized. The weight of the Q-value (congestion degree in the next lane) is neither part of the state definition nor a function in the possible next states. In our multi-objective traffic light control, we handle the drawbacks in the previous work, e.g., [6], [2] where the traffic adaptation is done now dynamically by changing the weather impact on speed parameters, also the state space representation is the same in size as the original controller, TC-1. We weight the reward of the flow rate objective and learn this weight by defining it as a function in the possible next states.

Since adverse weather conditions affects the traffic demand, safety, and the traffic flow operations [13], we choose the traffic objectives accordingly. The Q-function of the proposed multi-objective traffic light controller is given by:

$$Q(s,a)=\sum_{s'}\Pr(s,a,s')\Big(\sum_{objective_i}\big(W_i(s,a,s')*R_i(s,a,s')\big)+\gamma V(s')\Big).$$

Let the distance traveled by the vehicle in the current time step be equal to Δp (that is always positive). The first reward represents the flow rate and is given by: $R_{FR}(s,green,s')=1$ in case $tl'\neq tl$, otherwise, $R=0$. Given the number of blocks taken by the waiting vehicles and the length of the next lane are N, L, respectively. Let $W=N/L$, then the flow rate weight W_{FR}, is given by [6]:

$$W_{FR}(s,green,s')=\begin{cases}0, & \text{if } W\leq\theta,\\ 10*(W-\theta), & \text{if } \theta<W<=1,\\ 2, & \text{if } W>1.\end{cases}$$

θ is a threshold whose best value equals 0.8 [2]. If $tl'\neq tl$, W_{FR} will decrease when the next lane at tl' is free. In this case, $Q(s,green)$ will decrease and thus the gain of this policy will increase and accordingly the green phase length will increase that increases the traffic allowed to pass through, i.e., increasing the vehicles flow rate.

The second reward represents the ATWT and is given by: $R_{ATWT}(s,a,s')=1$ in case $\Delta p\simeq 0$, otherwise, $R=0$. The third reward represents the ATT and is given by: $R_{ATT}(s,a,s')=2^{-\Delta p}$. The fourth reward represents the AJWT and is given by: $R_{AJWT}(s,green,s')=0$ in case $tl'\neq tl$, otherwise, $R=1$. The fifth reward represents the safety and is given by: $R_S(s,a,s')=1/(\Delta p+1)$.

C. Bayesian Traffic Control

Wiering [1] learns the transition probability functions $\Pr(s'|s,a)$ and $\Pr(a|s)$ by just counting the frequency of the observed experiences. We implement the case where $\Pr(s'|s,a)$ changes over time due to the change in the environment dynamics in the congested traffic periods [11].

Our derivation of the posterior transition probability as a function of the simulation time step t is mentioned in details in [11]. In this paper, we estimate the weights of the next states/actions of our multi-objective traffic light control using the same Bayesian probability interpretation that we mention briefly in this section.

The primitive Bay's rule where A represents the posterior probability and B is the set of observed experiences $x_i's$ is given by: $\Pr(A|B)=\Pr(B|A)\Pr(A)/\Pr(B)$. The posterior probability at time $(t+1)$ is given by: $\text{Posterior}(t+1)=\text{Likelihood}(t+1)\text{Prior}(t+1)/\text{Normalizing Factor}(t+1)$.

We assume that $\text{Prior}(t+1)=\text{Posterior}(t)$, and the probability P_t represents either; (1) $\Pr(a|s)$: the posterior probability when fixing the state s with different action a or next state s', or (2) $\Pr(s'|s,a)$: the posterior probability when fixing the state/action pair (s,a) with different next states s'. In case of calculating $\Pr(a|s)$, the observed experience x_i equals 1 when the same situation (s,a,s') occurs and equals 0 when the same s with different a or s' situation occurs, in this case t equals the number of times the same state s occurs. In case of calculating $\Pr(s'|s,a)$, x_i equals 1 when the same situation (s,a,s') occurs and equals 0 when the same (s,a) with different s' situation occurs, in this case t equals the number of times the same situation (s,a) occurs.

$$\Pr(P_t|x_i)=\frac{\Pr(x_i|P_t)\Pr(P_t)}{\Pr(x_i)};\ \text{i = [1, \ldots, t+1]},$$

The joint probability $\Pr(x_i)$ of the observed experiences $x_i's$; i = [1, …, t+1] is equal to $\prod_{i=1}^{t+1}P_t^{x_i}(1-P_t)^{(1-x_i)}$, thus the posterior probability at time $(t+1)$ is given by:

$$\begin{aligned}f(t+1,P_t)&=\frac{L(t+1,P_t)f(t,P_t)}{\prod_{i=1}^{t+1}P_t^{x_i}(1-P_t)^{(1-x_i)}}\\&=\frac{\prod_{i=1}^{t+1}L(i,P_t)f(0,P_t)}{\prod_{i=1}^{t+1}P_t^{x_i}(1-P_t)^{(1-x_i)}};f(0,P_t)=1\end{aligned}$$

For an easier differentiation, we find $\ln f(t+1,P_t)$:

$$\begin{aligned}\ln f(t+1,P_t)=&\sum_{i=1}^{t+1}\ln(\prod_{j=1}^{i}P_t^{x_j}(1-P_t)^{(1-x_j)})\\&-\sum_{i=1}^{t+1}\ln(P_t^{x_i}(1-P_t)^{(1-x_i)}),\end{aligned}$$

Differentiating with respect to P_t and equating to 0, where $\ln f(t+1,P_t)$ and consequently $f(t+1,P_t)$ are maximum:

$$\begin{aligned}\frac{\partial\ln f(t+1,P_t)}{\partial P_t}=&\sum_{i=1}^{t+1}\sum_{j=1}^{i}\Big[\frac{x_j}{P_t}-\frac{(1-x_j)}{(1-P_t)}\Big]\\&-\sum_{i=1}^{t+1}\Big[\frac{x_i}{P_t}-\frac{(1-x_i)}{(1-P_t)}\Big],\end{aligned}$$

$$0 = \sum_{i=1}^{t}\sum_{j=1}^{i} x_j - \frac{P_t(t+1)(t+2)}{2} + P_t(t+1),$$

The posterior probability P_t as a function of t is given by:

$$P_t = \frac{2}{t(t+1)} \sum_{i=1}^{t}\sum_{j=1}^{i} x_j. \quad (1)$$

D. Impact of Adverse Weather on Traffic Conditions

Adverse weather conditions have a substantial impact on traffic speed, capacity, and flow rate [14], [15], [16]. However, the adaptation effects in the actual longitudinal driving behavior models need more investigation and a data-driven mathematical model is needed.

The results presented in [17] show that adverse weather conditions (specifically fog) led to a decrease in the desired speed, v_0. A substantial increase is observed in both of the distance to the front vehicle, s_0, and the minimum headway, T. The maximum acceleration, a, and deceleration, b, highly decrease after the start of the adverse weather condition.

In this paper, we have added in GLD the weather impacts on the IDM acceleration model parameters. The weather impacts on the speed parameters comply with the results presented in [17]. The added weather conditions can represent the weather in Egypt, specifically the weather of Alexandria city that include; light rain, normal rain, heavy rain, light fog, heavy fog, and sandstorm with $5\%, 10\%, 12\%, 25\%, 30\%, 36\%$ speed reduction than dry weather, respectively.

Cools *et al.* [15] study the impact of weather conditions on traffic intensity showing that snowfall, rainfall, and wind speed diminish traffic intensity while high temperatures increase traffic intensity. We simulate these two impacts by low and high vehicle generation frequencies, respectively.

V. PERFORMANCE EVALUATION

A. Co-learning Performance Indices

In order to examine the performance of our multi-objective controller, we need more elaborate performance indices than those originally existing in GLD. We use the same idea of co-learning driving policy proposed by Wiering in [1].

The first performance index is the co-learning ATWT. All vehicles are included in this statistics even those not arrived yet to their destination edge nodes. This is done by adding for those vehicles the expected trip waiting time $V(s)$ to the total waiting time experienced so far.

Given the set of entered vehicles is $V_{entered}$, the set of vehicles entered but not arrived so far is $V_{notArrived}$, the set of nodes (junctions or edge nodes) crossed by vehicle v is $N_{crossed}$, the time step in which v crosses node n is t_{cross}, the time step in which v engages the waiting queue leading to n is $t_{engageQ_n}$, the current time step is $t_{current}$, the time step in which v engages the waiting queue leading to the current node is $t_{engageQ_{current}}$, the expected trip waiting time of $v \in V_{notArrived}$ is $V(s_{current})$, and the number of vehicles entered so far is $N_{entered}$, the $ATWT_{colearn}$ is given by:

$$\begin{aligned} ATWT_{colearn} = \Bigg[& \sum_{v \in V_{entered}} \sum_{n \in N_{crossed}} (t_{cross} - t_{engageQ_n}) \\ & + \sum_{v \in V_{notArrived}} \Big((t_{current} - t_{engageQ_{current}}) \\ & + V(s_{current}) \Big) \Bigg] / N_{entered}. \end{aligned}$$

The second performance index is the co-learning ATT. Given the set of arrived vehicles is $V_{arrived}$, the set of vehicles entered but not arrived so far is $V_{notArrived}$, the time step in which v arrives its destination is t_{arrive}, the time step in which v starts its trip is t_{start}, the current time step is $t_{current}$, the expected trip time of $v \in V_{notArrived}$ is $V(s_{current})$, and the number of vehicles entered so far is $N_{entered}$, the $ATT_{colearn}$ is given by:

$$\begin{aligned} ATT_{colearn} = \Bigg[& \sum_{v \in V_{arrived}} (t_{arrive} - t_{start}) \\ & + \sum_{v \in V_{notArrived}} \Big((t_{current} - t_{start}) \\ & + V(s_{current}) \Big) \Bigg] / N_{entered}. \end{aligned}$$

The third performance index is the co-learning AJWT. Given the set of vehicles that crossed junctions is $V_{crossed}$, the set of vehicles entered but not crossed so far is $V_{notCrossed}$, the set of junctions crossed by vehicle v is $J_{crossed}$, the time step in which v crosses junction j is t_{cross}, the time step in which v engages the waiting queue leading to j is $t_{engageQ_j}$, the current time step is $t_{current}$, the time step in which v engages the waiting queue leading to the current junction is $t_{engageQ_{current}}$, the expected junction waiting time of $v \in V_{notCrossed}$ is $V(s_{current})$, the total number of vehicles crossed/not crossed is N_{total}, the $AJWT_{colearn}$ is given by:

$$\begin{aligned} AJWT_{colearn} = \Bigg[& \sum_{v \in V_{crossed}} \sum_{j \in J_{crossed}} (t_{cross} - t_{engageQ_j}) \\ & + \sum_{v \in V_{notCrossed}} \Big((t_{current} - t_{engageQ_{current}}) \\ & + V(s_{current}) \Big) \Bigg] / N_{total}. \end{aligned}$$

B. New Performance Indices

In the main GLD, the waiting vehicles in the edge nodes do not share in the ATWT performance index as they do not enter the network and consequently do not complete a trip. To avoid this situation, we reject the waiting vehicles outside the lane (queued in the edge nodes) and track the percentage of these rejected vehicles (the vehicles that are generated but not entered the city and will not enter anymore) and track as well these rejected vehicles occur in which edge nodes. We also have added a new performance index in GLD that is the relative throughput that equals the total number of arrived vehicles divided by the total number of entered vehicles. One more important performance index is the average vehicle speed that equals the total distance traveled by all vehicles (either have arrived or not arrived yet) divided by the total time spent in the network. This is the truest average that does not always equal the average of all vehicles average speeds.

C. Experimental Work

The original experiments in [4], [5] were done with fixed generation rates. In our experiment, the generation probability distributions change over time in order to simulate the realistic road situations of varying traffic amount entering and exiting a city due to the changing weather conditions.

We use the traffic network depicted in Fig. 1. The γ discount factor is set to 0.9 and a random traffic light decision chance is set to 0.01 for state space exploration purposes. We set all edge nodes to the same vehicle generation frequencies, and each edge node is set to the same IDM speed parameter settings according to the preset weather condition.

The duration of each simulation time step is 0.25 second. The results of this experiment is the average of ten independent runs. Every run has a seed equals to its starting PC clock time (milliseconds) and consists of 50,000 time steps.

At the simulation run time start, the inter-arrival vehicle generation distribution is set to $\mathcal{U}(a = 2, b = 4)$, this distribution leads to a congested traffic situation. In this period, we set the weather condition to *dry* corresponding to summery rush hours. At the simulation run time middle, the inter-arrival vehicle generation distribution is set to $Weibull(k = 20, \lambda = 20)$, this distribution leads to a free traffic situation. In this period, we set the weather condition to *sandstorm* that diminishes the traffic intensity.

Figures 2 and 3 compare the co-learning ATWT and ATT of our multi-objective Bayesian controller versus the TC-1 single objective controller [1] using the preset dynamic generation distributions and weather conditions, respectively.

Under the congested and adverse weather conditions, our controller significantly outperforms TC-1 (paired t-test, $P_{chance} < 0.0001$, by conventional criteria this difference is considered to be extremely statistically significant). The difference between the co-learning ATWT mean value when

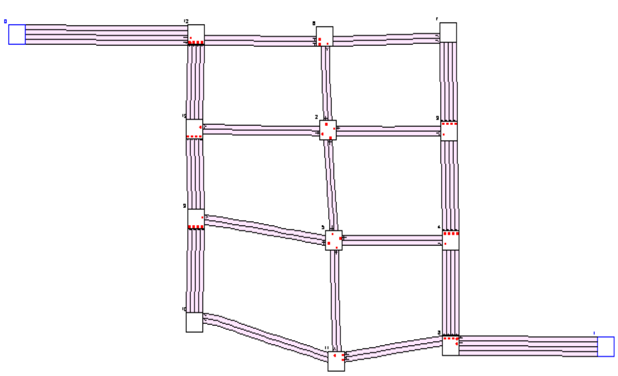

Fig. 1. Traffic network with 2 edge nodes, 10 traffic light nodes, and 2 nodes without traffic lights.

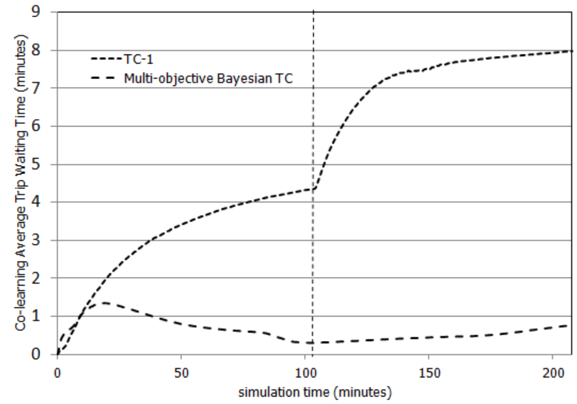

Fig. 2. Co-learning ATWT of the multi-objective Bayesian controller versus the TC-1 single objective controller.

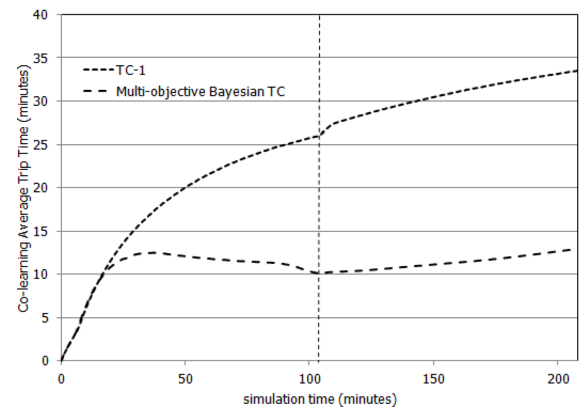

Fig. 3. Co-learning ATT of the multi-objective Bayesian controller versus the TC-1 single objective controller.

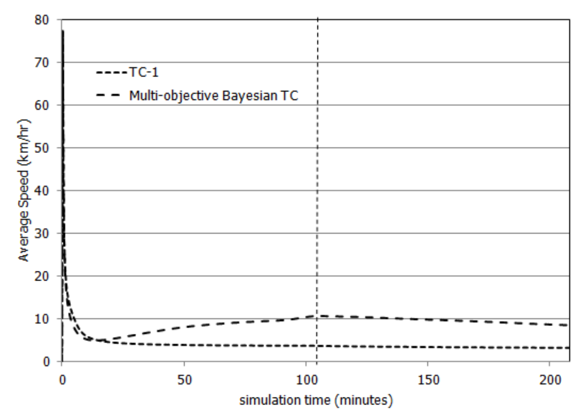

Fig. 4. Average speed of the multi-objective Bayesian controller versus the TC-1 single objective controller.

using the TC-1 single objective controller and the multi-objective Bayesian controller is equal to 4.547 minute.

Fig. 4 shows that vehicles have higher average speed when using the multi-objective Bayesian controller mainly before the start of the adverse weather conditions (where no actual need for safety deceleration).

Fig. 5 compares the traffic network throughput (i.e., the total number of arrived vehicles) for both controllers. This

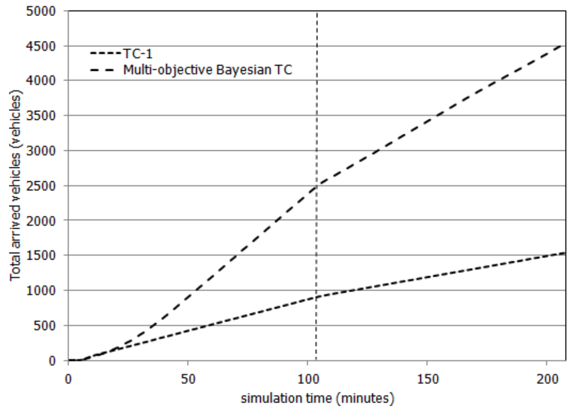

Fig. 5. Traffic network throughput of the multi-objective Bayesian controller versus the TC-1 single objective controller.

performance index is useful especially in evaluating the flow rate objective.

D. Discussion

We noticed that when a traffic congestion lasts for a long time period (especially near red signs where vehicles highly decelerate), the TC-1 controller and a single objective Bayesian based controller almost have the same performance. This is due to the Bayesian based controller learns the whole history, that can be avoided by learning partially from the history. For now, this is rare to happen in real world where congestion lasts for some limited time (which is typical to rush hours and usual adverse weather conditions) during which the performance measures will not overshot (as shown in [11]) this is due to the Bayesian based controller gives higher weights to the initial experiences as shown in (1).

We observed that our multi-objective controller gives better results when using the traffic network depicted in Fig. 1 (2 edge nodes with relatively long roads) that is due to vehicles take longer time to decelerate before approaching red signs rather than other dense networks (as in [1] that consists of 12 edge nodes with relatively short roads) in which vehicles accumulate at red traffic lights and consequently accumulate at edge nodes in a smaller time period.

VI. Conclusions and Future Work

In this paper, we develop a multi-objective traffic light control system that is adaptive to the changing environment and the non-stationarity of the road network. Under the long congested periods and adverse weather conditions, the single-objective controllers, e.g., [1], [11] can not perform well. Thus, using some innovative reward design, the multi-objective controller significantly outperforms the single objective controllers in the aforementioned road conditions.

As a future work, we want to make the weights of the multi-objective controller be function in the labeled roads (e.g., residential areas, schools, main streets). One more important objective is the environmental impact targeting a green planet. This impact needs a study on the role of our traffic light controller in minimizing the gas emissions produced by vehicles. For testing our traffic light controller behavior under further road non-stationarity, we want to check it against accidents risk when generating random perturbations that lead to sudden braking or lane-changing manoeuvres (e.g., a random obstacle might appear simulating a student passing a road in schools area that forces a driver to minimize the speed or change lane).

Acknowledgment

Special thanks are due to Dr Ahmed El-Mahdy, director of the Smart City Facility in Egypt-Japan University of Science and Technology (E-JUST), for fruitful discussions.

References

[1] M. Wiering, “Multi-agent reinforcement learning for traffic light control,” in *Proc. of the 17th International Conf. on Machine Learning (ICML2000)*, 2000, pp. 1151–1158.

[2] M. Steingröver, R. Schouten, S. Peelen, E. Nijhuis, and B. Bakker, “Reinforcement learning of traffic light controllers adapting to traffic congestion,” in *Proc. of the Seventeenth Belgium-Netherlands Conference on Artificial Intelligence Conference (BNAIC'05)*. Brussels, Belgium: Koninklijke Vlaamse Academie van Belie voor Wetenschappen en Kunsten, Oct. 2005.

[3] L. Kuyer, S. Whiteson, B. Bakker, and N. Vlassis, “Multiagent reinforcement learning for urban traffic control using coordination graphs,” *Machine Learning and Knowledge Discovery in Databases*, pp. 656–671, 2008.

[4] M. Wiering, J. Vreeken, J. V. Veenen, and A. Koopman, “Simulation and optimization of traffic in a city,” in *Proc. IEEE Intelligent Vehicle symposium (IV04)*, Parma, Italy, June 2004, pp. 453–458.

[5] M. Wiering, J. van Veenen, J. Vreeken, and A. Koopman, “Intelligent traffic light control,” Dept. of Information and Computing Sciences, Universiteit Utrecht, Tech. Rep. 2004-029, 2004.

[6] D. Houli, L. Zhiheng, and Z. Yi, “Multiobjective reinforcement learning for traffic signal control using vehicular ad hoc network,” *Journal on Advances in Signal Processing (EURASIP)*, vol. 2010, p. 7, 2010.

[7] T. L. Thorpe and C. W. Anderson, “Traffic light control using SARSA with three state representations,” IBM Corporation, Tech. Rep., 1996.

[8] M. Shoufeng, L. Ying, and L. Bao, “Agent-based learning control method for urban traffic signal of single intersection,” *Journal of Systems Engineering*, vol. 17, no. 6, pp. 526–530, 2002.

[9] B. Abdulhai, R. Pringle, and G. Karakoulas, “Reinforcement learning for true adaptive traffic signal control,” *Journal of Transportation Engineering*, vol. 129, p. 278, 2003.

[10] S. Cools, C. Gershenson, and B. DHooghe, “Self-organizing traffic lights: A realistic simulation,” *Advances in Applied Self-Organizing Systems*, pp. 41–50, 2008.

[11] M. A. Khamis, W. Gomaa, A. El-Mahdy, and A. Shoukry, “Adaptive traffic control system based on bayesian probability interpretation,” in *Proc. IEEE 2012 Japan-Egypt Conference on Electronics, Communications and Computers (JEC-ECC2012)*, march 2012, pp. 151–156.

[12] M. Treiber, A. Hennecke, and D. Helbing, “Congested traffic states in empirical observations and microscopic simulations,” *Physical Review E*, vol. 62, no. 2, pp. 1805–1824, 2000.

[13] T. Maze, M. Agarwai, and G. Burchett, “Whether weather matters to traffic demand, traffic safety, and traffic operations and flow,” *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1948, pp. 170–176, 2006.

[14] L. Goodwin, “Weather impacts on arterial traffic flow,” *Mitretek Systems Inc.*, 2002.

[15] M. Cools, E. Moons, and G. Wets, “Assessing the impact of weather on traffic intensity,” *Weather, Climate, and Society*, vol. 2, no. 1, pp. 60–68, 2010.

[16] H. Alhassan and J. Ben-Edigbe, “Highway capacity prediction in adverse weather,” *Journal of Applied Sciences*, vol. 11, pp. 2193–2199, 2011.

[17] R. Hoogendoorn, G. Tamminga, S. Hoogendoorn, and W. Daamen, “Longitudinal driving behavior under adverse weather conditions: adaptation effects, model performance and freeway capacity in case of fog,” in *Proc. IEEE 13th International Conference on Intelligent Transportation Systems (ITSC2010)*, 2010, pp. 450–455.