APPLICATION OF Q-LEARNING TO RAMP METERING IN CASES OF SIGNIFICANT CHANGES IN TRAFFIC DEMAND

B.Sc. Martin Gregurić
Research Assistant
Faculty of Transport and Traffic Sciences, University of Zagreb

Ph.D. Daniela Koltovska-Nečoska
Assistant Professor
Faculty of Technical Sciences, St. Kliment Ohridski University

Ph.D. Edouard Ivanjko
Assistant Professor
Faculty of Transport and Traffic Sciences, University of Zagreb

Abstract
The number of traffic participants has significantly increased in recent decades. Such a trend is also evident in road traffic. The result is that today's urban highways are under influence of the now increased traffic demand and cannot fulfil desired level of service anymore. Classical solution to this problem is to build new road infrastructure (urban and intercity highways, bypass roads and new segments of urban roads). That only increases the traffic demand following the saying: "If you built it, they will come!" Additional problem is that in most cases there is no more space available for infrastructure build up. Daily significant congestions appear, mostly in dense populated urban areas. To cope with the heavy congestions, new traffic control approaches are used. These approaches are solutions (services) from the domain of Intelligent Transport Systems, such as ramp metering, variable speed limit control, adaptive control of traffic lights, optimal control for consecutive intersections, etc. In this paper the control of on-ramp traffic flow on urban highways known as ramp metering is examined. To cope with significant daily changes of traffic demand various approaches with autonomic properties like self-learning are applied for ramp metering. One of the approaches with this property is reinforced learning. In this paper the Q-learning algorithm is applied to learn a local ramp metering control law in a simulation environment implemented in the microscopic simulator VISSIM. Proposed approach is tested in simulations with emphasis on a real world situation (typical part of a working day) containing significant changes in traffic demand.

1. INTRODUCTION
Congestion is an important issue in road traffic where it can significantly reduce the traffic users' level of service (LoS). Most significant negative impacts of the reduced LoS are delays in goods delivery, longer travel times in public, freight and private transport, etc. Delays induced by traffic congestion cause time losses for drivers and passengers, as well as increased fuel consumption. Considering this facts it is important to develop effective highway management control methods in order to mitigate congestion and restore originally planned LoS. Such traffic management control methods are considered under the scope of Intelligent Transport Systems (ITS) (Gregurić et al., 2014). Systems from the ITS domain are functionally built as a superstructure of several traffic control systems based on advanced optimization of transport processes with the use of information-communication infrastructure and devices. Today autonomic properties like self-learning, self-adaptation, self-configuration, etc. are also being added to improve LoS, adapt the parameters of traffic

controllers to the constantly changing traffic demand and to alleviate the work of the personal in traffic control centres.

Area where the on-ramp and mainstream flows are actually coming in interaction is known as the downstream bottleneck. Intense downstream bottlenecks occur as a consequence of platooned vehicle entry from on-ramps and lead to a traffic overspill onto the highway or to peaks in traffic demand (Papageorgiou et al., 2003). The application of ramp metering to control inflow of on-ramp traffic into mainstream can significantly improve the total travel time on urban highways (Papageorgiou and Kotsialos, 2003). Consequently, LoS is also improved. All under the assumption that the applied ramp metering algorithm correctly interprets current traffic data and can adapt to significant changes of traffic demand. The latter is important since traffic demand changes between morning and afternoon peak hours, moderate traffic during rest of daylight time and very low traffic demand during the night.

In this paper the Q-learning algorithm is applied to learn the local ramp metering control law in a simulation environment implemented in the microscopic simulator VISSIM. The design of a ramp metering algorithm with the capability to learn a local control strategy is described. The approach presented by the authors in (Ivanjko et al., 2015) is extended to enable its application in urban highways with significant changes in traffic demand. Newly extended approach is tested in simulations with emphasis on a real world situation containing significant changes in traffic demand. Goal of such simulation is to verify the proposed approach using a longer typical working day period containing traffic demand characteristics with significant changes. This should give better insight into the quality of the proposed approach versus existing short term verifications examining low and high traffic demands separately (Faresx and Gomaay, 2014).

2. RAMP METERING

Main idea of ramp metering is to resolve downstream congestions. Ramp metering uses special traffic signals at on-ramps to control the rate or size of vehicle platoons entering the mainstream flow according to the current traffic conditions (Papageorgiou et al., 2003). The information about current traffic conditions for a particular highway segment (traffic flow, speed, occupancy, etc.) is obtained by analysing real time traffic data collected from road sensors (inductive loops), cameras, etc. Sensors and traffic signalization are usually placed on the on-ramps and on the main road as presented in Fig. 1.

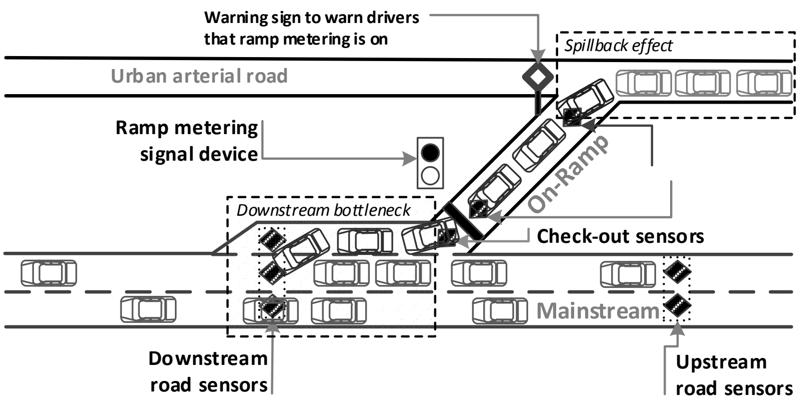

Figure 1. Downstream bottleneck and spillback effect location with basic ramp metering infrastructure (Gregurić et al., 2014)

The core part of ramp metering is an appropriate control algorithm, which produces decisions about the amount of vehicles coming from an on-ramp that are allowed to merge with the mainstream traffic flow. It is possible to divide ramp metering algorithms in two major categories (strategies): local (isolated) and coordinated. Local ramp metering algorithms take into account the traffic condition on a particular on-ramp and the nearby segment of the urban highway where they are applied. Drawback of local ramp metering algorithms is their unawareness of the overall traffic situation on the whole controlled highway segment. Today most used local ramp metering algorithm is ALINEA, due to its ratio between simplicity and efficiency (Hegyi et al., 2005). Core concept of ALINEA is to keep the downstream occupancy of the on-ramp at a specified level by adjusting the metering rate (Papageorgiou et al., 1991). Coordinated ramp metering algorithms take into account the overall highway traffic situation (Ghods et al., 2007). According to the overall highway traffic situation, the metering rate for every on-ramp is adjusted.

While the first ramp metering algorithm were purely reactive based control approaches with fixed parameters (Papageorgiou et al., 1991), today's approaches are more and more based on methods which include learning and adaptation capabilities (Mun Ng et al., 2013). The latter is crucial since traffic demand significantly changes during the day and the same control strategy can cope with this changes only if appropriate adjustment has been done. Such adjustment can include adaptation of control parameters, switching between several control strategies or including prediction of traffic parameters in the computation of the control input. Methods used for this purpose are based on neural networks, model predictive control, genetic algorithm, ant colony algorithm, etc. (Ghods et al., 2007, Hegyi et al., 2005, Koltovska and Bombol, 2014, Mun Ng et al., 2013).

3. MARKOV DECISION PROCESS AND TRAFFIC CONTROL

As mentioned, the core element of every traffic control systems is an appropriate control algorithm. To develop such a control algorithm one has to consider the underlying process. If the underlying process is related to road traffic it can be noticed that the resulting new state (traffic situation) is under the influence of random choices of traffic users (drivers and pedestrians) and computed choices of the decision maker (traffic controller). So, the new state of a traffic process describes the reaction to a control input (action a) applied to the current state (Davarynejad el al., 2011):

$$s_{n+1} \sim p\left(s_{n+1} \mid s_n, a_n\right), \tag{1}$$

where p is a probability distribution function over the state action space and s is the process state. Processes that can be described with such a model are called to be Markov decision processes (MDP) and the belonging probability distribution function is the Markov model of the whole system.

To describe a MDP a 5-tuple $\left(S, A, P, R, \gamma\right)$ is used where S is a finite set of states ($s \in S$), A is finite set of possible actions ($a \in A$), P presents the transition probability from a particular state s_n to a new state s_{n+1} if action a_n has been taken ($P^{a_n}\left(s_n, s_{n+1}\right) \geq 0$), R presents the reward received from the state transition, and γ is the discount factor ($\gamma \in \left[0,1\right]$). Regarding all possible state transitions the following condition is satisfied:

$$\sum_{s_{n+1}} P^{a_n}\left(s_n, s_{n+1}\right) = 1. \tag{2}$$

The reward function depends on the chosen action or of the so called policy function $\pi\left(s\right)$ applied on a particular state ($r\left(s, \pi\left(s\right)\right) \in R$). Is has to be defined appropriately for each underlying process that has to be controlled. Using the policy function the decision maker choses an optimal action to improve the current state of the underlying process. The discount factor γ represents the difference in importance between future and present rewards.

The problem of controlling a MDP can be defined as a problem to find the appropriate policy function that the decision maker will apply on a particular state achieving maximal reward. In traffic control the decision maker is often implemented as an intelligent agent (IA) and the policy function can be learned off-line or during operation (Koltovska and Bombol, 2014). The propriety of MDP that the effects of actions depend only on the current state can be used to teach the agent (Koltovska and Bombol, 2014). Applied agent would have the goal to find or learn the optimal policy function $\pi^*(s)$ which contains only the optimal state transition mappings from the whole state transition mapping space Π. Formally the state-value function of following a particular policy function $\pi \in \Pi$ to make the transition from state s_n to s_{n+1} can be expressed as the expected cumulative discounted reward value with the following equation (Davarynejad el al., 2011):

$$V^{\pi}(s) = E\left[\sum_{n=0}^{\infty} \gamma^k R\left(s_n, \pi(s_n)\right) \mid s_0 = s\right]. \tag{3}$$

The optimal state-value function can be defined as $V^*(s) = \max_{\pi} V^{\pi}(s)$ and it contains the optimal policy function $\pi^*(s)$. One of the approaches that can be applied to find the optimal policy function $\pi^*(s)$ is reinforcement learning (RL) using the Q-learning algorithm which is presented in the following chapter.

4. Q-LEARNING IN RAMP METERING

Reinforcement learning is one of the basic approaches of the technology based on the application of an IA. It enables the IA to act in a framework and gain new knowledge during operation. One of the significant achievements in RL was the development of the Temporal Differences Off-Policy Algorithm known as Q-learning. This algorithm provides the agent with an opportunity to learn the control policy by applying the learning rule:

$$\hat{Q}_n(s,a) \longleftarrow (1-\alpha_n)\hat{Q}_{n-1}(s,a) + \alpha_n[r + \gamma \max_{\alpha'} \hat{Q}_{n-1}(s',a')], \tag{4}$$

where the learning rate α_n is defined as

$$\alpha_n = \frac{1}{1 + visits_n(s,a)}, \tag{5}$$

and $\hat{Q}_n$ converges to Q^* (Gregurić et al., 2014) . Variable Q^* denotes the optimal action value function, $\hat{Q}_n(s,a)$ is the expected value of the previous defined value for a deterministic function case for an action a and state s, and $\hat{Q}_{n-1}(s',a')$ is the expected value of the previous defined value for the new action a' in the next state s'. The parameter γ is the discount rate in the range of $0 \leq \gamma \leq 1$, α_n is the learning rate, (s,a) presents the updated state and action during n iterations, and $visits_n(s,a)$ is the total number of visits for a state-action pair until the n^{th} iteration. The process of RL can be applied when a Markov decision process exists. In this context it is necessary to determine a set of states, a set of actions and a reward function to enable that a RL agent can learn the control policy.

4.1 States

The selection of the variables to describe the traffic process significantly varies in different applications of IA in traffic control (Koltovska and Bombol, 2014, Ivanjko et al., 2015, Faresx and Gomaay, 2014, Davarynejad el al., 2011). In this paper appropriate states that describe the overall traffic situation on a segment of an urban highway with an on-ramp. The overall traffic situation includes the mainstream and the on-ramp. Additionally, appropriate control input has to be included also. Mainstream can be described using traffic parameters like flow, speed, vehicle gap, density, etc. Density has been identified as the appropriate state for mainstream in several approaches for ramp metering (Papageorgiou et al., 2003, Papageorgiou et al., 1991, Papageorgiou and Kotsialos, 2003) so it is used in this approach also. The on-ramp traffic situation can be best described with the queue length since this state has the most significant influence on the on-ramp waiting time. Since the on-ramp traffic flow is controlled using a traffic light, its phase is the most suitable state. Within this research, the set of states *S* can be finally defined as:

$$S = \{(\phi, \rho, q); \quad \phi \in \{1,2,3\}, \quad \rho \in \{0,1,2\}, \quad q \in \{0,1,2\}\}, \tag{6}$$

where ϕ is the signal phase receiving the values $\{1,2,3\}$, ρ is density of the downstream mainstream traffic flow receiving the values $\{0,1,2\}$, and q is on ramp queue length receiving the values $\{0,1,2\}$. Detailed description of the mentioned states and representative values can be found in Table 1.

Table 1. Description of all states *S*

States *S*	Values	Description
Phases	1	Represents the "all green" phase with fixed duration of 3 s (one vehicle per green strategy)
	2	Represents the "all red" phase calculated by the ramp metering algorithm (extension of current phase duration)
	3	Represents the "ramp metering off" phase which is activated in the case of low mainstream density
AverageDensity Class	0	Downstream density is between 0 [veh/km] and 100 [veh/km]
	1	Downstream density is between 101[veh/km] and 200 [veh/km]
	2	Downstream density is larger than 201 [veh/km]
AverageQueue Class	0	On-ramp queue length is between 0 and 4 vehicles
	1	On-ramp queue length is between 5 and 7 vehicles
	2	On-ramp queue length is larger than 8 vehicles

4.2 Actions

Based on the information related to the detected state, the control agent takes an appropriate action. An agent makes a decision which action to apply on the traffic process every 3 seconds. For each state, the agent can only decide between two actions in the case of local ramp metering. First possible action is denoted with the value 1. Mentioned action suggests that it is necessary to stay in the current signal phase. Second possible action indicates the necessity to change the current signal phase and it is denoted by the value 2. The set of actions consist of two values only $\mathrm{A} \in \{1,2\}$ modelling the signal phase change. It has to be noticed here that for ramp metering only the green and red traffic light phases are used. From this reason two actions were enough.

4.3 Reward function

The reward function is a function that depends on the system's state and the action taken. Each action, derived by the agent, influences the environment. So upon the completion of an action, the environment is in a new state. The agent is rewarded if the action was good i.e. the new system state is closer to the desired state. Reward for a particular traffic solution is added in case when the on-ramp queue category has the value 0 or/and 1. Additional reward is added in the case when mainstream density reaches category 0.

To learn the control strategy, the RL agent requires a simulated traffic system environment. The simulation platform that is being used is the VISSIM microscopic simulator (Koltovska and Bombol, 2014). The program for communication among VISSIM simulator, the database and the RL algorithm is developed using the C# program language. A schematic representation of the implemented solution is given in Fig. 2 (Koltovska and Bombol, 2014).

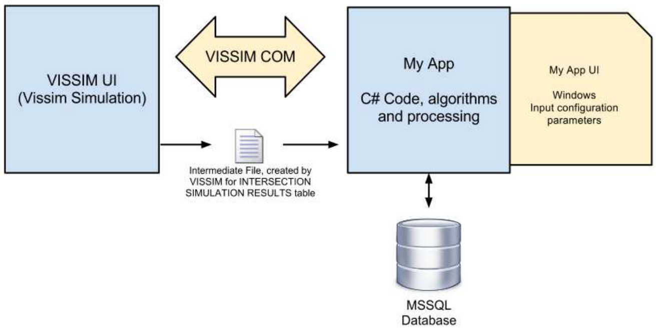

Figure 2. The process of communicating and interaction among the main elements

5. SIMULATION RESULTS

A highway model with one on-ramp is created in order to test the proposed local ramp metering approach based on the Q-learning algorithm. The mentioned model contains one signal head (traffic light with one red and one green light) and three groups of traffic sensors. First sensor is placed on the entrance of the on-ramp, the second sensor detects the queue length and the third sensor has the task to obtain mainstream traffic data. Graphical representation of the highway model used for simulation testing is presented in Fig. 3.

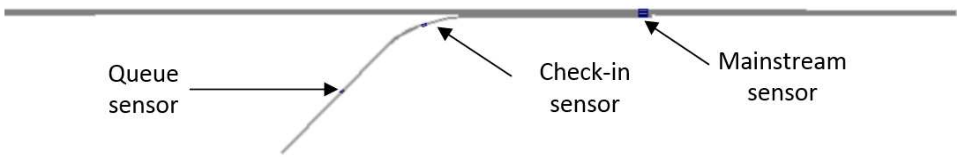

Figure 3. Graphical representation of the highway model

The duration of the simulation was set to 6 hours (~21600 seconds). Mainstream and on-ramp traffic demand values of the simulation model are reconstructed according to the traffic demand characteristics obtained from two inductive loops (two separate datasets) placed on the Ljubljana bypass (mainstream and one on-ramp). Dataset with higher traffic demand presents mainstream flow, and data set with lower traffic demand presents on-ramp flow. Both datasets obtained from the Ljubljana bypass are available as hourly traffic flow dataset, so they were interpolated in 5 minute interval traffic demand datasets using a spline function. So, every hour has twelve 5 minute intervals. The interpolated traffic demand datasets from the Ljubljana bypass are used as a template to build a simulation traffic demand datasets for on-ramp and mainstream flow.

Traffic demand values for every 5 minutes interval used in the simulation model are presented in Fig. 4. Mentioned traffic demand values describe the traffic situation between 6:00h and 12:00h in the morning. Both traffic demand datasets contain several significant and rapid changes of traffic demand according to the characteristic given in Fig. 4. This kind of change in traffic demand is used to evaluate the ability of the Q-learning ramp metering algorithm to turn-off itself in case of low traffic demand. Used profile of the on-ramp and mainstream traffic demands in Fig. 4 have two highlighted peaks with higher traffic demand. When the mentioned peaks are reached, second type ramp metering should provide higher rewards to the first two phases (green and red light phases) and no reward to the third phase, which turns the ramp metering off.

Simulation results with and without the use of ramp metering are compared. Two implementations of the Q-learning based ramp metering algorithms are used. First type provides ramp metering during the whole simulation run (Ivanjko et al., 2015). Second type proposed in this work conducts ramp metering control only during the periods with an increased traffic demand. Results of this comparative analysis are presented in Table 2. It has to be noticed that the simulator VISSIM computes total travel time as the sum of travel times of all active and arrived vehicles in a simulation scenario. This is the reason for the large total travel time values in Table 2 since the average mainstream travel time for one vehicle is about 1 min.

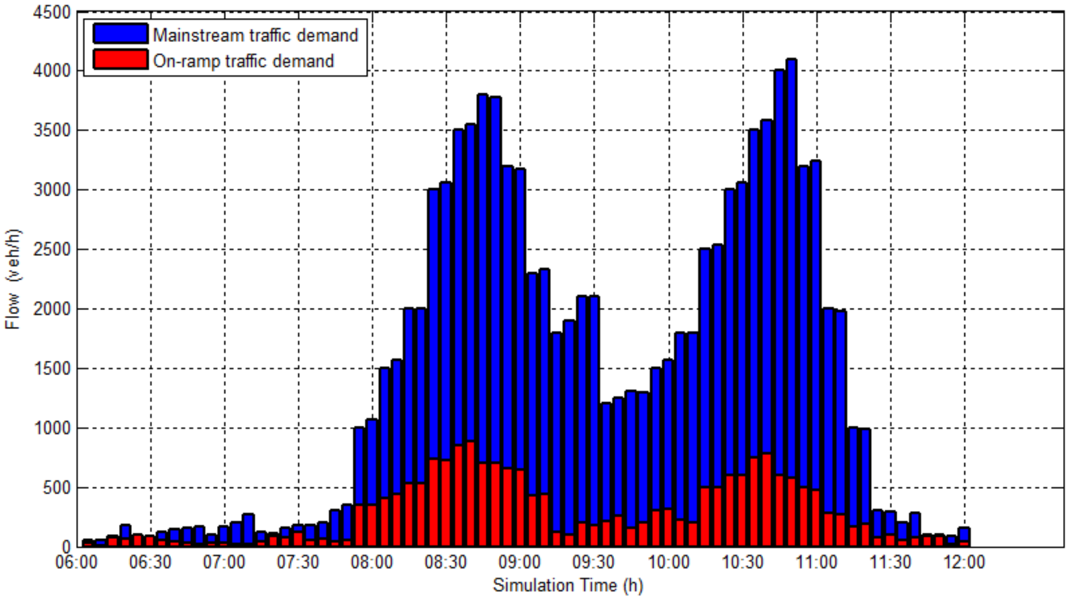

Figure 4. Traffic demand values for every 5 minute interval

Table 2. Comparative analysis between no ramp metering and Q-learning based ramp metering

Performance Measures	Control Method		
	No ramp metering	Q learning	
		First type	Second type
Average Mainstream Travel Time (s)	61.9	58.2	57.4
Average On-ramp Travel Time (s)	97.4	182.9	100.4
Average downstream Speed (km/h)	82.1	87.3	88.5
Average on-ramp Speed (km/h)	30.0	16	29.1
Average Speed of the whole highway (km/h)	70.19	61.56	73.61
Total Delay (h)	60.01	88.07	50.76
Total Travel Time (h)	200.84	228.88	191.81

According to the results given in Table 2 it is possible to conclude that the first type ramp metering algorithm based on Q-learning provides higher values of total, mainstream and on-ramp related travel time. Main reason for this result is the unnecessary reduction of the on-ramp flow entering the mainstream flow during periods with low traffic demand. Described behavior consequentially induces a larger total delay due to longer on-ramp queues created during the simulation time intervals with a low traffic demand. Even short periods of red light in the signal plan between two consecutive green light phases, may induce an increase in total delay at low traffic demand. This is the reason why this paper considers design of the second type ramp metering algorithm based on Q-learning. It contains the turn-off ability for cases with low traffic demand. This is enabled by adding a third phase into the control strategy of the first type ramp metering algorithm. The added third phase is a turn-off phase. It stops the ramp metering algorithm in the case of low traffic demand inducing a phase with green light only. The vehicles from the on-ramp can now enter mainstream without interruption. Changes in traffic demand are detected by measuring the mainstream density.

According to the results given in Table 2 it is also possible to conclude that the second type ramp metering algorithm based on Q-learning produces lower total, mainstream and on-ramp travel time compared to the first type. Furthermore, both ramp metering algorithms provide lower average mainstream travel time compared to the situation without ramp metering. This results can be explained by the on-ramp restrictions, which are the result of applying ramp metering algorithms. Second type ramp metering algorithm provides significantly lower average on-ramp travel time compared to the first type of ramp metering due to the added "turn-off" phase. This is resulting from the fact that ramp metering is turned off during low traffic demand periods. Drawback of this approach is higher average travel time on the on-ramps compared to the situation without ramp metering. In the other hand it produces decreased travel time for mainstream traffic.

Total Delay is highest in case of the first type ramp metering algorithm as it could be expected. The lowest total Delay is achieved by using the second type ramp metering algorithm. Mentioned result verify its ability to resolve congestion successfully without creating unnecessary long on-ramp queues in time periods with low traffic demand.

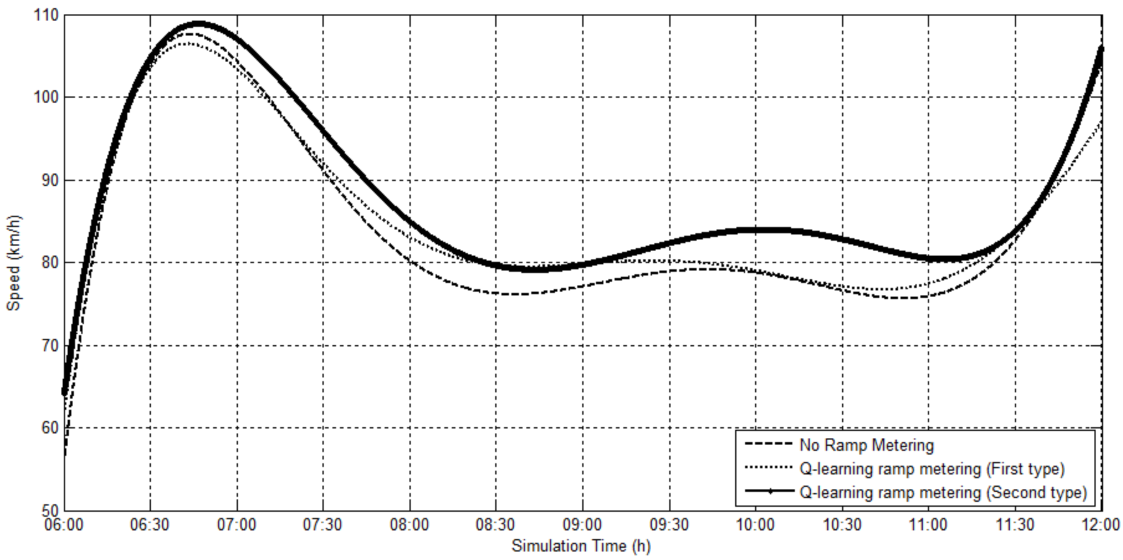

Figure 5. Comparative analysis of the mainstream traffic speed values

Additionally, this paper presents comparative analysis between mainstream speed and density. For the sake of clarity of the monitored traffic parameters change trends presentation a polynomial fit of the 6 order has been done and presented in Figs. 5 and 6 (Ivanjko et al., 2015). In Fig. 5 it is possible to see that the second type ramp metering algorithm achieves higher values of mainstream speed compared to other analyzed highway control strategies due to its less selective nature regarding on-ramp flow. First type of ramp metering algorithm produces higher speed compared to the situation without ramp metering, but slightly lower regarding the second ramp metering algorithm. More restrictive nature of the first type ramp metering algorithm compared to the s type is the main reason for this result. Furthermore, second type ramp metering algorithm achieves more homogenized mainstream speeds during the whole simulation run compared to other involved highway strategies.

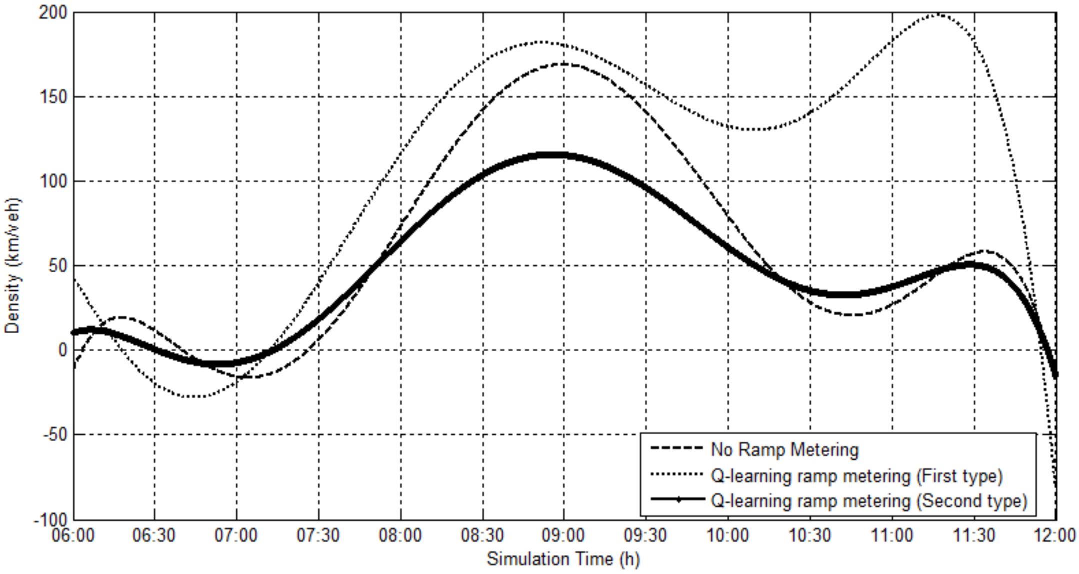

Figure 6. Comparative analysis of mainstream traffic density values

According to Fig. 6 mainstream density is lowest in the case of the second type ramp metering algorithm due to its selective application of ramp metering and better homogenization of mainstream speeds. Situation without ramp metering produces highest values of mainstream density. Reason for mentioned results is the uncontrolled access of vehicles from the on-ramp into mainstream. This action increase mainstream density and decreases mainstream speed. The contrary situation when ramp metering restricts the access of vehicles from the on-ramp into mainstream results with a decrease of mainstream density and increase in mainstream speed. Consequence is higher throughput of the urban highway allowing a short term vehicle wave to pass without creating significant congestion.

6. CONCLUSION AND FUTURE WORK

In this paper the Q-learning algorithm is applied to learn a local ramp metering control law in a simulation environment implemented in the microscopic simulator VISSIM. The first proof of concept results of an augmented Q-learning based ramp metering algorithm, which can cope with significant changes in traffic demand, are given and discussed. The motivation for this research lays in the application of an IA that can successfully cope with conditions of low and high traffic demand during a longer part of a typical working day. In such a case ramp metering is usually switched off during periods with low traffic demand. This has to be done manually or in fixed time intervals if no appropriate control strategy is applied. The result is lower LoS then the case were an automatic switching according to the current traffic situation is done.

Proposed approach was tested in simulations using a VISSIM based micro-simulation setup. Several types of quality measures based on travel time (total travel time, on-ramp travel time and mainstream travel time) are used to analyse the obtained results. Additionally, overall average mainstream speeds, mainstream density and total delay were also included into the analysis. First simulation results presented in this paper are promising. Improvements compared to the situation without ramp metering are noticeable in the improvement of mainstream travel time and average mainstream speed values. When significant traffic demand changes occur, the proposed second type ramp metering algorithm improves the urban highway LoS quality regarding all examined measures. It successfully turns the ramp metering off in cases of low traffic demand and turns it on when high traffic demand occurs.

Future work on this topic will include tuning of the applied reward system. Additionally, cooperation between several on-ramps and additional highway control systems like variable speed limit control based on learning an IA will be also examined.

AKNOWLEDGEMENTS

The research reported in this paper is partially funded by the European Union from the FP7 – Collaborative Project: “Intelligent Cooperative Sensing for Improved traffic efficiency – ICSI” (FP7-317671) and by the EU COST action TU1102 “Towards autonomic road transport support systems”.

REFERENCES

Ghods, A. H., Kian, A. R. and Tabibi, M. (2007) A Genetic-Fuzzy Control Application to Ramp Metering and Variable Speed Limit Control, IEEE International Conference on Systems, Man and Cybernetics, ISIC 2007, 10. Oct. 2007, Montreal, Canada, pp. 1723-1728

Hegyi, A., De Schutter, B. and Hellendoorn, H. (2005) Model predictive control for optimal coordination of ramp metering and variable speed limits, Transportation Research Part C, Vol. 13, No. 3, pp. 185–209

Papageorgiou, M., Diakaki, C., Dinopoulou, V., Kotsialos, A. and Wang, Y. (2003) Review of Road Traffic Control Strategies, Proceedings of the IEEE, December 2003, Vol. 91, No. 12, pp. 2043-2067.

Papageorgiou, M., Hadj-Salem, H., Blosseville, J. M. (1991) ALINEA: a Local Feedback Control Law for Onramp Metering, Transportation Research Record, 1320, 1991, pp. 58-64.

Papageorgiou, M. and Kotsialos, A. (2002) Freeway ramp metering: An overview, Proceedings of IEEE 5th Intelligent Transportation Systems Conference, Oakland, USA, pp. 271-281.

Gregurić, M., Ivanjko, E. and Mandžuka, S. (2014) New concepts for urban highways control, REAL CORP 2014. - Clever plans for smart cities, Vienna, Austria, 19.-20. May, 2014, pp. 423 – 431.

Koltovska, D. and Bombol, K. (2014) Intelligent Agent Based Traffic Signal Control on Isolated Intersections, TEM Journal, Vol. 3, No. 3, pp. 216-222.

Ivanjko, E., Koltovska-Nečoska, D., Gregurić, M., Vujić, M., Jurković, G. and Mandžuka, S. (2015) Ramp metering control based on the Q-learning algorithm, Cybernetics and Information Technologies, Special Issue on Control in Transportation Systems, Vol. 15, No. 5, pp. 88-97, DOI: 10.1515/cait-2015-0019.

Faresx, A. and Gomaay, W. (2014) Freeway Ramp-Metering Control based on Reinforcement Learning, 11th IEEE International Conference on Control & Automation (ICCA), June 18-20, 2014. Taichung, Taiwan, pp. 1226-1231.

Davarynejad, M., Hegyi, A., Vrancken, J. and van den Berg, J. (2011) Motorway Ramp-Metering Control with Queuing Consideration using Q-Learning, 14th International IEEE Conference on Intelligent Transportation Systems, Washington, DC, USA. October 5-7, 2011, pp. 1652-1658.

Mun Ng, K., Bin Ibne Reaz, M., Alauddin Mohd Ali, M. and Guy Chang, T. (2013) A brief survey on advances of control and intelligent systems methods for traffic-responsive control of urban networks, Technical Gazette, Vol. 20, No. 3, pp. 555-562.