Context-Aware Multi-modal Traffic Management in ITS: A Q-Learning based Algorithm

Adel Mounir SAID †, Ahmed SOUA*, Emad ABD-ELRAHMAN**†, Hossam AFIFI**

† National Telecommunication Institute-NTI, Egypt; * National Institute of Standards and Technology- NIST, USA; ** Institut Mines-Télécom IMT, TélécomSudParis, France.

amounir@nti.sci.eg, ahmed.soua@gmail.com, {emad.abd_elrahman, hossam.afifi}@telecom-sudparis.eu

Abstract— **Multi-modal traffic management in Intelligent Transportation Systems (ITS) aims to provide a more efficient traffic regulation to passengers and reduce congestion and obstruction in the roads. In spite of the outstanding progress made in this research filed; traffic management still a very challenging problem regarding the multiple factors that have to be taken into account in any proposed solution. To tackle this problem, this paper introduces a collaborative model based context awareness multi-modal traffic management aiming at providing an efficient way to manage the traffic inside a transportation station. In this model (Multi-Layers Stations: the stations that have different intersections for different means of transport), the traffic management is based on a Q-learning technique that takes into account the context awareness parameters to provide more potent decisions. The learning technique offers the opportunity to the system (transportation station) to adapt dynamically its decision (choice of the best transportation mean) based on feedbacks provided by the passengers traveling from that specified station and thus optimize their journey through the transportation network. The efficacy of our proposed technique is validated through extensive simulations for different layers of transport means like metros, trains, and buses. Our proposal holds for any ITS system decisions provided the availability of real-time traces about the passengers passing by any station.**

Keywords—ITS, Multi-modal Traffic, context awareness, Q-Learning.

I. INTRODUCTION

Traffic congestion is considered a serious problem that can dramatically affect the passengers' experience through the network by resulting in longer trip times, lost productivity, wasted energy due to the increased vehicular queuing, driver frustration, and safety issues. Intelligent Transportation Systems (ITS) are systems that aim to support transportation of humans and goods with information and communication technologies in order to efficiently and safely use the transport infrastructure and transport means (cars, trains, planes, and ships) [1].

One of the main key components of an ITS system is called the Advanced Traffic Management System (ATMS) [2]. It is responsible for the management of vehicles traffic flows. It consists of detectors, surveillance systems, and communication tools that aim to monitor and improve the traffic regulation and hence ensure a smooth flow for travelling users.

On the other hand, Context Awareness (CA) is defined as the conceptual abstraction that is used to analyze a collected data about certain circumstances. The use of the CA methodology in ITS could provide a rich service environment since it could be used to build ITS applications adapted to the demands of the passengers' requirements using the available transportation data such as [3-5]. In fact, the context awareness ITS applications are built based on collecting data from different sources such vehicles on the roads, passengers, GPS, etc. Moreover, the way of delivering the resultant crop of these applications in a convenient and efficient way is very important. Hence, applying context awareness in ITS makes it possible to go a step further in providing real-time feedbacks and tailoring control strategies to the actual state of the network. This would result in a better management strategy of the transportation network.

To grasp the potential advantages of using such context aware techniques in traffic management, we propose in this paper an adaptive ATMS data collection model for managing the traffic flows based on passengers' feedbacks. The main target is to choose/recommend the best transport mean to reach a destination at time (t). Our main goal is to alleviate the long trip delay, and to provide the possible comfort and best fare transport for travelling users. In addition, to adapt our proposal to the dynamic information provided by users, a reinforcement learning technique is proposed for the ATMS mode. Trip delay, transport comfort, and transport fare parameters are used in a reward formula for a better selection of transport means.

Moreover, we study the relations of population in such kinds of networks with different layers of the architecture and the effects of each layer on the other. This study needs statistics and behavior analysis of people coupled with some optimization algorithms and modeling techniques for traffic management decisions in such type of transportation networks. These statistics are collected from the traces of using metro (French underground train), train, and bus transportation means in France, as it will be described in the model validation section.

The rest of this paper is organized as follows: Section II introduces the relevant proposals carried out in ITS managements and decision-makings. Then, in Section III, Q-Learning model is presented with some highlights on motivations for the proposed algorithm followed by an underlining of the basic design of our proposal and its components. Afterward, the proposed technique is evaluated through extensive simulations using real-time statistics and discusses the obtained results.

II. Related Work

The term "Intelligent Agent" has been defined as *'The agent that could receive percepts from the environment and take actions'* [6]. In our study, the intelligent agent is equal to the transportation station. This latter will be able to make theoretical decisions based on feedbacks gathered from passing passengers. Based on these feedbacks, a learning system can optimize its decisions in order to propose the necessary guides for passengers passing this station.

We present hereafter some insights of the relevant proposals that tackled the issue of intelligent systems predictions based on reported feedbacks. Although the use of several approaches to investigate traffic management issues, the majority of these propositions are based on the use of reinforcement and Q-learning techniques [7-16]. Among those works, the time-arrival estimation technique introduced in [7] proposed a prediction engine system that built its visions and decisions based on the context behaviors of drivers and vehicles. It proposed a technique for increasing the estimation time of arrival depending on the observed patterns in traffic flow and driver travel behavior. Authors build a context pattern for the driver behavior. It estimates the arrival time of the vehicle from the previous travelled routes. The main weakness of this work is that authors did not propose backup routes for the driver to take in some specific situations such as rush hours, shortest route to the destination in case of emergency, etc. It could be useful only for tracking.

Authors in [8] propose a reinforcement learning-based solution for traffic control in case of the conventional pre-timed signal control and independent control agents. In this scheme, they provided a closed control loop for implementing the control agent learning based on the feedback reward. The control agent would achieve a convergence with the control police by interacting with the environment to learn and achieve the optimal mapping between the environment's state and the corresponding optimal control action.

Moreover they extended their work in [9] by using the reinforcement learning in a decentralized adaptive real-time traffic signal control system to minimize the total vehicle delay in the traffic network. Q-Learning agent learns the optimal mapping between the environment's state and the corresponding control action based on accumulating reward.

The Q-learning based traffic signal control models proposed in [9] deal with time varying and stochastic traffic flow problems. Moreover, authors in [11] used the multi-agent reinforcement learning based multi-cross roads traffic signal control to predict the traffic flow.

J.Songet et al. in [13] focus on situations where all the control agents coordinate with each other to adapt the traffic signals in roads. For this purpose, they propose the use of independent control agent mode and integrated mode coupled with the use of reinforcement learning to resolve the problem of agent communication. Simulation results outline the minimization of the intersection delay and travel time using their proposition.

Finally, in [10], authors investigate the problem of agent based self-optimization towards multiple polices and propose using distributed w-learning as a reinforcement learning to tackle this challenge. The goal behind their idea is to improve the performance of multiple policies deployed simultaneously in self-organizing urban traffic control.

Based on the previous study of existing works, we observe that the Q-learning technique has been used in a useful way for providing a traffic control and hence reducing the expected congestion in rush hours. On the contrary, these traditional techniques of alleviating the congestion on the roads do not solve the problem from its roots. Therefore, we believe that there is a missing part for completing the traffic congestion solutions and hence controlling the sources of these troubles on the roads. In other words, these solutions need to alleviate the congestion and delay problems starting from the traffic mean itself.

To fill this gap, we propose hereinafter a multi-modal traffic management technique based on a Q-learning algorithm. The learning process is fed by passengers' feedbacks in order to learn from their experience. The traffic analysis, based on number of users going in and out from the station, will help to perform the following two actions:

- People guidance information (best routing for people in the stations): through this action, the optimal way can be showed inside the intelligent station screens.
- Feedbacks for smart station in the new ITS environment and system prediction: those feedbacks will help in the future reactions in the new smart cities design and implementations (network planning, etc.).

III. The Proposed Model based Q-Learning

A. Q-Learning Model

When passenger reaches any Transport Station (TS) at a time (t), he has to choose a transportation mean to reach a destination. To facilitate the passenger travel, we propose a model for the Transportation Decision System (TDS) using a Q-learning based recommendation technique to offer to the passengers the best Transport Mean i (TMi) in function of several traffic parameters. The TDS model is shown in Fig. 1.

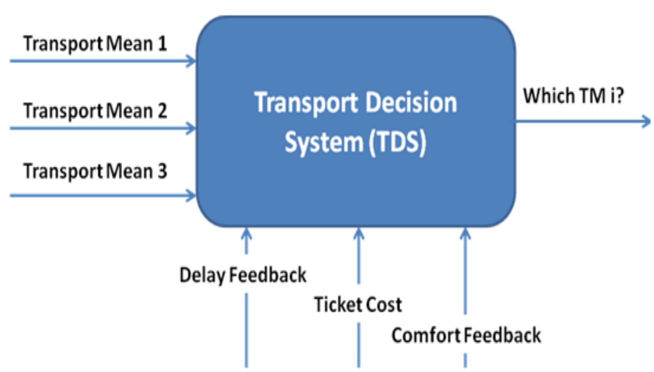

Fig. 1. TDS Model

The Q-learning process offers to the station the possibility to learn from its experience in guiding the passenger and thus it can optimize its interaction in the future. The experience here is expressed in terms of rewards and costs resulting from previous decisions. This technique is initiated at the TS profiting from the feedbacks of the different passengers passing from TS to *D* at time (*t*). When a passenger is traveling from TS to a destination taking a TMi, this latter will be rewarded or sanctioned based on several key parameters: the delay to reach the proposed destination, the passenger comfort feeling feedback, and the cost of the trip for each TMi. Thus, TS will optimize its TMs proposition through its experience during week days. In addition, passengers collaborate with the proposed TDS by their feedbacks to update and optimize its experience for more accurate decisions.

In this work, we model the decision problem inside TS as a Markov Decision Problem (MDP), which can be solved by reinforcement learning. The station decides at each state (hour of the day, or another duration time depending on the traffic) the best TMi to reach the destination based on its experience. After taking a decision, the station gets a reward or a cost from the environment. The Markov decision problem is defined as a tuple {*i, t, a, r*}.

- *i*: is the object under study. In our model, it corresponds to the TMi.
- *t*: is the states set. In our work, it corresponds to day hours or the time study.
- *a*: is the set of decisions a TS can takes: In the model, the action of TS is to select the best TMi that will fulfils the passengers' needs. In our case, the possible TMs are metro, train, and bus.
- *r*: is the immediate reward a passenger my receive after taking a decision *a*.

To solve this MDP problem, we propose to use a reinforcement-learning algorithm that is called the Q-learning technique.

A $Q_i(t,a)$ matrix is used to store the learned reward/sanction for each state and action pair for a TMi at time (*t*). For example $Q_i(t,a)$ is the expected reward for taking an action (a), which corresponds to using TMi at time (*t*). The updating function of $Q_i(t,a)$ is defined as:

$$Q_i(t,a) = (1-\alpha) * Q_i(t,a) + (1-\alpha) * \left(r_i(t) + \gamma * max_{a'} Q_i(t+1,a')\right) \quad (1)$$

Where:

- α: denotes the learning rate which models here how quickly the Q-values can change with the dynamic feedbacks given by the users.
- γ: refers to the discount factor. It models the fact that immediate reward is or is not more valuable than future reward. If this value is high, future rewards are more valued than immediate ones. In the opposite case, the learning algorithm count immediate rewards more strongly.
- r_i: represents the expected immediate reward of choosing TMi at time (*t*).
- $max_{a'} Q_i(t+1,a')$: models the maximum expected future reward when the system reaches the state (*t*+1) after taking the decision a.

The most important challenge to successfully achieve the learning performance is to define the suitable reward function. In fact, the TDS at each station will use this function to update its forwarding and decision or guiding policy.

For the immediate rewards, we consider the most relevant parameters effective in the choice of the suitable transportation mean. The first factor is the delay of the passenger trip relative to the number of stations to pass in order to reach the final station *D*. The reward should be higher for a trip with fewer intermediate stations. Second, the decision is also based on the cost of the trip. We suppose that passengers appreciate more the cheapest transportation mean. The third proposed factor is the comfortability coefficient of using a TMi at time (*t*). This parameter reflects the degree of comfortability feedback of each passenger when using the chosen TMi. The comfortability here means for example the number of passengers, the number of chairs, etc. Thereby, a passenger will be more interested with a more comfort mean. Finally, the respect of arrival time will be also a crucial parameter that will affect the decision of the TS. For this reason, when the TMi arrives after its planned arrival time, it will receive a sanction. In the other case, it will receive a reward. Moreover, in case all the TMs are delayed, the bonus will be added to the TMs according to the lowest delay achieved and the TMi of a highest delay will have no bonus.

Based on these decision factors, we formulate the reward function as follows:

$$r_i(t) = \frac{1}{\left(\frac{B_1(t)}{nb_{Stations}}\right)} + \left(\frac{Cost}{B_2}\right) + (B_3(t) * Comfortability) + B_4(t) \quad (2)$$

Where B_1 is the coefficient related to the trip duration in minutes according to the use of TMi, and nb_stations is the number of station between TS and *D*. B_2 is a coefficient that reflects the cost of the TMi ticket relative to a mean value, which equals to 1. The ticket value is fixed and independent on the daytime or weekdays. B_3 is the passenger feedback about the TMi comfortability (satisfaction percentage). The comfortability coefficient parameter is the default value, which equals to 1. B_1 and B_3 are function of time. B_4 is a bonus added to the low transport means that has lower delays. The transport mean of a maximum delay has zero bonus.

The reward function, as shown in Eq. (2), considers the most important parameters that can influence the transport station decision to choose the most suitable transport mean for the passenger at time (*t*).

Thus, more reward is assigned to the TMi that offers less traveling time, cost fees, and ensures at the same time good traveling conditions (respect of time and comfortability). In our model, we assume the TDS is applied in a station containing

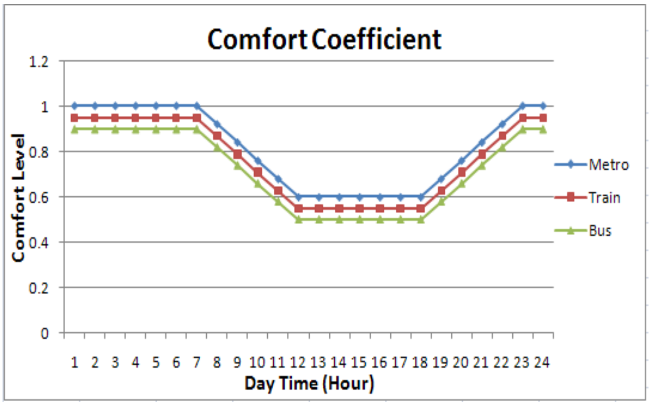

Fig. 2. Passengers comfortability feedback

three different TMs (metro, train, and bus). Each of these transportation means has its ticket cost relative to an average cost (default tariff cost) and comfortability coefficient (based on users experiences). We consider the passengers' comfortability feedback as shown in Fig. 2.

These results have been obtained through several observations. Obviously, the metro is supposed to be more comfort than other TMs since it is more frequent. The train comes in the second rank even if it accommodates more chairs and then the bus.

For the last parameter (respect of arrival time), we also have relied on real statistics that give us an insight on how these different transportation means respect or not their arrival time. These statistics depend on real traces collected as it will be described in the following section for three contiguous working days gathered from SNCF site [5].

We can obviously see in Fig. 3, 4 and 5 that in average the metro is the best TMi that respects its arrival times. However, in some rush hours, the metro can exceed its schedule and then have to be penalized (as also all the other transportation means) during the decision procedure. The TDS optimizes its decisions based on the experience of transportation systems during the weekdays. TDS updates its decisions calculations (rewards / sanctions) based on the collected passengers' feedbacks and hence the system experience.

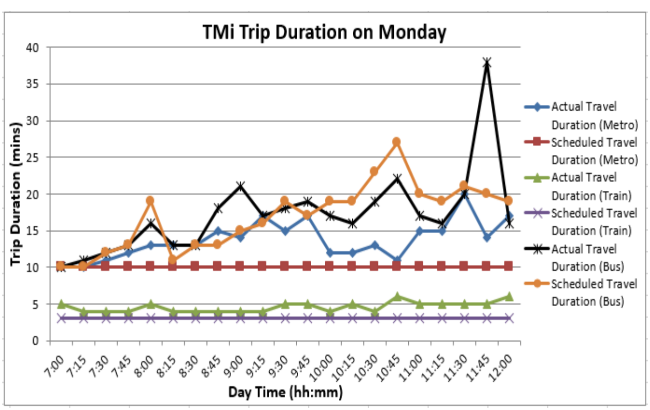

Fig. 3. Metro, Train, and Bus trip duration on Monday

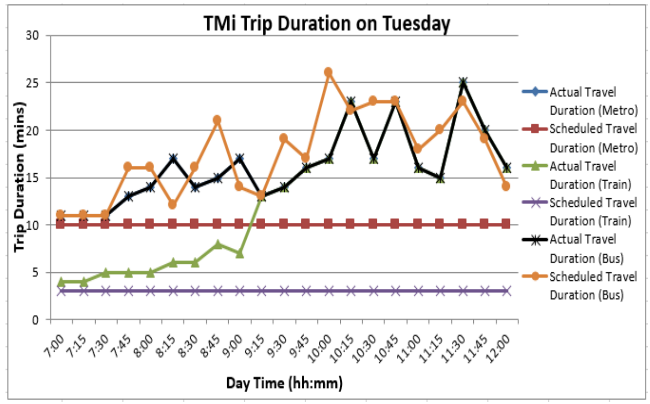

Fig. 4. Metro, Train, and Bus trip duration on Tuesday

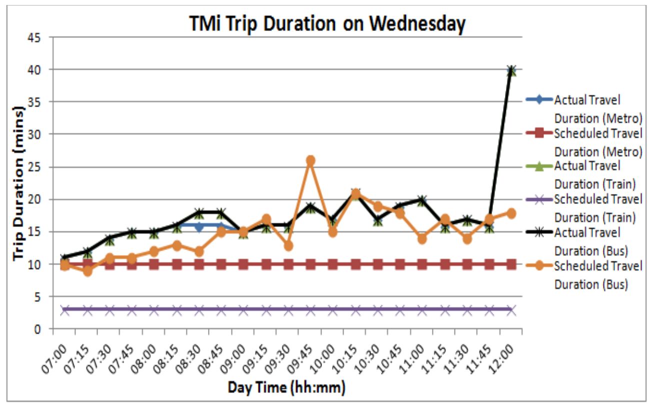

Fig. 5. Metro, Train, and Bus trip duration on Wednesday

B. The Decision System Algorithm

The information will be gathered in real-time from users in different means of transporting systems as feedbacks. This retrieval of information could be collected from each user from the phone using a specific very simple application. Then, it will feed the Decision System Algorithm *DSA(t)* as follows in Alg.1:

Alg.1 DSA(t): Algorithm Overview

For each TMi, collect the passengers feedback at their arrival stop
L1, L2, and L3 are 3 lists initialized by NULL
Q_i is initialized based on the previous knowledge of the traffic jam for every day and is rewarded based on the current time (t) by time interval "τ"
N_i is the available TMs at the current time (t)
if "$N_i \neq 0$" **then**
  for $j \in N_i$ **do**
    at time (t) there is a comparison between the TMi as follows:
    if "Transport_means (j) ≠ 0 and (reward (j) > reward(i))" **then**
      $L1 \leftarrow j$
    end if
    if "Transport_means (j) = 0 and (reward (j) > reward (i))" **then**
      $L2 \leftarrow j$
    end if
    if "Transport_means (j) ≠ 0 and (reward(j) < reward (i)" **then**
      $L3 \leftarrow j$
    end if
  end for

```
if "L1 ≠ 0 & L2 ≠ 0" then
        The next Transport_mean is chosen from L1 of the largest Q-value
end if
   if "L1 = 0 & L2 ≠ 0" then
        The next Transport_mean is chosen from L2 of the largest Q-value
end if
if "L1 = 0, L2 = 0, & L3 ≠ 0" then
        The next Transport_mean is chosen from L3 of the largest Q-value
end if
else
    /* N_i is empty */
    At time (t) a negative reward is generated for that TMi
end if
at time (t+1) the reward for every TMi is computed after the interval "τ" based on step (2)
at time (t) the Q-value is updated based on Eq. (1)
```

IV. Performance Validation

In this section, we investigate the performance of our proposed TDS model in terms of delays and rewards. We carried out the evaluation using Matlab. The simulation model is built according to the following specifications:

- The case study is the trip between Gare-du-Nord and Chatelet (Paris-France).
- Transport means available are metro, train, and bus.
- The days studied are three contagious days (Monday, Tuesday, and Wednesday) as a sample case study.
- The day's hours studied from 7:00am until 12:00pm quarterly.
- The default comfortability coefficient for the metro, train, and bus are 0.6, 0.8, and 0.5 consequently. This includes the capacity, design satisfaction, periodicity, and flexibility of using each transport mean.
- TMi cost is considered unity and it value is redeemed relatively between the transport means (Metro, Train, and Bus). The TMi costs are 0.7, 1, and 0.7 for the Metro, Train, and Bus consequently.
- Trip duration Traces are also collected from Sytadin traffic monitoring site [17] for private routes.

Simulation 1: the first simulation highlights the traffic situation study in the first day (Monday). We simulated the busiest hours of the day per quarter. The trip duration of using each mentioned transport mean is shown in Fig.3. From the figure, using buses has the largest delay compared to metros and trains at the same day hours. Fig. 6 shows the transport means Q-value of our proposed TDS model. A higher Q-value represents higher performances for the concerned transport mean. As shown in the figure, metro is recommended to be used along all the studied hours and then train. Buses are not recommended in this period.

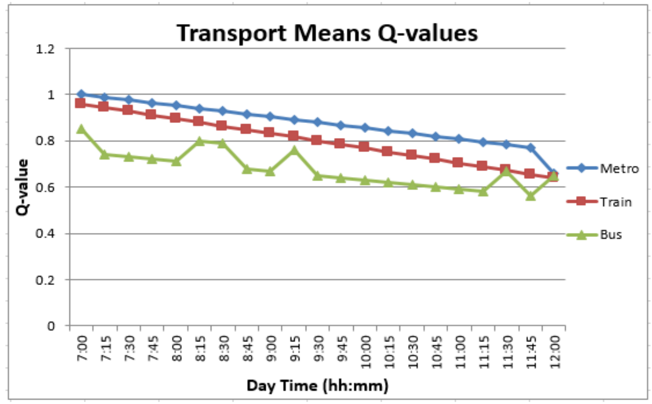

Fig. 6. Traffic means Q-value on Monday

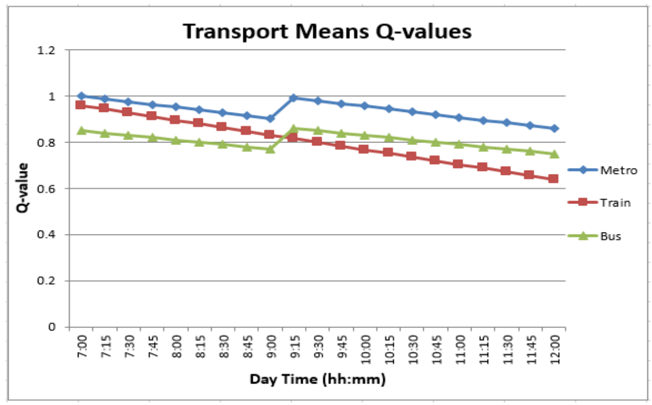

Fig. 7. Traffic means Q-value on Tuesday

Simulation 2: In this second part of the evaluation study, we are interested on another important metric, which is the delay to reach the final destination. The trip duration of using each mentioned transport mean on second day (Tuesday) is shown in Fig. 4. As illustrated by this figure, buses have also in this scheme the largest delay compared to metros and trains at the same day hours since roads at this time of day are greatly affected by traffic congestion. Fig.7 shows the transport means rewards of our proposed TDS model. Metro achieves the highest reward and hence it is recommended to be used along all the studied hours and then train. The recommendation of using train and bus are fluctuated in the middle of the studied hours.

Simulation 3: On the third day (Wednesday), the transport means have the trip duration in minutes as shown in Fig. 5. According to that, Fig. 8 shows the transport means rewards of our proposed TDS model. Metro has the highest reward and hence it is recommended to be used along all the studied hours and then train as in the previous simulations. Moreover, train and bus have the second recommendation until the time 8:15. Then, train recommendation decreases to the third category after that time.

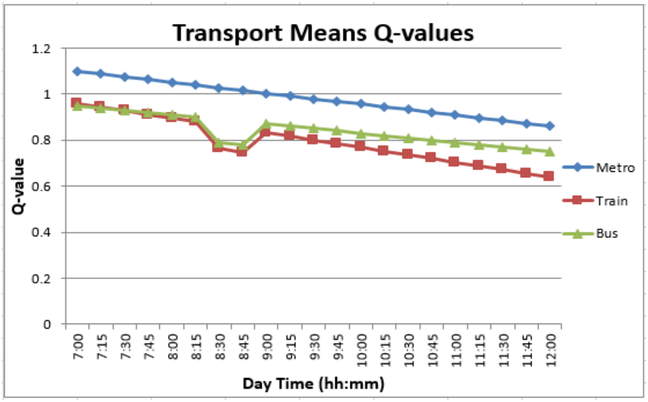

Fig. 8. Traffic means Q-value on Wednesday

The simulations results show that the metro has the best reward in most days' hours. This is because metro doesn't suffer of congestion like buses on roads. Also, it has a lower tariff than trains for the same source and destination stations and it is more frequent than buses and trains. If the traffic conditions changed, the result will be different for any reason or event such as weekends, holidays, employees' strikes, which occur frequently in France, etc.

Moreover, the simulation results can be modified according to the proposed passengers' satisfaction shown in Fig.2. This observation confirms the dynamicity of our model that holds for any transportation model and it can be modified to fit new traffic environment.

V. Conclusion

This paper proposed a new multi-modal traffic management solution that can be applied in ITS systems. The proposed algorithm considered the Q-Learning technique for learning enforcement by time and based on the gathered information in real time from the passengers crossing stations. Moreover, this decision aims to build a smart transportation station and a real time passengers guide using Markovian model. This latter was tested and simulated using Matlab and fed by real statistics gathered from real traffic sites in France. Simulations results corroborate the efficiency of our model in term of giving the travelling passenger the best transport mean.

For our future research work, we will focus on a three dimension modeling approach where we add the time to the user-location based on user-mobile statistics in order to build a new graph for multi-modal traffic behavior. This future enhancement will construct a more complete and universal traffic management framework.

References

[1] ETSI EN 302 665, "Intelligent Transport Systems (ITS);Communications Architecture" V1.1.1, Sept. 2010.

[2] A. K. Runyoro, I. Zlotnikova, J. Ko, "Towards automated road information framework - A case study of Tanzania," 13th International Conference on Control, Automation and Systems (ICCAS), doi: 10.1109/ICCAS.2013.6703906, Gwangju, pp. 267 – 272, Oct. 2013.

[3] R. Meier et al., "Towards Delivering Context-Aware Transportation User Services," IEEE Intelligent Transportation Systems Conference (IEEE ITSC 2006), Toronto, doi: 10.1109/ITSC.2006.1706769, pp. 369-376, Sept. 2006.

[4] Transport & Traffic Tweets: http://www.twitter.com, last seen Jun 2014.

[5] SNCF Transilien: http://www.transilien.com, last seen June 2014.

[6] S.J. Russel and P. Norvig, "Artificial Intelligence: A Modern Approach", Third Edition, Prentice Hall, 2010.

[7] D. Faganand R. Meier"Intelligent time of arrival estimation," IEEE Forum on Integrated and Sustainable Transportation System (FISTS), vol., no., Vienna, doi: 10.1109/FISTS.2011.5973595, pp.60 - 66, June 2011.

[8] S. El-Tantawy*et al.*,"Multiagent Reinforcement Learning for Integrated Network of Adaptive Traffic Signal Controllers (MARLIN-ATSC): Methodology and Large-Scale Application on Downtown Toronto", IEEE Transactions on Intelligent Transportation Systems, vol.14, no. 3, pp. 1140 – 1150, Sept. 2013.

[9] S. El-Tantawyand B. Abdulhai"An agent-based learning towards decentralized and coordinated traffic signal control," 13th International IEEE Conference on Intelligent Transportation Systems (ITSC), Funchal, doi: 10.1109/ITSC.2010.5625066,pp. 665–670, Sept. 2010.

[10] I. Dusparic and V. Cahill "Distributed W-Learning: Multi-Policy Optimization in Self-Organizing Systems," Third IEEE International Conference on Self-Adaptive and Self-Organizing Systems SASO '09, San Francisco,pp. 20–29, Sept. 2009.

[11] J. Song andZ. Jin "Q-learning based multi-intersection traffic signal control model," International Conference on System Science, Engineering Design and Manufacturing Informatization (ICSEM), doi: 10.1109/ICSSEM.2011.6081298, vol. 2, pp. 280–283,Oct. 2011

[12] J. Zhang*et al.*, "Data-Driven Intelligent Transportation Systems: A Survey," IEEE Transactions on Intelligent Transportation Systems, vol. 12, no. 4, pp. 1624–1639, Dec. 2011.

[13] J.Songet.al, "Implementing traffic signal optimal control by multiagent reinforcement learning," International Conference on Computer Science and Network Technology (ICCSNT), Harbin, vol. 4, doi: 10.1109/ICCSNT.2011.6182495, pp. 2578–2582, Dec. 2011.

[14] S. El-Tantawy and B. Abdulhai, "Multi-Agent Reinforcement Learning for Integrated Network of Adaptive Traffic Signal Controllers (MARLIN-ATSC)", 15thInternational IEEE Conference on Intelligent Transportation Systems (ITSC), Anchorage, doi: 10.1109/ITSC.2012. 6338707, pp. 319 – 326, Sept. 2012.

[15] A. Salkham and V. Cahill "Soilse: A decentralized approach to optimization of fluctuating urban traffic using Reinforcement Learning," 13th International IEEE Conference on Intelligent Transportation Systems (ITSC), Funchal, doi: 10.1109/ITSC.2010. 5625145, pp: 531–538, Sept. 2010.

[16] M. Abdooset. al, "Traffic light control in non-stationary environments based on multi agent Q-learning," 14thInternational IEEE Conference on Intelligent Transportation Systems (ITSC), Washington, doi: 10.1109/ITSC.2011.6083114,pp:1580–1585, Oct. 2011.

[17] Sytadin traffic en temps réel: http://www.sytadin.fr/, last seen Dec. 2014