Journal of Siberian Federal University. Engineering & Technologies, 2019, 12(1), 113-125
~ ~ ~

УДК 004.8

Energy Efficient Dynamic Bid Learning Model for Future Wireless Network

Abdulkarim Oloyede*[a], Nasir Faruk[a], Lukman Olawoyin[a] and Olayiwola W. Bello[b]
[a]*Department of Telecommunication Science*
University of Ilorin
Ilorin, Nigeria
[b]*Department of Information and Communication Science*
University of Ilorin
Ilorin, Nigeria

Received 19.09.2017, received in revised form 30.01.2018, accepted 18.03.2018

In this paper, an energy efficient learning model for spectrum auction based on dynamic spectrum auction process is proposed. The proposed learning model is based on artificial intelligence. This paper examines and establishes the need for the users to learn their bid price based on information about the previous bids of the other users in the system. The paper shows that using Q reinforcement learning to learn about the bids of the users during the auction process helps to reduce the amount of energy consumed per file sent for the learning users. The paper went further to modify the traditional Q reinforcement learning process and combined it with Bayesian learning because of the deficiencies associated with Q reinforcement learning. This helps the exploration process to converge faster thereby, further reducing the energy consumption by the system.

Keywords: Q Reinforcement Learning, Spectrum Auction, Dynamic Spectrum Access, Bayesian Learning.

Citation: Oloyede A., Faruk N., Olawoyin L., Bello O.W. Energy efficient dynamic bid learning model for future wireless network, J. Sib. Fed. Univ. Eng. technol., 2019, 12(1), 113-125. DOI: 10.17516/1999-494X-0035.

* Corresponding author E-mail address: Oloyede.aa@unilorin.edu.ng, faruk.n@unilorin.edu.ng

Энергоэффективная обучаемая модель динамической ставки для будущей беспроводной сети

Абдулкарим Олоиеде[a], Насир Фарук[a], Лукман Олавоин[a] и Олаивола В. Белло[б]
[a]Департамент телекоммуникационных наук
Университет Илорин
Илорин, Нигерия
[б]Департамент информационных и коммуникационных наук
Университет Илорин
Илорин, Нигерия

В статье представлена энергоэффективная обучающая модель аукциона частот, основанная на его динамическом процессе. Предложенная модель обучения базируется на искусственном интеллекте. Рассмотрены требования по установлению цены предложения на основе информации о предыдущих ставках других пользователей в системе. Применяя Q-обучение, можно сократить количество потребляемой энергии за файл, отправленный для пользователей обучения. Описаны изменения традиционного процесса Q-обучения и объединение его с Байесовским обучением из-за недостатков Q-обучения. Это помогает ускорить процесс поиска, тем самым уменьшая потребление энергии системой.

Ключевые слова: Q-обучение, аукцион частот, динамический доступ к спектру, Байесовский вывод.

I. Introduction

The conventional fixed spectrum allocation, which wastes the spectrum resources, is one of the main causes of congestion on the radio spectrum [1-3]. This scheme was widely accepted and used because to an extent, the fixed spectrum allocation scheme prevents interference from others to whom the frequency band is not allocated [2]. Previously, this method of allocation worked perfectly however, due to an increase and variation in the demand for the use of the radio spectrum, this scheme is leading to "*artificial spectrum scarcity*" [4]. This is because the paucity of the radio spectrum depends on location (space) and the time of the day (time) [5-7]. In addition to this, fixed spectrum allocation creates spectrum holes also known as white space thereby, degrading the spectral efficiency of the radio spectrum. These disadvantages among others led to the concept of Dynamic Spectrum Access (DSA) as proposed in [8-10]. Based on the concept of DSA, the author of [11] also proposed an auction model as a fair process that can be used in accessing the radio spectrum in an opportunistic manner. Spectrum auction is quite attractive because it allows spectrum holders to share their underutilised spectrum band for economic benefit [12, 13]. Furthermore, the use of dynamic spectrum allocation in combination with spectrum auction for short term allocation of the radio spectrum allows both the user and the wireless service provider to gain market knowledge about the radio spectrum [14]. Spectrum auction has been previously proposed in [15-18] however, some crucial problems such as the all-important energy efficiency are still ambiguous in most of the existing spectrum auction models.

As a result of the effects of climate change, which is necessitating energy conservation, energy efficiency is quite crucial for any future wireless network. Hence, this work proposes an energy efficient auction model for future wireless network. The idea behind the proposed model is an automated auction process, which would not always require much input from the users such as the proposed auction model in [19, 20]. Furthermore, in the proposed model in [21], which proposed an auction process for DSA, it was evidently shown that the use of software would be an advantage and it should be adopted in generating the bids during an auction process. This paper proposes the use of learning algorithms to aid the auction process. This is necessary because, in the proposed model in [22, 23], the auctioneer assumed a prior knowledge of the bid equation. However, these assumptions cannot always be made in real systems, because the auctioneers are most likely going to be unaware of the number of the interested bidders until the process of bid submission has been accomplished. Therefore, such an assumption cannot always be generalised. In addition to this, this work adopted the use of an automated software system because most users lack the global knowledge of a wireless network such as the amount of radio resources needed. Therefore, in this work, we propose the use of artificial intelligence. In this model, the users learn how to bid effectively and be able to win the future bidding process using the information available after each auction process. This process is also referred to as machine learning. The use of Machine learning during an auction process helps in eliminating the additional delay that might be introduced into the system as a result of periodic the auction process in dynamic spectrum allocation. This is because the reality is that an auction process requires some time to process the submitted bids and to determine the winner(s). However, delay is an important metric in wireless communication. The delay tolerance of a system depends on the application. A data packet based communications system has more tolerance for delay when compared with the voice communications system.

In this work, we propose a model for DSA using a concept which is based on the first price auction and the use of a reserve price. The reserve price is introduced based on the advantages of the reserve price as outlined in [24]. At the start of an allocation period, a first priced auction process is carried out. In this auction process, each of the user that wants to access the radio spectrum submit a bid and a number of users which is same as the number of channels that is available emerges as the winners. The auction process is explained in more details in the modelling section. Hence, it is a multi-winner auction process. We assume two types of users; those that bid myopically are one group of users while the other group of users are the learning users. A learning user learns the appropriate bid that is above the set reserve price, which is set by the Wireless Service Provide (WSP) using artificial intelligence. Myopic bidders, however places a bid value randomly. The process of setting the reserve price is explained later in the modelling section. Two types of learning methods are examined in this work while examining the delay associated with each model. This paper also compares the amount of energy that is consumed when the users are learning with the non-learning users. This is examined based on Q reinforcement learning and then Bayesian learning used in conjunction with the Q reinforcement learning.

The rest of this paper is as organised as follows: Section II provides the related work. In section III, the energy model, utility function and learning models used in this work are introduced. In Section IV, the simulation model adopted in this work is explained. Results and discussion are provided in section V which is followed by the conclusions in the last section.

II. Background and Related Work

Reinforcement learning (RL) is a type of machine learning that allows software agents to learn the ideal behaviour within a context [25]. It uses feedback from the stochastic environment during the learning process to maximize performance by balancing between conflicting considerations [26]. Generally, RL process uses a form of additive reward and penalty. This reward and penalty are sometimes with a discount factor as proposed in [27]. Its learning process involves trials known as an exploration for the learning process to converge effectively before the process of exploration, which is the use of the learnt values. The learning agent usually uses the set of information learnt at time t based on an action a_t ($a_t \in A(S_t)$) to transit from one state (S) to another at time $t + 1$. A learning agent receives a reward (r_t) based on the action taken. The goal of the learning agent is to derive a maximum possible reward ($\pi : S \rightarrow A$) form the action taken. Where $A(S_t)$ and π is the number of available actions at time t respectively. Reinforcement learning was previously used in [28].

An auction process for dynamic spectrum access was considered in [15]. For a DSA model based on an auction process, learning of the appropriate bid value is rather essential than a user guessing a value because the load on the wireless system varies and the budget of users also varies therefore, if users can learn the appropriate bidding price based on the traffic load, then the users are able to maximise whatever the available budget is by bidding at periods only when the user can afford to pay. This helps to reduce the amount of the energy that is wasted when bids are rejected.

III. The Energy Model, Utility Function and Learning Process

A. The Energy Model

The energy model adopted in this work is represented using state 1 to 4 as shown in Fig. 1 and explained below:

1. A user with packets to send wakes up into the OFF state, such user subsequently submits a bit to the auctioneer.

2. A winning user changes into the ON state from the OFF state.

3. Such a winning user is in the ON state until after transmission or until when the packet of the user is dropped. Packet dropping is usually as a result of the bid being below the minimum price set by the auctioneer or due to poor channel conditions.

4. The transmission process is completed by the user moving back to the OFF state.

B. Utility Function

The utility function expresses the satisfaction derived by the user in placing a bid. The users are assumed to be price sensitive hence, it is in the best interest of the users to win with the least possible

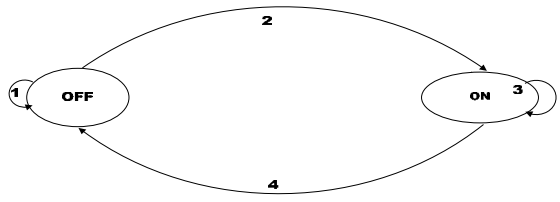

Fig. 1. Energy and system model as a two state Markov chain

amount. An exponential utility function is assumed due to its characteristics of an increase, which is rapid in nature. A user whose bid falls short of the reserve price has a utility with a value of zero while the winning user with the least value (b_{min}) has the highest utility value.

$$u_i(T)=\begin{cases} 2^{\frac{b_i}{b_{min}}}-1 & If\ a\ bidder\ wins \\ 0 & otherwise \end{cases}. \quad (1)$$

C. Q Reinforcement Learning

Reinforcement learning is one form of the several machine learning techniques used in wireless networks [1]. In reinforcement learning, the players involved use a wide range of methods to draw on the prior experience. The policy adopted by the player usually affects the converging speed at which the users learn. Q reinforcement learning, is a kind of reinforcement learning which assigns values to pairs of the action state which is usually used to model Markov decision processes (MDP). Every state has a number of possible actions and the reward earned depends on the action taken (Scalar reward). The reward is based on how close the action is to the best action. Therefore, a temporal difference is used to estimate the optimal value. Take for example, in an auction process with reserve price where users want to win with the least possible amount. The closer the user's bid is to the reserve price the higher the user's reward. In a bidding process with reserve price the acceptable values of the winning bids can be any value till above the reserve price and a difference in price unit as small as 0.01 can make a difference between a winning bid and the loosing bid.

The Q reinforcement learning approach used in this work is similar to [10]. This work assumes that the only information held by the users is the information about their own bid history. They do not have the bid history of the other user in the system. It is also assumed that a user can only submit one bid in a bidding round. We also use a first price sealed bid auction with a reserve price. It is assumed that the users in the system have no prior knowledge of the reserve price. This is determined by the WSP according to the process explained in [29].

In the proposed model, a bidding user submits a bid to the spectrum broker who is also the service provider. Based on the bid submitted, the user obtains a utility value using the utility equation (1). It is assumed in this paper that users want to win the auction process with a high value of utility. however, the closer the user's bid is to the reserve price the lesser the probability that the user wins the bid therefore, the challenge faced by the bidder is to decide the appropriate value of utility that allows the bidder to be among the winning bidders.

The objective of the learning user is to obtain and optimal π^* that maximises the total expected discounted reward given by

$$V^{\pi} = E\left\{\sum_{t=0}^{\infty} \beta^t r\big(s_t, \pi(s_t)\big) | s_o = s\right\}. \quad (2)$$

Where E is the expectation operator, $0 \leq \beta < 1$ is a discounted factor and π is a policy $S \rightarrow A$ For a policy π a Q value is defined as the expected discounted reward for executing a at state s and the following policy π is given as:

$$Q^{\pi}(s,a) = R(s,a) + \beta \sum_{s^*} Pr_s, s^*(a) V^{\pi}(s^*). \quad (3)$$

Where $R(s, a)$ is the old value of the Q value. It can be clearly seen from equation (2) to equation (3) that in order to obtain the optimal policy π^*, the information of s is vital. In this work, in order to simplify the learning process, we assume that users can only pick a specific bid value form a bin and each bin is associated with a traffic load. Each bin has a range of values as explained further in the modelling section. We associate with a traffic load because the reserve price depends on the traffic load in the system and the reserve price determines if the bid is accepted or rejected. Ideally in a real world the bid value in each of the bins could be associated with the quality of service that the user expects from the WSP. During an exploration period, each user explores a range of available bid values and keep a record as shown in Eq.(4)

$$B_p = (b_j, N_t, u, N_w, W_{\%}, N_m\ ). \tag{4}$$

Where b_j is the j^{th} value of the bid in the p^{th} bin, N_t is the number of trial with a maximum value of τ, u is the utility derived from the bid value using equation (3), N_w is the number of times the user used the bid value and wins, $W_{\%}$ is the winning percentage of the user when using the specific value and N_m is the multiplication of $W_{\%}$ by u. N_m is then subtracted from N_m^*. Where N_m^* is the optimal reward function. The final reward used in updating the Q table is given as

$$\begin{array}{cccccc} P/J & 1 & 2 & 3 & 4 \ldots & j \\ 1 & x_{11} & x_{12} & x_{13} & x_{14} \ldots & x_{1j} \\ 2 & x_{21} & x_{22} & x_{23} & x_{24} & x_{2j} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ p & x_{p1} & x_{p2} & x_{p3} & x_{p4} & x_{pj} \end{array} \tag{5}$$

Where x_{pj} the value of N_m of bid j in bin p. During the Q exploration period, the user first try all the available values in all the bins by setting τ to a low number and the user does not try bins in which all the available range of values gives $W_{\%}$ of zero. This is to reduce the exploration period. Take for example, when $j = 5$ and $p = 5$. In the tuple below, where the reserve price is assumed to be varying between 4 and 4.5.

$$P_1 = [1\ ,\ 1.5, 2, 2.5, 3]\ P_2 = [3.5, 4, 4.5, 5, 5.5]. \tag{6}$$

At the beginning of the exploration period when τ is set to 10% of maximum τ. The N_m in the record for P_1 are all be zero while some of P_2 would be zero and others would have different percentages based on the reserve price. In such scenario, the user would not further explore the values in P_1.

The Q Reinforcement Learning Algorithm

1: Users pick a bid value randomly from the bin;

2: The utility function of the user based on the bid is calculated;

3: Other record is also calculated e.g $W_{\%}$ and $N_{m;}$

4: After 10% of τ trials, the user would try from only bids from bins that are more than 20% and has a value of $W_{\%} > 20$;

5: After τ trials the user picks the value that gives max $W_{\%}$.

D. Bayesian framwork for Rainforcement Learning

Delay is a critical factor in a communication network, depending on the application. The delay associated with random exploration in Q reinforcement learning sometimes leads to a convergence

point that is not optimal. A non-optimal convergence point can be arrived at if the exploration period is not long enough. Hence, there is a need to enhance the exploration process in Q learning. One method that could be used in enhancing Q learning approach is the Bayesian learning. In Bayesian learning algorithm, the learning agent makes a decision based on the most likely event that is going to happen using prior experience. This allows for a faster and smoother movement from exploration behavior to exploitation behavior. Bayesian learning uses the Bayes' theorem in the exploration stage as shown below.

$$P(A\backslash B) = \frac{P(B\backslash A)\times P(A)}{P(B)}. \tag{7}$$

Where $P(A)$ is the prior probability distribution of hypothesis A, the $P(B)$ is the prior probability of the training data B and $P(B\backslash A)$ is the probability of A giving B. Bayes learning is appropriate for bid learning because usually a commodity on sale would have a guide price and buyers offers a bid not too far away from the guide price. It also offers several advantages over the Q learning model. Firstly, it provides an ideal format to reach a compromise between the exploration and the exploitation stage by giving information on states in which the player might not have explored. Secondly, it allows an incorporation of the prior determination of the user's transition. However, it has a disadvantage, which is, it can only be applied when prior information is available.

The prior knowledge used in this work is generated by the system based on previous experience. This is similar to what is adopted in [11]. We assume that the transition matrix is sparse, as only a certain number of the bid values in the bins are non-zero. Similar assumption was made in [11]. The posterior distribution represents the uncertainty in the system, for the transition probability for each of the state action pair, as explained in the Q reinforcement learning.

The Bayesian Learning Algorithm

1: Prior probability is given for all the possible state;
2: Q reinforcement learning is carried out;
3: Bayes rule is applied;
4: Users pick the action with the highest utility and winning percentage.

IV. Modelling

In order to evaluate the performance of the proposed learning algorithms, we model a bidding scenario using a multi winner first price sealed bid auction with reserve price similar to [4], where the spectrum is allocated dynamically. We model an uplink scenario with one WSP and N users and M available channel in each of the cells. The users require a transmit channel in order to transmit data files via a central base station. We assume the users are connection oriented and the user to whom the channel is allocated uses the channel for a fixed period of time, before releasing it and another auction round is carried out. We assume a Poisson arrival process with z users out of the N possible users requesting for a channel by submitting a bid (b_1, b_2, b_3 … b_z) and the value of z varies depending on the traffic load on the system. The bid value is chosen randomly from the possible bid value in the bin. Based on the bids submitted, K winners which is the number of available channels in the system at time T emerges.

$$K \subseteq z \subseteq \mathrm{N}. \tag{8}$$

The user is allocated the channel, provided the value bided is above the reserve price. The reserve price is dynamic depending on the traffic load and it is calculated by the WSP as shown below.

$$r = \frac{C_f}{M}, \tag{9}$$

$$C_f = \frac{z}{K}. \tag{10}$$

Where M is the total number of channels in the system. After submitting the bid the learning process as explained earlier begins. We assume only one of the user is learning to win the bid with the minimum possible amount and others are just bidding myopically. The system flow chart is as shown below (Fig. 2).

V. Results and Discusion

The parameters used in the modelling is given in table 1. The reserve price is set in such a way that it varies within a bin depending on the traffic load e.g when the traffic load is 2 or 4 the reserve price can only vary from 0.35 to 0.45 and 0.45 to 0.55 respectively. We also assume all the users are transmitting at the same power and transmission rate.

Fig. 3 shows the value of N_m in percentage when the traffic load on the system is 4 Erlang and τ is initially set to 500 in then 2000 in Fig. 3(a) and 3(b) respectively. When τ is 500 it can be seen that using a bid value of 50 price unit gives the best value of N_m. However, this is not the optimal converging value as it can be seen that when, τ was further increased to 2000 the optimal value changes to 5. With a low value of τ, the learning does not converge at the best value because, utility function and bid acceptance value are like opposing each other. A user bidding the highest value wins, but has a close

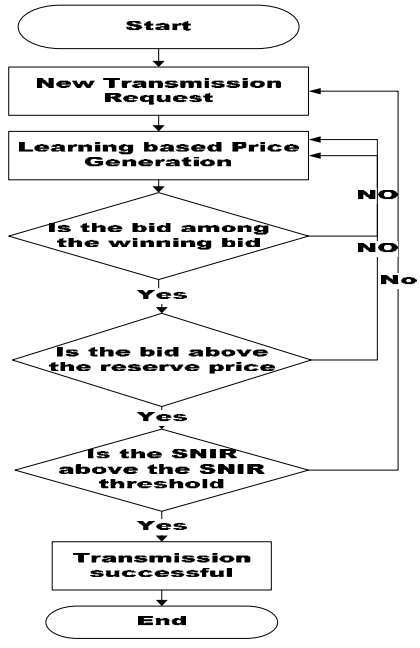

Fig. 2. System flow Chart

Table 1. Parameters Used

Parameter	Values
Cell radius	2 *km*
$SNIR_{Treshold}$	1.8 *dB*
Interference threshold	-40 *dBm*
Users(N)	50
File Size	2 *Gbits*
Frequncy Reuse Factor	3
Hight Of Base Station	15 *m*
Bin 1	[0.30-0.35]
Bin 2	[0.35-0.45]
Bin 3	[0.40-0.50]
Bin 4	[0.45-0.55]
Bin 5	[0.55-0.65]

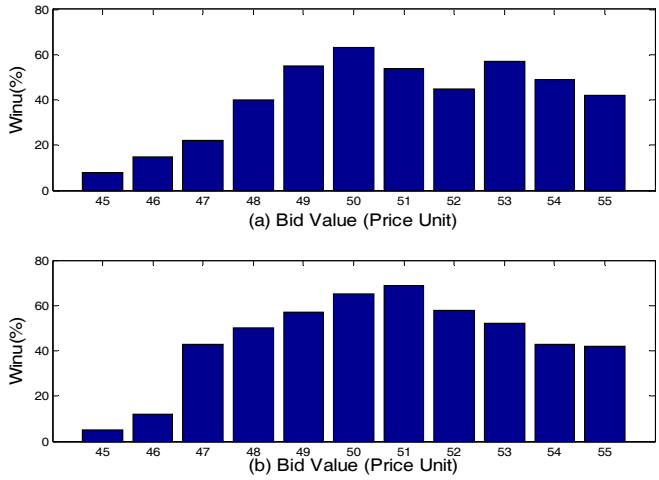

Fig. 3. Bid Value for Q Reinforcement Learning with 500 Events and 2000 Events

to zero utility value, thereby having a low value of $W_{\%}$. However, as more trial is carried out, this effect balances each other out. It is difficult to know the appropriate value to set τ too but our results showed that when using Q reinforcement learning the value of τ must be sufficiently large enough to allow for optimal convergence.

However, using the Bayesian approach with reinforcement learning allows for faster convergence as it can be seen from Fig. 4 that the optimal convergence point is same for when τ is 500 and 2000. This shows that the Bayesian approach to reinforcement learning converges faster than traditional Q reinforcement learning.

We observe the system performance in terms of the amount of energy consumed per file sent by the learning user and the users that are not learning. Fig. 4(a) shows the results for the warm up stage and Fig. 4(b) for steady state scenario. The warm-up scenario involves both exploration and exploration

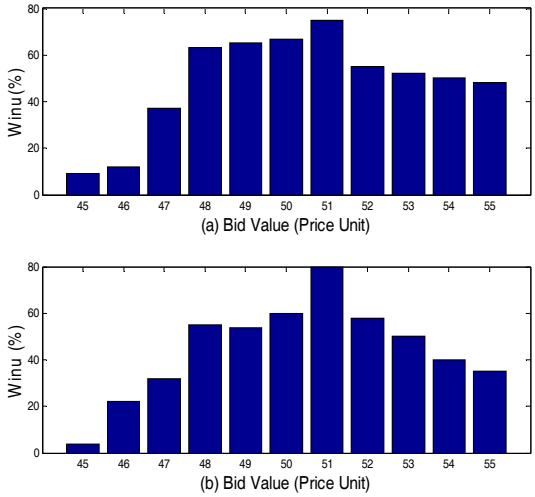

Fig. 4. Bayesian Learning for 500 Events and 2000 Events

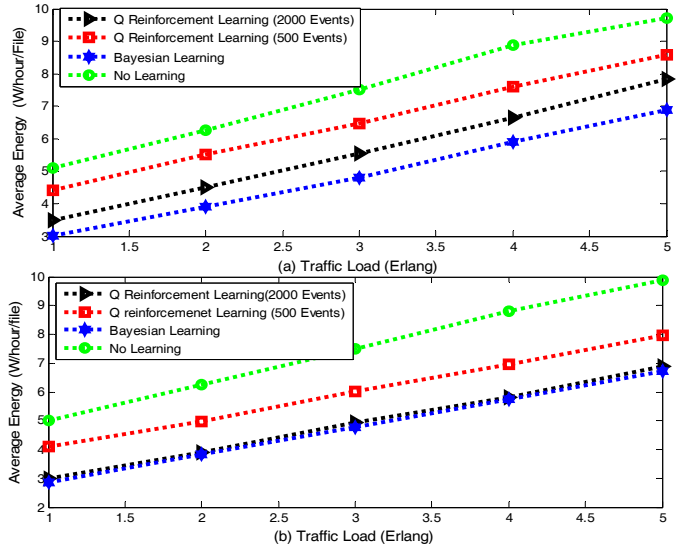

Fig. 5. Average Energy consumed for the Learning User

while in the steady state the users have finished learning so it involves only exploitation (output from the exploration stage are discarded).

From Fig. 5, it can be seen that the average energy consumed increases with the traffic load in the system because, as the traffic load increases the reserve price and collusion in the system also increases thereby increasing the average energy per file sent. Fig. 5(a) presents the warm up results. Using Q reinforcement learning with 2000 events consumed more energy than 500 events because of the increases in exploration value as it can be seen from 5(b) when the steady state is reached that 2000 events of Q reinforcing learning performs just as good as Bayesian learning but Q learning with 500 events consumes more energy because it does not converge at the best value.

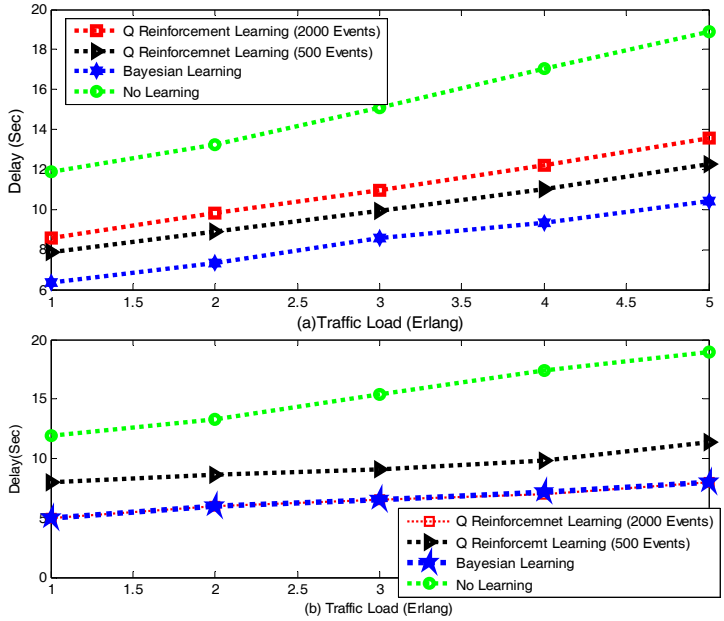

Fig. 6. Average delay for Learning User

Fig. 6 shows the average delay per file sent, with 6(a) being the results from the warm-up stage while Fig. 6(b) is for the steady state. It can be seen that if 2000 events is carried out, Q reinforcement learning perform just as well as Bayesian learning in the exploitation stage, but if we consider the warm-up process, Q reinforcement learning introduces more delay into the system.

VI. Conclusions

This paper examines the use of learning based on the auction process for dynamic spectrum access. The paper used the Bayes rule to bias the exploration towards the most promising state, thereby reducing the cost of exploration. The results shows that the use of the Bayes rule helps in reducing the amount of energy consumed, delay and the convergence speed reduction. Q reinforcement learning used in this work is effective and flexible if a large number of trials can be performed, however much faster learning process can be obtained if the reinforcement learning is combined with Bayesian learning.

References

[1] Nissel R. and Rupp M. Dynamic Spectrum Allocation in Cognitive Radio: Throughput Calculations, in *IEEE International Black Sea Conference on Communications and Networking (BlackSeaCom), Varna, Bulgaria*, 2016.

[2] Patel H., Gandhi S. and Vyas D.A *Research on Spectrum Allocation Using Optimal Power in Downlink Wireless syste*m, 2016.

[3] Cao B., Zhang Q. and Mark J.W. Conclusions and Closing Remarks, in *Cooperative Cognitive Radio Networking*, ed: Springer, 2016, 95-97.

[4] Freyens B.P. and Alexander S. Policy objectives and spectrum rights for future network developments, *Dynamic Spectrum Access Networks (DySPAN), 2015 IEEE International Symposium on*, 2015, 229-240.

[5] Bhattarai S., Park J.-M. J., Gao B., Bian K. and Lehr W. An Overview of Dynamic Spectrum Sharing: Ongoing Initiatives, Challenges, and a Roadmap for Future Research, *IEEE Transactions on Cognitive Communications and Networking,* 2016, 2, 110-128.

[6] Abdelraheem M., El-Nainay M. and Midkiff S. F. Spectrum occupancy analysis of cooperative relaying technique for cognitive radio networks, *Computing, Networking and Communications (ICNC), 2015 International Conference on*, 2015, 237-241.

[7] Mahajan R. and Bagai D. Improved Learning Scheme for Cognitive Radio using Artificial Neural Networks, *International Journal of Electrical and Computer Engineering (IJECE),* 2016, 6, 257-267.

[8] Zhu J. and Liu K.J.R. Cognitive Radios For Dynamic Spectrum Access – Dynamic Spectrum Sharing: A Game Theoretical Overview, *Communications Magazine, IEEE,* 2007, 45, 88-94.

[9] Zhu J. and Liu K. J. R. Multi-Stage Pricing Game for Collusion-Resistant Dynamic Spectrum Allocation, *Selected Areas in Communications, IEEE Journal on,* 2008, 26, 182-191.

[10] Subramanian A.P., Al-Ayyoub M., Gupta H., Das S.R. and Buddhikot M.M. Near-Optimal Dynamic Spectrum Allocation in Cellular Networks, *New Frontiers in Dynamic Spectrum Access Networks, 2008. DySPAN 2008. 3rd IEEE Symposium on*, 2008, 1-11.

[11] Jia J., Zhang Q., Zhang Q. and Liu M. *Revenue generation for truthful spectrum auction in dynamic spectrum access*, presented at the Proceedings of the tenth ACM international symposium on Mobile ad hoc networking and computing, New Orleans, LA, USA, 2009.

[12] Boukef N., Vlaar P.W., Charki M.-H. and Bhattacherjee A. Understanding Online Reverse Auction Determinants of Use: A Multi-Stakeholder Case Study, *Systèmes d'information & management,* 2016, 21, 7-37.

[13] Nehru E.I., Shyni J.I.S. and Balakrishnan R. Auction based dynamic resource allocation in cloud, *Circuit, Power and Computing Technologies (ICCPCT), 2016 International Conference on*, 2016, 1-4.

[14] Pilehvar A., Elmaghraby W.J. and Gopal A. Market Information and Bidder Heterogeneity in Secondary Market Online B2B Auctions, *Management Science,* 2016.

[15] Hyder C.S., Jeitschko T.D. and Xiao L. Bid and Time Strategyproof Online Spectrum Auctions with Dynamic User Arrival and Dynamic Spectrum Supply, *Computer Communication and Networks (ICCCN), 2016 25th International Conference on*, 2016, pp. 1-9.

[16] A. Gopinathan, N. Carlsson, Z. Li, and C. Wu, «Revenue-maximizing and truthful online auctions for dynamic spectrum access,» in *2016 12th Annual Conference on Wireless On-demand Network Systems and Services (WONS)*, 2016, 1-8.

[17] Khaledi M. and Abouzeid A. Optimal Bidding in Repeated Wireless Spectrum Auctions with Budget Constraints, *arXiv preprint arXiv:1608.07357,* 2016.

[18] Lai W.-H., Polacek P. and Huang C.-W. A Posted-Price Auction for Heterogeneous Spectrum Sharing under Budget Constraints, *proceedings of the 9th EAI International Conference on Bio-inspired Information and Communications Technologies (formerly BIONETICS) on 9th EAI International*

Conference on Bio-inspired Information and Communications Technologies (formerly BIONETICS), 2016, 320-323.

[19] Fraser S.A., Lutnick H. and Paul B. Automated auction protocol processor, ed: Google Patents, 1999.

[20] Mori M., Ogura M., Takeshima M. and Arai K. Automatic auction method, ed: Google Patents, 2000.

[21] Oloyede A. and Grace D. Energy Efficient Soft Real Time Spectrum Auction for Dynamic Spectrum Access, presented at the 20th International Conference on Telecommunications Casablanca, 2013.

[22] Oloyede A. and Dainkeh A. Energy efficient soft real-time spectrum auction, *Advances in Wireless and Optical Communications (RTUWO)*, 2015, 113-118.

[23] Yin J., Shi Q. and Li L. Bid strategies of primary users in double spectrum auctions, *Tsinghua Science and Technology,* 2012, 17, 152-160.

[24] Vincent D. R. Bidding off the wall: Why reserve prices may be kept secret, *Journal of Economic Theory,* 1995, 65, 575-584.

[25] Abel D., MacGlashan J. and Littman M.L. Reinforcement Learning As a Framework for Ethical Decision Making, *Workshops at the Thirtieth AAAI Conference on Artificial Intelligence*, 2016.

[26] Patel M.A.B. and Shah H.B. Reinforcement Learning Framework for Energy Efficient Wireless Sensor Networks, 2015.

[27] Jiang T. Reinforcement Learning-based Spectrum Sharing for Cognitive Radio, PhD thesis, Department of Electronics, Univeristy of York, 2011.

[28] Lorenzo B., Kovacevic I., Peleteiro A., Gonzalez-Castano F.J. and Burguillo J.C. Joint Resource Bidding and Tipping Strategies in Multi-hop Cognitive Networks, *arXiv preprint arXiv:1610.02826,* 2016.

[29] Oloyede A. and Grace D. Energy Efficient Bid Learning Process in an Auction Based Cognitive Radio Networks, *Paper accepted in Bayero Univeristy Journal of Engineering and Technology(BJET),* 2016/02/02, 2016.