

A Novel Approach of Cognitive Base Station with Dynamic Spectrum Management For High-speed Rail *

Qingting Wu, Yiming Wang, Zhijie Yin, Hongyu Deng, Cheng Wu[†]
School of Urban Rail Transportation, Soochow University, Suzhou, China

Abstract

The characteristic of fast movement in high-speed rail seriously affects the stability of vehicular wireless communication. Applying cognitive technology to individual users often brings frequent channel switch and inefficient blind learning. To address these issues this paper proposes a novel concept of Cognitive Base Station (CBS), which has the capability of forecasting spectrum holes and assigning spectrum to individuals. We then give the model of cognitive base station and evaluate the performance in our simulation platform within high-speed rail environment. The experiment results further prove that the model can significantly improve the performance of vehicular communication.

1 Introduction

With the development of era, the demand for rail transit is rapidly increasing. When travelling on train, the passengers always hope to enjoy better communication quality and faster data access service. European Rail Traffic Management System (ERTMS) is a revolution in railways to guarantee the communication, which is consist of European Train Control System (ETCS) and a mobile-communications network optimized for railways called GSM-R.

GSM-R is the Global System for Mobile Communications-Railway in the worldwide and is dedicated to provide the bidirectional radio bearer for the train signaling systems, which operates in a 4MHz band (876-880 MHz for uplink and 921-925 MHz for downlink) [Sniady and Soler, 2012]. It is possible to divide the authorized band into 19 channels of 200KHz width in each GSM-R group. The rail line is covered with GSM-R groups and each consists of many GSM-R cells. A single GSM-R cell can use only few of the channels in a round robin manner, because the same channel cannot be reused by neighboring cells due to interference. Each cell is equipped with a base station. The base station is made up of building baseband unit (BBU) and radio remote unit (RRU). RRU is

always deployed outside along the railway and BBU is inside. One BBU is connected to multiple RRUs. BBU and RRU are used to process baseband signal and radio frequency signal, respectively. To ensure the communication between RRU and passengers, two vehicular stations (VS) are installed on the top and final carriages of the train. The network architecture is illustrated in Fig. 1 [Isheng Zhao *et al.*, 2013], [Tian *et al.*, 2012]. The GSM-R system consists of base transceiver stations (BTS) along the railway lines and embedded GSM-R mobiles connected to antennas on the roof of the trains. The train has to be permanently connected to the trains control center. This connection has a high priority level, and if the modem connection is lost, the train stops automatically [Dudoyer *et al.*, 2012].

However, under the circumstance of high-speed railway [Zhang *et al.*, 2012], vehicular communication often shows unstable, even sometime dreadful [Ai *et al.*, 2014]. Usually, when the speed is up to 350 kilometers per hour, there unavoidably arises some issues, such as Doppler shifts, fast cell switching and the penetration loss [Zhou and Ai, 2014]. The Doppler shifts results from the relative motion between a vehicle and a base station. Doppler Effect becomes another pivotal factor degrading system performance, which increases randomness of received signal [Liu *et al.*, 2011], [Li and Zhao, 2012], [Dybala and Radkowski, 2013]. The high speed operation of the train leads to fast cell switching. As a train moves across the footprint of the satellite beam, the receiving signal level may vary, especially towards the edge of the beam, which significantly impacts service rates even causing service drops [Li *et al.*, 2013], [Alkayal and Saada, 2013]. The fully enclosed body structure with good sealing property of the high-speed train results in penetration loss. Typically, the terminals inside the train connect to the base stations along the railway tracks via wireless links, in which the large penetration loss will directly degrade the communication link quality and decrease the cell coverage [Zhu *et al.*, 2013], [Liu *et al.*, 2012]. Furthermore, Federal Communications Commission (FCC) released the investigation on the usage of spectrum In 2003. It suggested that the authorized band in 3 – 6GHz range is less than 0.5% utilized on average. And so is the band below 3GHz, which is less than 35% [Commission and others, 2003]. Just based on these viewpoints, it is necessary to introduce a novel architecture for high-speed vehicular communication to address the issues

*Project supported by the National Nature Science Foundation of China (No. 61471252) and the Natural Science Foundation of Jiangsu Province (No. BK20130303).

[†]Corresponding Author: cwu@suda.edu.cn

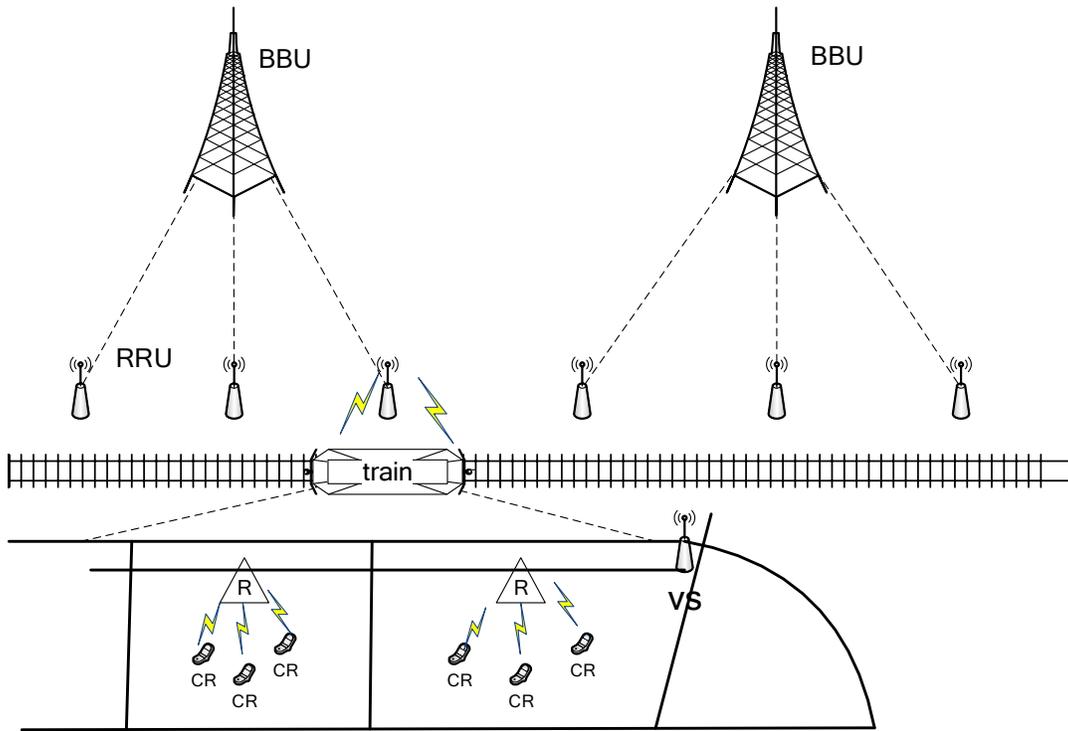


Figure 1: Networks architecture for the high-speed rail communication system.

from individual user's high-speed movement along the rails and the inefficiency in the spectrum usage.

In recent years, a lot of researchers used cognitive radio (CR) to improve the performance of wireless communication. The basic idea of CR networks is that the unlicensed devices (also called cognitive radio users or secondary users) need to vacate the spectrum band once they detect the licensed devices (also known as primary users). Simon Haykin defined the CR as an intelligent wireless communication system that is aware of its environment and uses the methodology of understanding-by-building to learn from the environment and adapt to statistical variations in the input stimuli [Haykin, 2005]. Letaief presented a cognitive space-time-frequency coding technique that can opportunistically adjust its coding structure by adapting itself to the dynamic spectrum environment [Letaief and Zhang, 2009]. Soyeon Kim proposed a CR operational algorithm for mobile cellular systems, which was applicable to the multiple secondary user environment [Kim and Sung, 2014]. These results proved CR technology can significantly reduce interference to licensed users, while maintaining a high probability of successful transmissions in a cognitive radio (CR) ad hoc network.

There are few publications about applying CR to the field of urban rail transit. Wu proposed a wireless cognitive model for high-speed individuals' spectrum management and show a small performance improvement in wireless communication [Wu *et al.*, 2015]. Although using cognitive radio in high-speed-railway has improved the performance, there are still so many issues that are open to address:

(1) Most of the cognitive radio users usually sense in the

same environment and each user is independent. So they compete each other for the spectrum resources, which leads to blind learning and frequent conflicts.

- (2) The rail transit contains a large number of CR users. While every user senses the environment, the system works with heavy workload and high computational complexity.
- (3) The operations of mutual competition and cooperation between the CR users interfere with not only primary users, but also themselves and their neighbors.
- (4) Spectrum holes in each base station are different. It would inevitably occur spectrum handoff.

For addressing the above issues, we try to propose a novel model of cognitive base station in the paper. Our proposed CBS attempts to use the authorized bands for railway without interrupting PUs. The CBS model should satisfy the following conditions:

- (1) The CBS can forecast spectrum holes according to its experience and assign spectrum to individuals within its range of coverage. In this way, the computational complexity of the entire network can be reduced.
- (2) The rail transit runs daily over a fixed route according to its timetable. The CBS can take the advantage of these characteristics, cooperate with each other to forecast spectrum holes on the whole route.

This paper is organized as follows. We first introduce the concept of cognitive base station and its mathematical model in Section 2. Section 3 then applies the novel CBS model

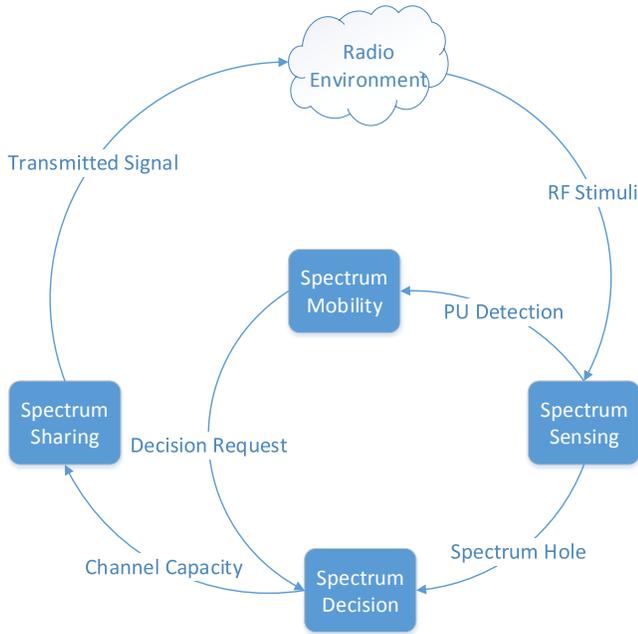


Figure 2: The cognitive cycle of a cognitive base station.

with RL into the scenario of high-speed rail, and propose the cooperation mechanism of multiple CBS agents. The experimental simulation results are given in Section 4. We conclude this paper in Conclusion.

2 Cognitive Base Station Model

Our proposed CBS is deployed along the railway, which works as a spectrum assigner. It learns from feedback received through interactions with an external environment and assigns spectrum to the passengers in the range of coverage. We consider each CBS to be an agent, which has four spectrum management functions: spectrum sensing, spectrum mobility, spectrum decision and spectrum sharing [Chkribene and Hamdi, 2015], [Lee and Akyildiz, 2012]. Fig. 2 gives the steps of the cognitive cycle within the framework of CBS, which is formed by the spectrum-aware operations. Each CBS agent uses reinforcement learning to operate spectrum management. All of the agents can sense the environment, obtain its own current state about spectrum usage, and communicate with each other for the purpose of cooperation. They then make decision according to its own state and the whole network situation, then use spectrum mobility to choose actions. Finally, these CBS agents continue to send its new state to the other neighbor CBS agents.

We assume that our cognitive radio network along high-speed rail consists of a collection of CBS agents and CR user agents. Each CBS agent has its own PUs and available spectrums. The CBS agents undertake decisions on choosing the spectrum independently of the CR user agents in the range. A choice of spectrum by the CBS agent i is essentially the choice of the frequency represented by $f^i \in F$. The CR user agents continuously monitor the spectrum that the CBS agent choose in each slot time. We assume perfect sensing,

in which, the CBS agents correctly infer the presence of the PUs if the former lies within the PUs' transmission range.

- Long-term Awareness of Spectrum Usage

Characterizing the spectrum bands based on their activity, and in particular, learning about the utilization of the channel is a key function of the CR users. Online learning algorithms must be developed that allow the CBS agents to continuously gather information about its radio environment, and construct a utilization function. Apart from simply classifying the spectrum as *busy* or *available*, it is beneficial if a probability distribution of the anticipated transmission/silent durations of the PUs can be derived. We propose a tightly integrated reinforcement learning equipped link layer protocol to schedule the transmissions between CBS agents and CR user agents over time.

- End-to-End Learning

Distributed networks rely on multihop forwarding of packets between a source-destination pair. Each CBS agent on this path learns of its own spectrum environment over time, and this information can be leveraged at the start and end points of the path to make optimal decisions regarding the spectrum choices and routing options. As an example, spectrum switching costs locally at a node affects end-to-end delays. While spectrum characteristics can be locally inferred, the specific choice of the spectrum at each link to minimize intra-path switching must be undertaken at the end points of the path. We explore ways to share this learning and spectrum awareness obtained by a node between its local neighbors, and subsequently over multiple hops to the destination. The cost of this learning and the benefits are investigated as part of this project.

3 SPECTRUM MANAGEMENT BASED COGNITIVE BASE STATION

3.1 The Q-Learning

Reinforcement learning, which is inspired by psychological learning theory from biology [Waltz and Fu, 1965], enables the agent to learn behavior through trial-and error interactions with a dynamic environment [Sutton and Barto, 1998]. The classical reinforcement algorithm is Q-Learning, the process of which is as follows [Puterman, 1994]. On each step of interaction the agent chooses an action according to the external environment based on its current state. As a result, the action changes the environment and receives a reward. The agent need to develop a policy, that maximizes the long-run measure of reinforcement.

The classic reinforcement learning algorithm is formulated as follows. At each time t , the agent perceives its current state $s_t \in S$ and the set of possible actions A_{s_t} . The agent chooses an action $a \in A_{s_t}$ and receives from the environment a new state s_{t+1} and a reward r_{t+1} . Based on these interactions, the reinforcement learning agent must develop a policy $\pi : S \rightarrow A$ which maximizes the long-term reward $R = \sum_t \gamma r_t$ for MDPs, where $0 \leq \gamma \leq 1$ is a discounting factor for subsequent rewards. The long-term reward is

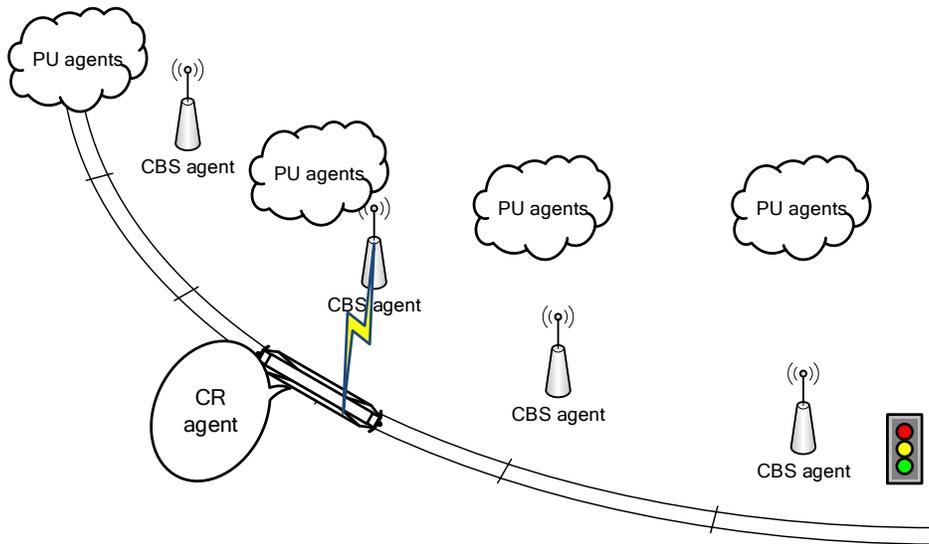


Figure 3: The cognitive base station within the high-speed-rail transportation.

the expected accumulated reward that the agent expects to receive in the future under the policy, which can be specified by a *value function*. In this way, the Q-learning can calculate an update to its expected discounted reward, $Q(s_t, a_t)$ as follows:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha[r_t + \gamma \max_a Q(s_{t+1}, a) - Q(s_t, a_t)]$$

where γ is the discount factor such that $0 \leq \gamma < 1$. The agent stores the state-action values in a table Q [Wu *et al.*, 2010], [Jiang *et al.*, 2011], [Bkassiny *et al.*, 2013].

Recently the reinforcement learning has attracted increasing interest in the machine learning and artificial intelligence communities. Kadam *et al.* applied the Q-Learning into routing data in Wireless Sensor Network scenario to route data efficiently from one source to multiple mobile sinks [Kadam and Srivastava, 2012]. It turned out that the algorithm can extend the network lifetime.

3.2 Application to Cognitive Base Station

We illustrate the high-speed railway environment with CBS agents along the way in Fig. 3. We further model a cognitive radio network as consisting of a set of Cognitive Base Stations, denoted CBS , a set of primary users, denoted PU , and a set of available frequencies, denoted SP . We assume that the topological structure of a given network is fixed.

Spectrum holes vary due to the behavior of PUs, which causes the change of environment. CBS agents can perceive the states within the environment. The state of an CBS agent is the current spectrum of its transmission. The state of the multi-agent system includes the state of every CBS agent. We therefore define the state of the system at time t , denoted s_t , as

$$s_t = (\vec{s}_p)_t$$

, where \vec{s}_p is a vector of spectrums across all agents. Here sp_i are the spectrum on the i th agent and $sp_i \in \vec{SP}$. Nor-

mally, if there are m spectrums, we can use the index to specify these spectrums. In this way, we have $\vec{SP} = \{SP_1, SP_2, \dots, SP_m\}$.

At a particular time and a particular state, the CBS will take action according to learning results to either switch channel or transmit. At time t we define $a_t = k$, where k is the action that CBS chooses at time t and

$$k \in \{switch_to_channel_1, switch_to_channel_2, \dots, switch_to_channel_m, transmit_data\}.$$

Once the CBS agent has detected any active PU, it would take action to channel switching. We use the Q table to store state-action values. At time t , the state is sp_t and the action is k , then we can calculate the value $Q(sp_t, k)$ by the above Q-learning formulas. If PU is detected, the CBS agent would switch to the other available spectrum with the largest Q -value.

The reward is the estimate for spectrum usage availability on a CBS agent. The different network situation results in different rewards as follows.

- *CR-PU interference*: If a PU's activity occurs in the spectrum shared by any CR user, and in the slot same selected for transmission, then a high penalty of -15 is assigned. The intuitive meaning of this is as follows: We can avoid the collisions among the CR users using the mediation from the CBS agents. However, the concurrent use of the spectrum with a PU goes against the principle of protection of the licensed devices, and hence, must be strictly avoided.
- *Successful Transmission*: If none of the above conditions are observed to be true in the given transmission slot, then packet is successfully transmitted from the sender to receiver, and a reward of $+5$ is assigned, which is found experimentally to give the best results.

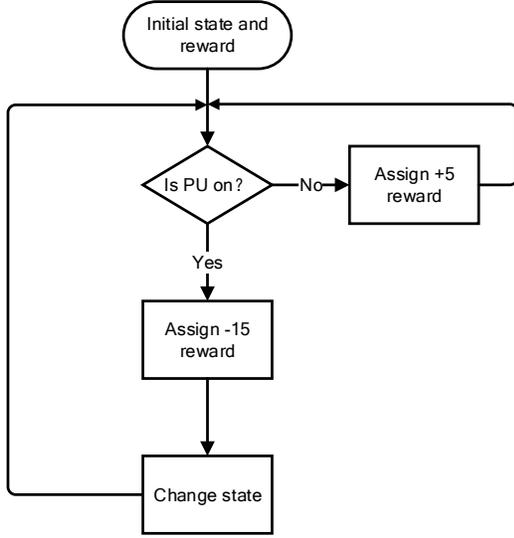


Figure 4: The Q-learning process on CBS model.

Once detected the primary user, a harsh punishment will be given. Otherwise, a positive reward will be assigned. Fig. 4 illustrates the proposed process, and Algorithm 1 describes our algorithm for implementing the Q-learning on CBS agent.

4 EXPERIMENTAL SIMULATION

4.1 Experimental Design

In this section, we describe preliminary results from applying our reinforcement learning based approach to the cognitive radio model. To detect the PUs correctly is the necessary prerequisite. The overall aim of our proposed learning based approach is to allow the CBS agents to decide on an optimal choice of spectrum so that (i) PUs are not affected, and (ii) CR users share the spectrum in a fair manner. These two rules are to simulate the public's behaviors in Urban Rail Transit Environment. That is, those bands that are frequently occupied by licensed users are rarely utilized because of open areas or relatively closed environment, and the public can opportunistically use band resources with a same probability.

Our novel CBS network simulator within the framework of high-speed rail has been designed to investigate the effect of the proposed reinforcement learning technique on the network operation. The implemented *ns-2* model is composed of several modifications to the physical, link and network layers in the form of stand-alone C++ modules. The PU Activity Block describes the activity of PUs based on the on-off model, including their transmission range, location, and spectrum band of use. The Channel Block contains a channel table with the background noise, capacity, and occupancy status. The Spectrum Sensing Block implements the energy-based sensing functionalities, and if a PU is detected, the Spectrum Management Block is notified. This, in turn causes the device to switch to the next available channel, and also alert the upper layers of the change of frequency. The Spectrum Sharing Block coordinates the distributed channel access, and calculates the interference at any given node due to the ongoing

Algorithm 1 Pseudo code of Q-learning on CBS

Main()

Initialize state s_t and action a_t and their \vec{Q} value;

repeat

 Q-learning(s_t, a_t, \vec{Q})

until all episodes are traversed

Q-with-Kanerva(s_t, a_t, \vec{Q})

repeat

 Take action s_t , observe reward r_t , get next state s_{t+1}

 Get $Q(s_t a_t)$ from the Q -table;

for all actions a^* under new state s_{t+1} **do**

 Generate the state-action pair $s_{t+1} a_{t+1}$ from state s_{t+1} and action a^*

 Get $Q(s_{t+1} a_{t+1})$ from the Q -table;

end for

$\delta = r + \gamma * \max Q(s_{t+1} a_{t+1}) - Q(s_t a_t)$

$\Delta \vec{Q} = \alpha * \delta$

$\vec{Q} = \vec{Q} + \Delta \vec{Q}$

$s_t = s_{t+1}$

if random probability $\leq \epsilon$ **then**

for all actions a^* under current state s_t **do**

$a_t = \text{argmax}_a Q(s_t a_t)$

end for

else

$a_t = \text{random action}$

end if

until s_t is terminal

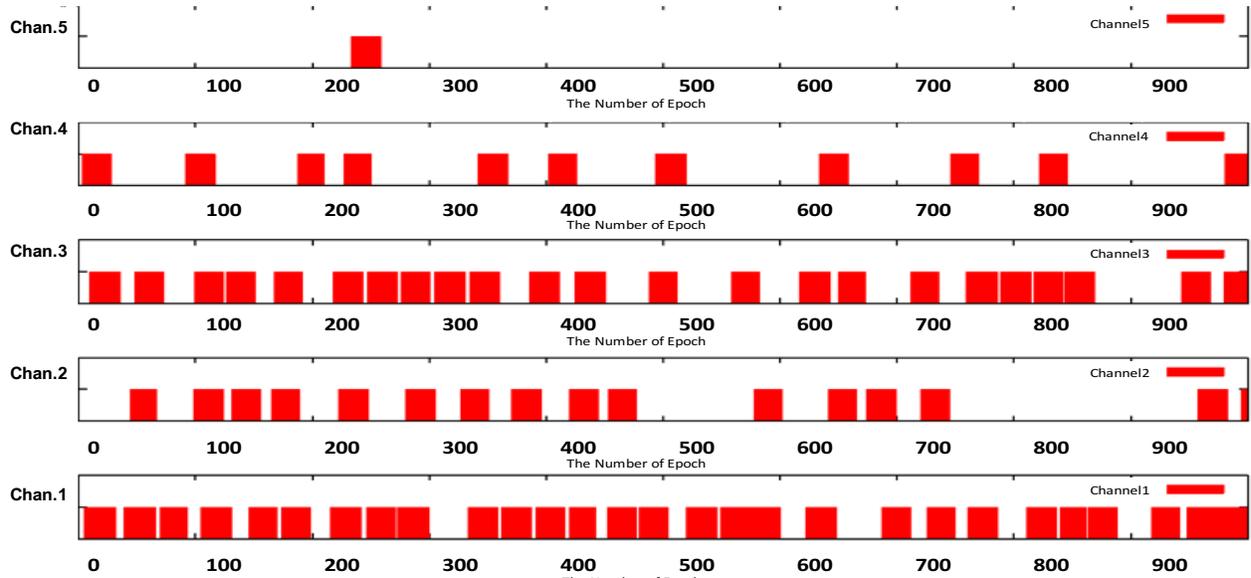
transmissions in the network. The Cross Layer Repository facilitates the information sharing between the different protocol stack layers.

We conduct our experiment in the following scenario: there are 2 trains which take on 21 passengers for each and 5 CBS agents aside the railway. The average speed of train is $10m/s$. We have 10 primary users in the range of each CBS. The activity of primary users is based on ON-OFF model and each primary user is assigned the spectrum randomly from 5 spectrums (small network) or 10 spectrums (large network). The CBS agent senses the spectrum holes per 0.1 second and assigns available spectrum to CR user agent. The simulation parameters are summarized in Table 1.

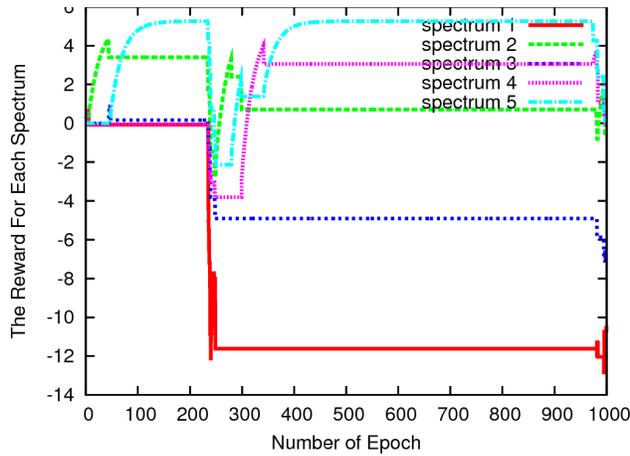
4.2 Experimental results

We compare the performance of our CBS with reinforcement learning (CBS-RL) scheme with the CBS with Round-Robin scheme (CBS-RR), which is a typical way in GSM-R system. The Round-robin (RR) scheme employs the principle that once a spectrum is not available, the agent switches to next channel in equal portions and in circular order, handling all switches without priority (also known as cyclic executive). This method is simple, easy to implement, and starvation-free. In our RL-based scheme, the exploration rate ϵ is set to 0.2, which we found experimentally to give the best results. The initial learning rate α is set to 0.8, and it is decreased by a scaling factor of 0.995 after each time slot.

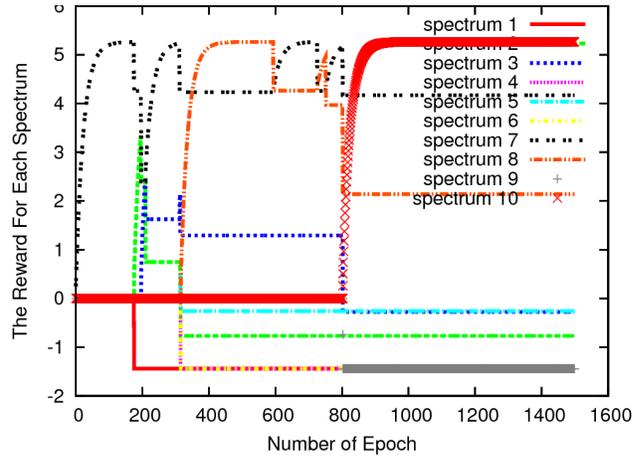
Figure 5(a) shows an example about the distribution of



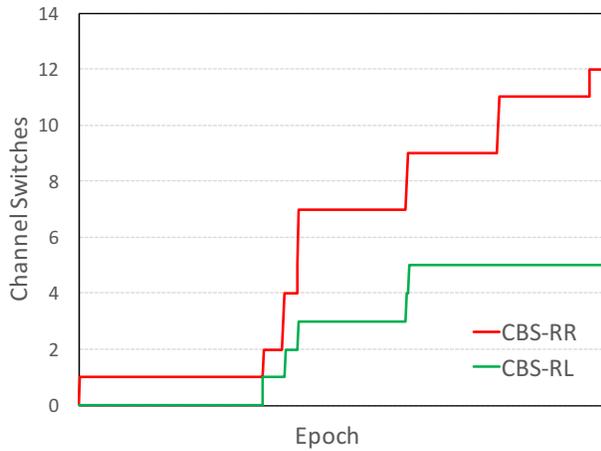
(a) An example about the distribution of spectrums occupancy on CBS with 5 spectrums.



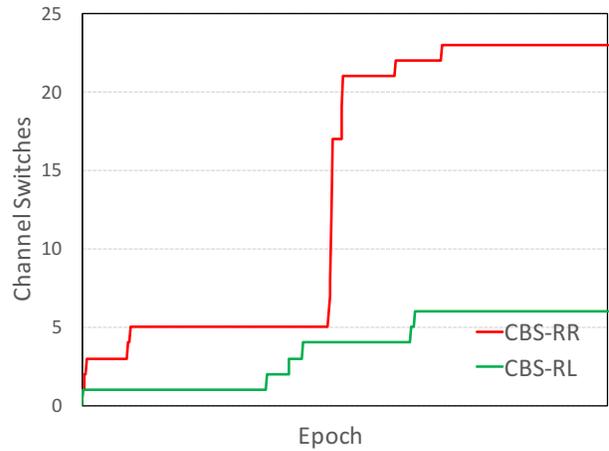
(b) Average rewards for 5 spectrum bands



(c) Average rewards for 10 spectrum bands



(d) Cumulative number of channel switching for 5 spectrum bands



(e) Cumulative number of channel switching for 10 spectrum bands

Figure 5: CBS simulations with RL and RR schemes.

Table 1: Simulation Parameters

| Parameters | Values |
|----------------------------------|----------------|
| Topology size | X:7000m Y:500m |
| Number of passengers | 42 |
| Number of primary users | 50 |
| Number of cognitive base station | 5 |
| Speed | 10m/s |
| Number of spectrums | 6 |
| Bandwidth | 2000000Hz |
| Simulation time | 1000s |

spectrums occupancy on the CBS with 5 spectrums. Spectrums occupancy on CBS follows the ON-OFF model: the ON mode is in the normal distribution with the parameter $\mu = 25$, and the OFF mode is in the exponential distribution with the parameter β . the value of which is randomly generated.

Figure 5(b) and 5(c) show the average rewards received by CBS agent across all spectrums using the CBS-RL scheme. The result in Figure 5(b) shows that after learning over 1000 epochs, Channel 5 receives the largest positive reward of approximately +5.5, while Channel 1, 2, 3 and 4 gets a reward of approximately -11.8, +0.7, -5.1 and +3.3. The results indicate that our approach pushes the CBS agents to gradually achieve higher positive rewards and choose more suitable spectrum for their transmission. The results also indicate that the reward tends to be suitable to the distribution of spectrums occupancy. A similar trend is observed in Figure 5(c), with Channel 10 receiving the highest average reward of approximately +5.2.

Figure 5(d) and 5(e) show the cumulative number of channel switching using CBS-RL and CBS-RR schemes. The result in Figure 5(d) shows the average number of channel switches for the small topology. We observe that after learning, the CBS-RL scheme tends to decrease number of channel switching to 5, while CBS-RR keeps the channel switches to approximately 12. For the large topology in Figure 5(e), the CBS-RL scheme reduces the channel switches to 6, while CBS-RR keeps the channel switches approximately 23. The results indicate that our proposed CBS-RL approach can keep the channel switches lower than the CBS-RR approach and converge to an optimal solution.

5 CONCLUSIONS

To address the issues of frequent channel switches and inefficient blind learning in high-speed rail, we propose a novel concept of Cognitive Base Station, which has the capability of forecasting spectrum holes and assigning spectrum to individuals. Our simulation results prove that after autonomous learning, the CBS-RL scheme can forecast spectrum holes. In this way, our proposed model can significantly improve the performance of vehicular communication, which can decrease cell-switching and unsuccessful transmission.

References

[Ai *et al.*, 2014] Bo Ai, Xiang Cheng, Thomas Kurner, Zhangdui Zhong, Ke Guan, Ruisi He, Lei Xiong, David W

Matolak, David G Michelson, and Cesar Briso-Rodriguez. Challenges toward wireless communications for high-speed railway. *Intelligent Transportation Systems, IEEE Transactions on*, 15(5):2143–2158, 2014.

- [Alkayal and Saada, 2013] Faisal Alkayal and Johnny Bou Saada. Compact three phase inverter in silicon carbide technology for auxiliary converter used in railway applications. In *Power Electronics and Applications (EPE), 2013 15th European Conference on*, pages 1–10. IEEE, 2013.
- [Bkassiny *et al.*, 2013] Mario Bkassiny, Yang Li, and Sudharman K Jayaweera. A survey on machine-learning techniques in cognitive radios. *Communications Surveys & Tutorials, IEEE*, 15(3):1136–1159, 2013.
- [Chkirkbene and Hamdi, 2015] Zina Chkirkbene and Nouredine Hamdi. A survey on spectrum management in cognitive radio networks. *International Journal of Wireless and Mobile Computing*, 8(2):153–165, 2015.
- [Commission and others, 2003] Federal Communications Commission et al. Facilitating opportunities for flexible, efficient, and reliable spectrum use employing cognitive radio technologies. *Et docket*, (03-108):05–57, 2003.
- [Dudoyer *et al.*, 2012] Stephen Dudoyer, Virginie Deniau, Ricardo Adriano, MN Ben Slimen, Jean Rioult, Benoît Meyniel, and Marion Berbineau. Study of the susceptibility of the gsm-r communications face to the electromagnetic interferences of the rail environment. *Electromagnetic Compatibility, IEEE Transactions on*, 54(3):667–676, 2012.
- [Dybala and Radkowski, 2013] Jacek Dybala and Stanislaw Radkowski. Reduction of doppler effect for the needs of wayside condition monitoring system of railway vehicles. *Mechanical Systems and Signal Processing*, 38(1):125–136, 2013.
- [Haykin, 2005] Simon Haykin. Cognitive radio: brain-empowered wireless communications. *Selected Areas in Communications, IEEE Journal on*, 23(2):201–220, 2005.
- [isheng Zhao *et al.*, 2013] isheng Zhao, Xi Li, Yi Li, and Hong Ji. Resource allocation for high-speed railway downlink mimo-ofdm system using quantum-behaved particle swarm optimization. In *Communications (ICC), 2013 IEEE International Conference on*, pages 2343–2347. IEEE, 2013.
- [Jiang *et al.*, 2011] Tianzi Jiang, David Grace, and Paul D Mitchell. Efficient exploration in reinforcement learning-based cognitive radio spectrum sharing. *Communications, IET*, 5(10):1309–1317, 2011.
- [Kadam and Srivastava, 2012] Kaveri Kadam and Navin Srivastava. Application of machine learning (reinforcement learning) for routing in wireless sensor networks (wsns). In *Physics and Technology of Sensors (ISPTS), 2012 1st International Symposium on*, pages 349–352. IEEE, 2012.
- [Kim and Sung, 2014] Soyeon Kim and Wonjin Sung. Operational algorithm for wireless communication systems using cognitive radio. In *Communication, Networks and Satellite (COMNETSAT), 2014 IEEE International Conference on*, pages 29–33. IEEE, 2014.

- [Lee and Akyildiz, 2012] Won-Yeol Lee and Ian F Akyildiz. Spectrum-aware mobility management in cognitive radio cellular networks. *Mobile Computing, IEEE Transactions on*, 11(4):529–542, 2012.
- [Letaief and Zhang, 2009] Khaled Ben Letaief and Wei Zhang. Cooperative communications for cognitive radio networks. *Proceedings of the IEEE*, 97(5):878–893, 2009.
- [Li and Zhao, 2012] Jinxing Li and Youping Zhao. Radio environment map-based cognitive doppler spread compensation algorithms for high-speed rail broadband mobile communications. *EURASIP Journal on Wireless Communications and Networking*, 2012(1):1–18, 2012.
- [Li et al., 2013] Ying Li, Lei Lei, Zhangdui Zhong, and Siyu Lin. Performance analysis for high-speed railway communication network using stochastic network calculus. In *Wireless, Mobile and Multimedia Networks (ICWMMN 2013), 5th IET International Conference on*, pages 100–105. IET, 2013.
- [Liu et al., 2011] Qiuyan Liu, Miao Wang, and Zhangdui Zhong. Statistics of capacity analysis in high speed railway communication systems. *Tamkang Journal of Science and Engineering*, 14(3):209–215, 2011.
- [Liu et al., 2012] Liu Liu, Cheng Tao, Jiahui Qiu, Houjin Chen, Li Yu, Weihui Dong, and Yao Yuan. Position-based modeling for wireless channel on high-speed railway under a viaduct at 2.35 ghz. *Selected Areas in Communications, IEEE Journal on*, 30(4):834–845, 2012.
- [Puterman, 1994] M. L. Puterman. Markov decision processes. In *Wiley*, 1994.
- [Sniady and Soler, 2012] Aleksander Sniady and Jose? Soler. An overview of gsm-r technology and its shortcomings. In *ITS Telecommunications (ITST), 2012 12th International Conference on*, pages 626–629. IEEE, 2012.
- [Sutton and Barto, 1998] R. Sutton and A. Barto. *Reinforcement Learning: An Introduction*. Bradford Books, 1998.
- [Tian et al., 2012] Lin Tian, Juan Li, Yi Huang, Jinglin Shi, and Jihua Zhou. Seamless dual-link handover scheme in broadband wireless communication systems for high-speed rail. *Selected Areas in Communications, IEEE Journal on*, 30(4):708–718, 2012.
- [Waltz and Fu, 1965] M. D. Waltz and K. S. Fu. A heuristic approach to reinforcement learning control systems. In *IEEE Transactions on Automatic Control*, 10:390-398., 1965.
- [Wu et al., 2010] Cheng Wu, Kaushik Chowdhury, Marco Di Felice, and Waleed Meleis. Spectrum management of cognitive radio using multi-agent reinforcement learning. In *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems: Industry track*, pages 1705–1712. International Foundation for Autonomous Agents and Multiagent Systems, 2010.
- [Wu et al., 2015] Cheng Wu, Yiming Wang, Xiang Qiang, and Zhaoyang Zhang. Adaptive spectrum management of cognitive radio in intelligent transportation system. In *Applied Mechanics and Materials*, volume 743, pages 765–773. Trans Tech Publ, 2015.
- [Zhang et al., 2012] Jiayi Zhang, Zhenhui Tan, Xiaoxi Yua, Haibo Wang, and Linwen Zhang. Review of public broadband access systems for high-speed railways and key technologies. *Journal of the China Railway Society*, 34(1):46–53, 2012.
- [Zhou and Ai, 2014] Yuzhe Zhou and Bo Ai. Quality of service improvement for high-speed railway communication-s. *Communications, China*, 11(11):156–167, 2014.
- [Zhu et al., 2013] Xiangqian Zhu, Shanzhi Chen, Haijing Hu, Xin Su, and Yan Shi. Tdd-based mobile communication solutions for high-speed railway scenarios. *Wireless Communications, IEEE*, 20(6):22–29, 2013.