Intelligent Power Controller of Wireless Body Area Networks Based on Deep Reinforcement Learning

Peng He[1,2,3](✉), Zhenli Liu[1,2,3], Lei Fu[1], Zhongyuan Tao[1,2,3], Jia Liu[1,2,3], Tong Tang[1,2,3], and Zhidu Li[1,2,3]

[1] School of Communication and Information Engineering,
Chongqing University of Posts and Telecommunications, Chongqing, China
`hp6500@126.com`,
`1132780762@qq.com`, `761386167@qq.com`, `1300093246@qq.com`,
`1747543987@qq.com`,
`{tangtong,lizd}@cqupt.edu.cn`
[2] CN Key Laboratory of Optical Communication and Networks in Chongqing,
Chongqing, China
[3] Key Laboratory of Ubiquitous Sensing and Networking in Chongqing,
Chongqing, China

Abstract. Wireless Body Area networks allow groups of tiny sensors to communicate for purpose of medical applications. With the progress of sensor manufacture and artificial intelligence, abundant wearing devices are produced and applied with powerful intelligence functionalities. In wireless body area networks, battery energy capacity and inter-network interference are two serious threats to restrict the raise of performance. In this work, we focus on the power controlling theme in wireless body area networks. First, we introduce the primer overview of the deep-Q-Network algorithm, which is the method utilized in this work. Second, we present our communication system which is composed of two interfered WBANs. Third, we show how to design the power controller based on the deep-Q-network algorithm. The results reveal that our proposed power controller significantly decreases energy consumption by sacrificing little throughput performance.

Keywords: Wireless body area network · Power controller · Deep Q network

---

Supported by the National Natural Science Foundation of China (Grant No. 61901070, 61871062, 61771082, 61801065), partially supported by the Science and Technology Research Program of Chongqing Municipal Education Commission (Grant No. KJQN201900611, KJQN201900604), and partially supported by Program for Innovation Team Building at Institutions of Higher Education in Chongqing (Grant No. CXTDX201601020).

Y. Chen et al. (Eds.): BICT 2020, LNICST 329, pp. 239–249, 2020.
https://doi.org/10.1007/978-3-030-57115-3_21

1 Introduction

The rapid growing demands of healthcare application promote the development of wireless body area network (WBAN) using various sensors, which are distributed on-body or imbedded (in-body), obtaining the fatal physiological data to assist medical cures [1]. These sensors play an important role in the activities of WBAN, obtaining several important physiology data as diverse as electrocardiogram (ECG), electroencephalograph (EEG), etc. Emerging networks and communication technologies are fully considered in the applications of healthcare including disease prediction and detection in order to provide better service for patients. New generation of WBANs is significantly improved in terms of transmission rate, latency, QoE and security.

The development of WBAN still faces plenty of challenges. A serious one is the limitation of battery energy, which restrains the durability of the networks. In a WBAN, distinct activities of nodes consume various level of energy. For instance, the power-on, sensing, communication, storing and computation are general processes which rapidly consume life of sensors. In addition, frequency switch of sensors is another type of killer, which significantly increases the capability demand of batteries.

To improve energy efficiency, one available way is utilizing dynamic power controller in different processes of the communication. The concept of dynamic controller has been proposed for a long time. It can be classified into types of open loop and closed loop based the concept of controlling engineering. In wireless networks, power controlling methods are utilized to realize different goals. For example, CDMA cellular networks apply power controlling in order to minish self-interference and increase the channel capacity. Differently, OFDMA cellular networks apply power controlling in order to decrease uplink energy consuming and inter-cell interference. Dynamic power controller can be deployed in different places to allocate the transmission power of the nodes. The major target of this work is to use dynamic power controller to minish inter-network interference and decrease energy consumption of WBANs.

Artificial intelligence is attracting great attentions due to the recent progress of deep learning and reinforcement learning. By combining the two learning methods, deep reinforcement learning (DQN) has been proposed which owns advantages of both ones [2]. DQN method has been successfully applied for vast scope of areas including WBAN systems for purpose of intelligent medical applications.

In this paper, we focus on dynamic power controlling issue in wireless body area network. To motivate this work, the proposed communication system is composed of two WBANs. In a communication, physiological data is obtained by two groups sensor nodes, relayed by the central nodes and finally uploaded onto the cloud, online doctors or hospitals. The contribution of this work is listed as follows, First, we propose a dynamic power controller named Deep-Q-Network-based Power Controller (DQNPC), which increases adaptation of WBAN nodes to environment. Second, we validate the feasibility and efficiency of proposed DQNPC in a simulation of two WBANs.

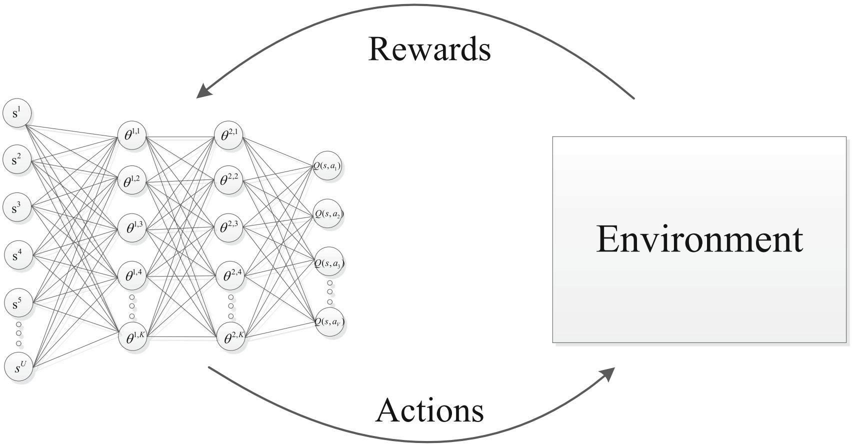

Fig. 1. The proposed DQN framework

2 Related Work

Body area networks covers different research scopes which can be imitated from area of wireless communication. For example, [3] focused on security topics of WBAN where a state-of-art survey of related authentication techniques is presented. [4] proposed an improved energy efficient routing protocol for wireless body area sensor networks to obtain energy-saving and reliable performance. [5] proposed a wearable sensor node with solar energy harvesting and Bluetooth low energy transmission that enables the implementation of an autonomous WBAN. [6] applied finite-difference time-domain concept to study dynamic channel characters of wireless body area network during the walking process.

A group of literatures studied the power controlling issue for WBAN. For instance, [7] proposed a game-theoretic approach for power control and relay selection in WBAN. [8] applied a deep learning tool for channel prediction to transmit controlled power in wireless body area networks. [9] proposed a novel power controller to mitigate inter-network interference in WBANs to increase the maximum achievable throughput with the minimum energy consumption.

Deep Reinforcement Learning (DRL) is an emerging powerful tool which plays a significant role in decision and prediction behaviours of various applications. Deep Q-learning (DQN) is an typical example of DRL, which is a type of off-line method which is developed based advanced Markov Decision Process [10]. Compared with traditional optimization methods such as dynamic programming, DQN help networks to solve uncertainty and stochastic problems in a accurate manner [11]. The framework of DQN maintains progress once it comes out. As early as 1992, work of [13] proved that DQN algorithm converges on action-values with probability 1. In [12], the step size selection of DQN training was investigated in order to optimize the learning process. In [14], a multi-agent DQN algorithm was proposed for general general-sum Markov games.

3 A Primer of Deep Reinforcement Learning

In this section, we will present a primer framework of DQN methods, as shown in Fig. 1. DQN is a typical off-line class of reinforcement learning. The network nodes deploy DQN and receive rewards based on a group of observed states, and then decide optimized actions in order to adapt the environment. Compared with traditional reinforcement learning, DQN utilizes deep neural network to compute strategies instead of Q tables. As a result, DQN is capable of dealing with larger scopes of observed data with a rapid and effective way. In this work, the proposed DQN method contains the following elements (Fig. 2):

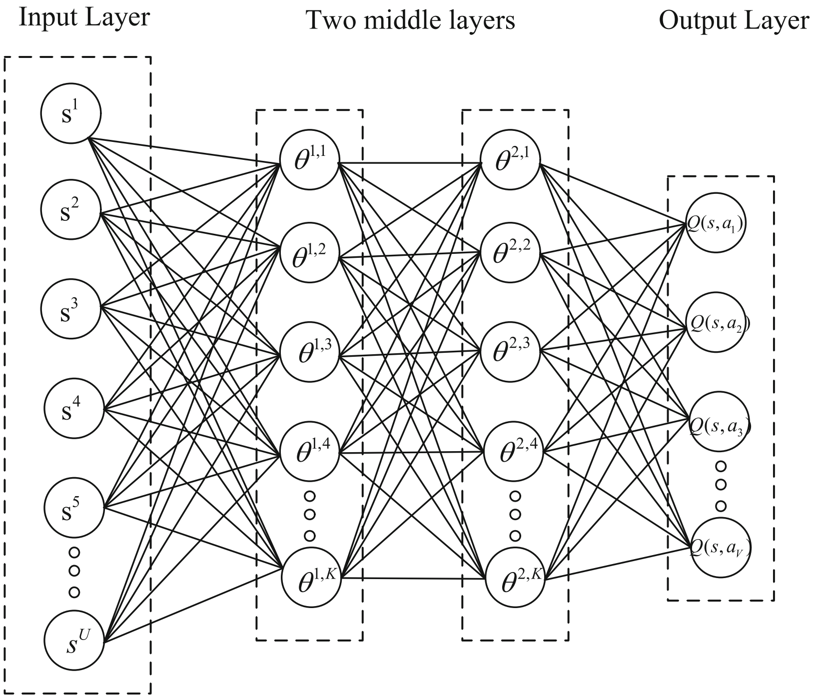

Fig. 2. Utilized neural network with 4 layers including status sector as the input layer, policies as the output layer, and 2 middle layers for computation.

1) Statuses (s_t). The statuses of desired target may be discrete or continuous quantifiable indicators.
2) Actions (a_t). The actions are defined as the active changes of statuses, such as the jump between discrete statuses or changes of continuous statuses.
3) Policies (π). The polices indicate mapping between the environment and actions.

4) Rewards (R_{s_t,a_t}). The rewards are desired gain of certain index when actions are carried out.
5) Strategies ($Q_{\pi}(s_t, a_t, \theta)$). The Strategies are made based on the polices, which guide the selection of actions in order to obtain maximum rewards.

The above-mentioned elements work together and promote the learning process of WBAN nodes. In traditional reinforcement learning, strategies $Q_{\pi}(s_t, a_t)$ alters in every time slot. However, it is slow or even unable to form convergence when $t \to \infty$ is caused by significant stochastic property of environment. In the DQN Method, a neural network is introduced to improve the convergence speed of learning. The weight is denoted by Θ and further vectored as $\theta = vec(\Theta)$. Accordingly, the strategies are upgraded from $Q_{\pi}(s_t, a_t)$ to $Q_{\pi}(s_t, a_t, \theta)$. To obtain the optimal strategies, the proposed neural network is composed of one input layer (s_t), two hidden layers and one output layer $Q_{\pi}(s_t, a_t, \theta)$. See Fig. 3. In the utilized neural network, the activation function is chosen as $\sigma(x) = 1/(1 + e^{-x})$ [15].

The DQN framework contains the training and verification processes. In the training process, the major target is to minimum the mean-squire error loss function of θ which is given by,

$$\min_{\theta_t} \; M(\theta_t) := E_{s,a}[(z_t - Q_{\pi}(s_t, a_t; \theta_t))^2] \tag{1}$$

where $\mathrm{z(t)} := \mathrm{E_{s'}}[\mathrm{r_{s,s',a}} + \gamma \max_{a'} \mathrm{Q_{\pi}(s', a'; \theta_{t-1}) | s_t, a_t}]$ indicate the estimated function in a duration t, with current state s, previous state s' and action a.

In the training process, the vectored weights θ update in each duration based on the stochastic gradient decent algorithm,

$$\theta_{\mathrm{t+1}} := \theta_{\mathrm{t}} - \kappa \nabla L_t(\theta_{\mathrm{t}}) \tag{2}$$

where experience of change of θ_t is stored in a buffer $\aleph$. In addition, the state-action value function is defined with Bellman equation,

$$\mathrm{Q}^*_{\pi}(\mathrm{s_t}, \mathrm{a_t}) := \mathrm{E_{s'}}[\mathrm{r_{s,s',a}} + \gamma \max_{a'} \mathrm{Q}^*_{\pi}(\mathrm{s'}, \mathrm{a'}) | \mathrm{s_t}, \mathrm{a_t}] \tag{3}$$

Note that γ is the discount factor which locates in area of $(0, 1)$. In order to implement the training process, the goal is to find a strategy to maximize $\mathrm{Q}^*_{\pi}(\mathrm{s_t}, \mathrm{a_t})$.

4 Power Controlling Scheme

4.1 System Model

As shown in Fig. 3, we consider two WBANs in the proposed system, which is respectively denoted by C_1 and C_2. Each WBAN is personal-centric deployed which contains one central node (S_c) and several sensor nodes ($S_{s,i}, i = 1, 2...n$). The sensor nodes are on the surface of skin or embedded inside the body, that are capable of collect physiological data via sensing technologies. The central node is a type of intelligent device, which is assumed capable of receiving data from the sensor nodes and delivery them to the cloud network or medical server.

The development of WBAN faces several challenges. A serious challenge of WBAN is the inter-network interference, that is caused by interfered transmitting power of nearby nodes. It is known that inter-network interference is serious when neighbouring WBANs work in a narrow space, which restrains the network performance such as bit error rate (BER), network throughput, etc. In our system model, we consider two WBANs as the simplest study case.

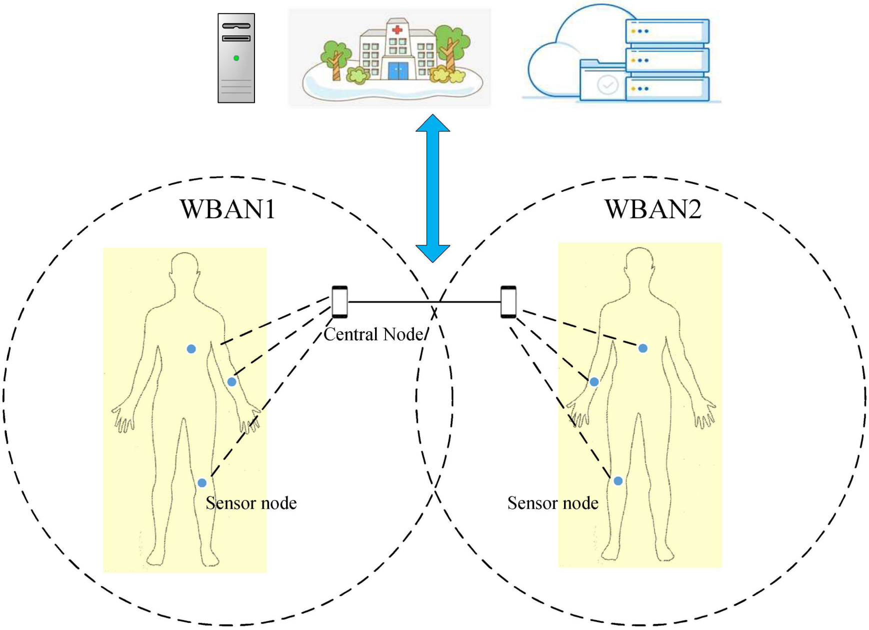

Fig. 3. Proposed system model which contains two WBANs.

Let $x_i(t)$ be the transmitted signal from the i^{th} sensor node of WBAN1, the received signal of central node in WBAN1 is expressed by,

$$y(t) = \sum_{i \in C_1}^{N} x(t) * h_i(t) + \sum_{j \in C_2} x(t) * h_{ij}(t) + n(t) \qquad (4)$$

where $n(t)$ denotes channel noise which is model by AWGN, N is the total number of corresponding links. $h_i j(t)$ denote the fading distribution function of signal between node i and j among different WBANs. $h_i(t)$ denotes the fading distribution function of signal in a same WBAN. We apply the Signal to Interference plus Noise Ratio (SINR) to calculate the interference of node i during time slot τ, that is expressed by,

$$\gamma_i^\tau = \frac{H_i^\tau p_i^\tau}{\frac{1}{B}\sum_{i \neq j} H_{ij}^\tau p_j^\tau + N^\tau} \tag{5}$$

where H_i^τ and H_{ij}^τ respectively means the frequency function of $h_i(t)$ and $h_i j(t)$ using Fast Fourier Transform Algorithm. B is the bandwidth utilized in the communication. p_i is the transmission power of the i^{th} sensor node. N^τ is the frequency function of the noise. The link capacity of the node i is calculated using Shannon formula:

$$U_i^\tau = Blog\frac{1}{1+\gamma_i^\tau} \tag{6}$$

4.2 Power Controller Design

The proposed power controller is designed based on the DQN method presented in the previous section, namely Deep-Q-learning-based Power Controller (DQNPC). The power controller is assumed to be deployed the central node of each WBAN, that regulate the allocation of transmission power through message exchange with sensor nodes. By applying the formula of (5), the target of the power controlling theme is to minimize the inter-network interference as,

$$max\sum_\tau\sum_i U_i^\tau \tag{7}$$

$$min\sum_\tau\sum_i p_i^\tau \tag{8}$$

The state vector of the DQNPC is defined as,

$$s^\tau = [\gamma_i^\tau, U_i^\tau, p^\tau] \tag{9}$$

Here γ_i^τ, U_i^τ and p^τ respectively denote SINR, link capacity and transmitting power during time slot τ. We assume that γ_i^τ can be estimated based on the received data of central node. U_i^τ is calculated on the basis of (6). We also assume p^τ can be obtained in the transmitter sensor nodes, and sent to the central node via the feed-back link. In addition, we ignore the interference between distinct sensor node in one WBAN, since exiting modulation and multi-access technologies have been studied to deal with this problem.

Reward function $R(t)$ is the key factor of DQN method which promotes the learning process. We use a logarithmic equation to design the reward function $R(t)$ based on the target of (8),

$$R(t) = log\frac{\sum_\tau\sum_i U_i^\tau}{\sum_\tau\sum_i p_i^\tau} \tag{10}$$

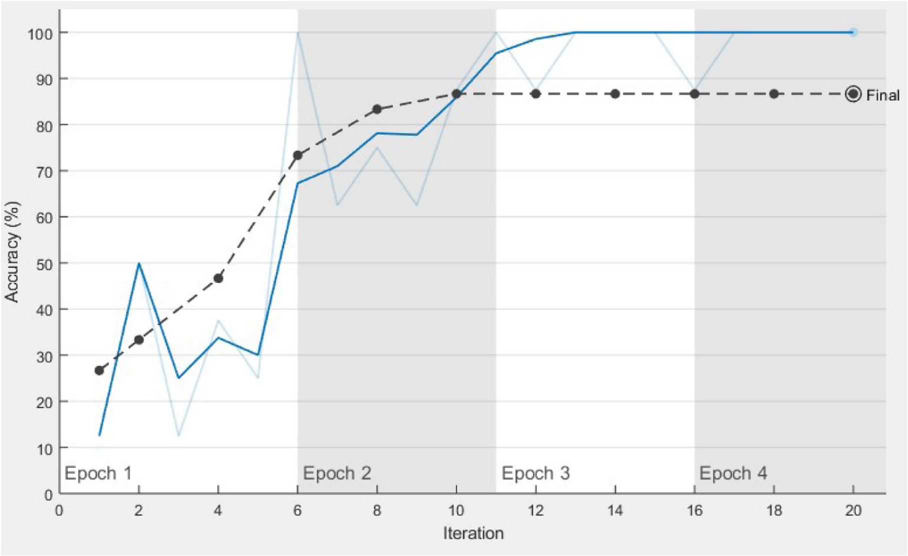

Fig. 4. The accuracy of DQNPC in the training and validation process. The controller restrains in 4 epochs

5 Performance Evaluation

In this section, we first describe how to design the simulation configuration and list major simulation parameters. Then we show the simulation results based on our proposed controlling scheme.

5.1 Simulation Configuration

In this work, we apply the MATLAB 2019a simulator to validate the proposed DQNPC. MATLAB 2019a has added packages to support the simulation of deep learning and reinforcement Learning. We compare our proposed DQNPC with existing power controller, which is FPC. FPC is the fix power controller, which statically allocate the transmission power in each time slot.

Major parameters of the simulation is as follows. The transmission power locates from −50 to 50 dBm. The noise power is set as −100 dBm. The communication bandwidth is set as 10 KHz. The number of nodes in the two WBANs changes from 5 to 15. Discount factor γ is set as 0.9. Learning rate of DQNPC is set as 0.5.

5.2 Numerical Result

Figure 4 presents the accuracy performance of DQNPC in 4 epochs. The blue full line is the training accuracy and the black dotted line is the validation

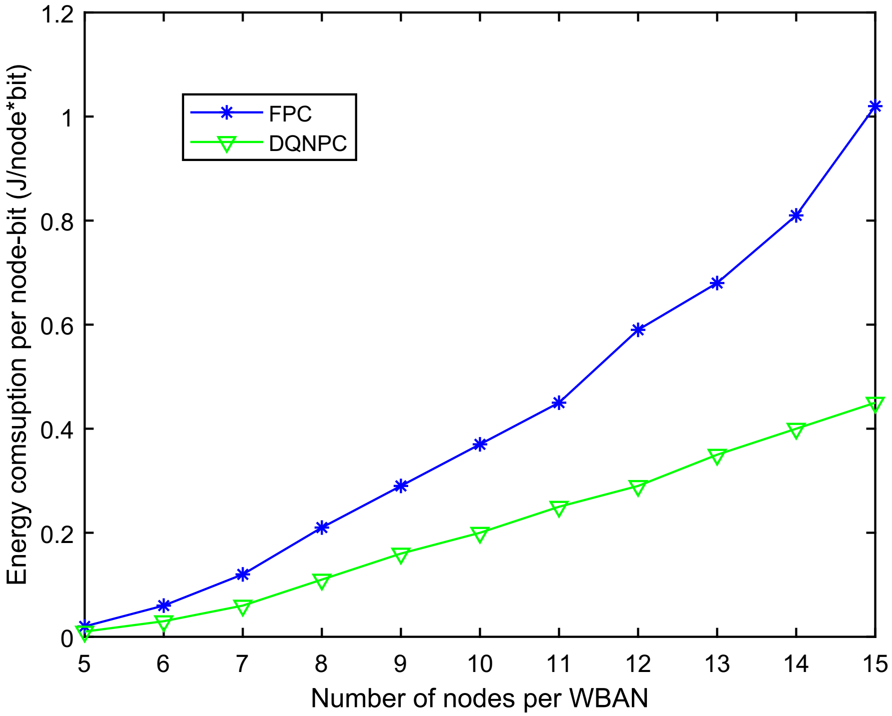

Fig. 5. Average energy consumption per bit and node with increase number of nodes

accuracy. It can be clearly seen that training accuracy has some fluctuation with the increase of epochs, and finally settles in 100% in the third epoch. Differently, validation accuracy increase smoothly and settles in 86.67%. The results indicate that our proposed DQNPC is feasible and with low complexity.

Figure 5 shows how energy consumes in transmission of unit bit for each node. The result reveals that applying FPC will lead the rapid increase of energy consumption for WBANs, since increase of the nodes results in heavier interference, and nodes may spend more time and energy for successful communication. However, proposed DQNPC is more intelligent to dynamically alleviate the inter-network interference. As a results, the energy consumptions decrease.

Figure 6 shows the relation between average node throughput and number of nodes per WBAN. It can be seen that node throughput decrease with the increase of node number due to the increased interferences. By comparing the performance of two algorithms, we can see that throughput of FPC is a little higher than that of DQNPC. Such difference of node throughput decrease with the number of nodes, and become unconspicuous when node number is large. We infer that DQNPC increases the performance of energy consumption by sacrificing limit performance of throughput.

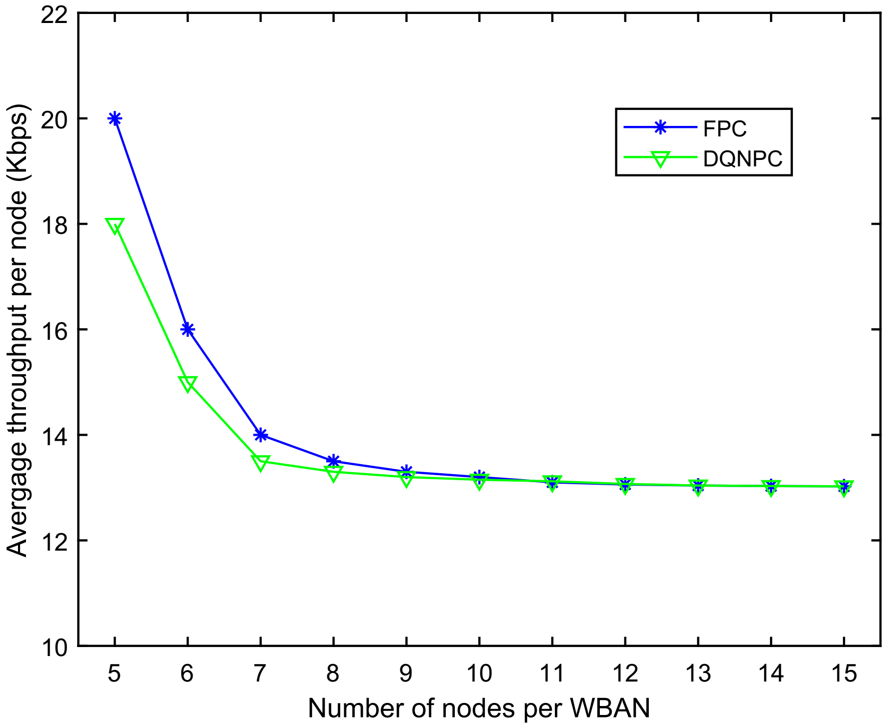

Fig. 6. Average node throughput with increase number of nodes

6 Conclusion

In this work, we studied the dynamic power controlling scheme of WBAN. Two WBANs were configured in our proposed network, and we analzed the inter-network interference. We proposed a Deep-Q-Network-based Power Controller (DQNPC) to address the energy and interference problems of WBAN. The results indicate that the proposed power controller is feasible to configured in WBANs. The results also reveal that proposed power controller decreases the energy assumption by sacrificing little performance of throughput. In the future, we will study the performance of the DQNPC in more WBANs.

References

1. Naranjo-Hernàndez, D., Callejón-Leblic, A., Lučev Vasić, Ž., Khan, M.A., et al.: Past results, present trends, and future challenges in intrabody communication. Wirel. Commun. Mob. Comput. **2018**, 1–40 (2018)
2. Luong, N.C., Hoang, D.T., Gong, S., et al.: Applications of deep reinforcement learning in communications and networking: a survey. IEEE Commun. Surv. Tutor. **21**, 3133–3174 (2019)

3. Hussain, M., Mehmood, A., Khan, S., Khan, M.A., Iqbal, Z.: A survey on authentication techniques for wireless body area networks. J. Syst. Arch. **101**, 101655 (2019)
4. Javaid, N., Abbas, Z., Fareed, M.S., Khan, Z.A., Alrajeh, N.: RE-ATTEMPT: a new energy-efficient routing protocol for wireless body area sensor networks. Procedia Comput. Sci. **19**, 224–231 (2013)
5. Wu, T., Wu, F., Redout, J.M., Yuce, M.R.: An autonomous wireless body area network implementation towards IoT connected healthcare applications. IEEE Access **5**, 11413–11422 (2017)
6. Mohamed, M., Joseph, W., Vermeeren, G., Tanghe, E., Cheffena, M.: Characterization of dynamic wireless body area network channels during walking. EURASIP J. Wirel. Commun. Netw. **2019**(1), 1–12 (2019). https://doi.org/10.1186/s13638-019-1415-3
7. Moosavi, H., Bui, F.M.: Optimal relay selection and power control with quality-of-service provisioning in wireless body area networks. IEEE Trans. Wirel. Commun. **15**(8), 5497–5510 (2016)
8. Yang, Y., Smith, D.B., Seneviratne, S.: Deep learning channel prediction for transmit power control in wireless body area networks. In: International Conference on Communications (ICC), pp. 1–6. IEEE, Shanghai (2019)
9. Kazemi, R., Vesilo, R., Dutkiewicz, E., Liu, R.: Dynamic power control in wireless body area networks using reinforcement learning with approximation. In: International Symposium on Personal, Indoor and Mobile Radio Communications, pp. 2203–2208. IEEE, Toronto (2011)
10. Puterman, M.L.: Markov Decision Processes: Discrete Stochastic Dynamic Programming. Wiley, Hoboken (2014)
11. Hausknecht, M., Stone, P.: Deep recurrent Q-learning for partially observable MDPS. In: AAAI Fall Symposium Series (2015)
12. Dabney, W.C.: Adaptive step-sizes for reinforcement learning. Ph.D. dissertation (2014)
13. Watkins, C.J., Dayan, P.: Q-learning. Mach. Learn. **8**(3–4), 279–292 (1992). https://doi.org/10.1007/BF00992698
14. Hu, J., Wellman, M.P.: Nash Q-learning for general-sum stochastic games. J. Mach. Learn. Res. **4**(Nov), 1039–1069 (2003)
15. Mismar, F.B., Brian, L.E., Ahmed, A.: Deep reinforcement learning for 5G networks: joint beamforming, power control, and interference coordination. arXiv preprint arXiv:1907.00123 (2019)