DDPG-Based Throughput Optimization with AoI Constraint in Ambient Backscatter-Assisted Overlay CRN

Xueli Jia, Kechen Zheng *, Kaikai Chi and Xiaoying Liu

School of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310023, China; 2111912158@zjut.edu.cn (X.J.); kkchi@zjut.edu.cn (K.C.); xiaoyingliu@zjut.edu.cn (X.L.)

* Correspondence: kechenzheng@zjut.edu.cn

Abstract: The combination of ambient backscatter (AB) communications (ABCs) and RF-powered cognitive radio networks (CRNs) deals with challenges of both energy supply and spectrum shortage, and improves the network performances. With the expansion of wireless networks, many applications raise requirements for both high-throughput and timely data. Driven by these facts, we study the long-term throughput optimization of the secondary network in the AB-assisted overlay CRN (ABO-CRN), ABCs, and CRNs with the age of information (AoI) constraint, which is a novel metric for measuring the freshness of data received by receivers. Due to the dynamic environment, complete knowledge of the environment could not be obtained. Then, the deep deterministic policy gradient (DDPG), a deep reinforcement learning (DRL) method that addresses decision issues in both continuous and discrete spaces, is deployed to address the throughput optimization. We consider the impacts of time and energy allocation on the reward when the AoI constraint can not be satisfied, and develop the corresponding reward functions. Furthermore, we analyze the impacts of the minimum throughput requirement and maximum allowable AoI on the throughput and AoI of the secondary networks in the ABO-CRN, ABCs, and CRNs. We compare the throughput optimization scheme under the AoI constraint with two baseline schemes (i.e., throughput-optimal (T-O) and AoI-optimal (A-O) baseline schemes), and the simulation results show that the throughput of the ABO-CRN is close to the optimal throughput of the T-O baseline scheme, and the AoI of the ABO-CRN is close to the optimal AoI of the A-O baseline scheme.

Keywords: cognitive radio networks; ambient backscatter; age of information; DDPG

1. Introduction

Nowadays, the number of wireless devices has increased year by year, and the amount of spectrum demands has also increased [1]. However, the amount of spectrum is limited, and the majority of spectrum has been allocated to licensed users (primary users, PUs) as the licensed spectrum. PUs occupy the licensed spectrum at a certain time and places, and the licensed spectrum may not be occupied for a long time. In this case, the utilization ratio of the licensed spectrum is low [2]. In addition, due to the limited size, wireless devices can not carry large-capacity batteries, and frequent replacement of batteries is not allowed under certain circumstances, such as the inside of chimneys and the inside of bodies [3]. Therefore, issues of spectrum shortage and energy supply attract a large number of researchers. The RF-powered ambient backscatter (AB)-assisted cognitive radio networks (CRNs) (AB-CRNs) are introduced to improve the spectrum utilization ratio and alleviate the difficulty of energy supply. We introduce the RF-powered AB-CRNs from the following three aspects: the energy harvesting technology, the CRNs, and the AB communications (ABCs).

Energy harvesting technology, which allows wireless devices to scavenge energy from the environmental energy sources, is an efficient way to address the difficulty of energy supply [2,4]. The energy sources can be solar, wind, heat, and RF signals, etc. Some energy sources are unstable and conditionally available. Taking solar as an example,
harvesting energy from solar depends on weather conditions, and can be used during the daytime. In addition, the devices powered by solar are bulky and costly [5], and most wireless devices can not carry huge solar panels. Differently, harvesting energy from RF signals has lots of advantages, i.e., stability, ubiquity, and controllability [6], and has been applied in a variety of wireless network fields. In [7], Yang et al. investigated the application of energy harvesting technology in wireless body area networks. Chi et al. [8] aimed to minimize the RF energy provision of time-division-multiple-access (TDMA) and non-orthogonal-multiple-access (NOMA) schemes in wireless powered communication network. In [9], Hoang et al. explored security and energy harvesting requirements in MIMO-OFDM networks. Authors in [10] and [11] applied energy harvesting technology in wireless sensor networks (WSN). Azarhava et al. [10] studied the energy harvesting WSN with TDMA scheme. Ghosh et al. [11] maximized the energy efficiency by proposing the upper confidence bound based algorithm.

Cognitive radio (CR) technology is introduced to deal with the issue of spectrum shortage. Secondary users (SUs), i.e., cognitive users, are allowed to access the licensed spectrum opportunistically without causing harmful influence to the PUs, which improves the spectrum utilization ratio [12]. SUs are able to execute various transmission modes in the CRNs, such as overlay mode and underlay mode. When secondary transmitters (STs) transmit data in overlay mode, they exclusively occupy the licensed spectrum when PUs are not using the licensed spectrum, and immediately leave the licensed spectrum when PUs begin to transmit. When STs transmit data in underlay mode, STs could access the licensed spectrum even when the licensed spectrum is used by PUs. However, the secondary transmission should not lower the transmission quality of PUs, hence the transmit power of STs can not exceed the power threshold that PUs tolerate [13]. RF-powered CRNs allow SUs to harvest energy from RF signals, and access the licensed spectrum following various transmission modes. As a result, RF-powered CRNs are able to alleviate both spectrum shortage and energy insufficiency, and have now been extensively studied. In [14], Zheng et al. divided a unit circular area into three concentric circle regions: overlay region, underlay region, and energy harvesting region, and examined spectrum-limited and energy-limited scenarios. In the cooperative CRN, Papadopoulos et al. [15] enhanced the overall throughput performance by two network coding algorithms. In [16], Rathee et al. considered security threats in the CRNs, and proposed a secure hand-off mechanism for the emulation attack of cognitive users.

Although the RF-powered CRNs have excellent performance in solving the problems of spectrum shortage and low utilization ratio, as well as energy supply difficulties, whether the transmission performance of the CRNs could be improved depends on the energy harvesting from the RF signals of primary users and the spectrum usage [17]. For example, if a ST harvests little energy, or its energy storage capacity is too small to store sufficient energy, the ST lacks energy to transmit data, or spends much time harvesting energy, degrading the throughput performance of the secondary network. Furthermore, if the PU occupies the licensed spectrum with high probability, the ST has few opportunities to transmit data. Driven by these facts, AB technology is proposed to alleviate difficulties of both spectrum shortage and energy insufficiency. AB mode transmission is performed by using ambient RF signals, such as TV signals, and it does not need extra spectrum. In addition, after receiving the RF signal, the devices with AB functionality map the data to a series of reflective states by adjusting the load impedance of the antenna, and then reflect the RF signals to the receiver according to the matching degree between the antenna impedance and the load impedance. The operation of reflection consumes little energy [18]. Ye et al. [19] investigated the ABCs with multiple channel links, and minimized the outage probability. Liu et al. [20] utilized coherent and non-coherent orthogonal space-time block codes to improve the backscatter efficiency in ABCs with multiple antennas. In [21], Madavani et al. studied a full-duplex ABC, and optimized the throughput of the minimum AB device and the overall throughput.

However, throughput of AB mode transmission is relatively low, especially when the primary RF signals are weak. Considering the characteristics of CR technology and
AB technology, the combination of AB technology and RF-powered CRNs is potentially helpful for dealing with spectrum shortage and energy insufficiency. The RF-powered AB-CRN are energy-saving and spectrum-saving [22]. As far as we know, Hoang et al. in [22] initially introduced the RF-powered AB-CRN, and analyzed the throughput of the AB-assisted overlay CRN (ABO-CRN) and the AB-assisted underlay CRN (ABU-CRN). Extended from [22], Zhuang et al. [23] studied the RF-powered AB-CR-NOMA networks, where STs perform underlay mode transmission. In [24], Zhu et al. investigated the distributed resource allocation in AB-CRN.

In [22,23], authors investigated popular metrics, such as throughput. Besides, the age of information (AoI), a novel metric that measures the freshness of data received at the receiver, has also been extensively studied in recent years [25]. Different from delay metric, AoI focuses on the data timeliness of the receiver, while delay metric focuses on that of the transmitter. Authors in [26–28] investigated the AoI minimization in the CRNs. In [26], Leng et al. utilized the partially observation Markov decision process to analyze the AoI performance in several cases. In [27], Gu et al. studied the AoI in overlay and underlay scenarios, and analyzed the effect of critical generation rate of the primary IoT on the secondary IoT. In [28], Wang et al. took the collision constraint into account to minimize the long-term average AoI. The AoI minimization challenge in ABCs is also a fascinating and important research topic. In [29], Abbas et al. focused on the minimization of AoI in the backscatter communications, and introduced several algorithms for the AoI minimization.

With the expansion of wireless networks, timely delivery is required [30], and in the scenarios, such as cyber-physical system, low-quality but timely data is useless [31]. However, the throughput optimization can not guarantee the freshness of data, and AoI optimization can not guarantee the quality of data. Liu et al. [32] investigated the AoI minimization under throughput requirements in the multi-path network. Kadota et al. [33] proposed a low-complexity scheduling algorithm for the AoI minimization with throughput constraints of the wireless network. Bhat et al. [34] studied the throughput maximization under the AoI constraint in fading channels.

Obviously, compared with the optimization of a single frame, long-term optimization is more practical. In practice, the network environment is dynamic, and the channel quality, such as the channel gain, varies with the frame. The optimization of a single frame is limited by the channel quality, and affects that of the subsequent frames. The short-term optimization ignores the connection between the optimization of the current frame and that of subsequent frames, which degrades the performance of the network. However, the long-term optimization takes the aforementioned connection into account, and provides more practical decision to enhance the performance of the network. Taking the throughput optimization of a single frame by energy management as the example, the throughput increases with the consumed energy, hence consuming all the available energy of the frame is optimal for the short-term optimization. However, consuming all the available energy in the frame with poor channel quality leads to the lack of available energy in the frame with good channel quality. Therefore, the long-term optimization is more practical than the short-term optimization. Due to the dynamic and uncertain network parameters, the complete knowledge about the network could not be obtained in advance. Some traditional methods are incapable of addressing challenges with too many dynamic and unpredictable environmental parameters. Deep reinforcement learning (DRL) has been proved as an effective way to tackle the challenge [35,36]. When applying value-based DRLs, the action space for DRLs has to be discrete. If the discrete methods are improper for the scenario, vital information may be lost, or the action space dimension may be too large [37]. Different from value-based DRLs, DRLs based on the policy gradient are able to deal with the problems of continuous spaces, and have no need to discretize the action space. Policy gradient-based DRLs have been applied into the field of the wireless networks [38].

Taking the limitations summarized in Table 1 into account, we conclude the novel-ties and contributions as follows. Considering the problems of spectrum shortage and energy insufficiency, we focus on AB-CRN, while a majority of researches in AB-CRN evaluated the network performances such as throughput, energy consumption, delay, etc.,
and ignored the data freshness of the secondary receivers (SRs). Driven by the fact, we optimize the long-term throughput of the secondary network in the ABO-CRN with the AoI constraint, in order to guarantee the high throughput and data freshness of the secondary network. According to our knowledge, we are the first to study the optimization of both throughput and AoI in the research area of AB-CRNs. The main contributions are summarized as follows.

- In order to achieve the long-term throughput optimization of the secondary network with the AoI constraint, we utilize deep deterministic policy gradient (DDPG), a DRL based on the policy gradient, to find the optimal policy for jointly managing time and energy of STs. Considering the impacts of time and energy allocation on the reward when the AoI constraint cannot be satisfied, we develop the corresponding reward functions with respect to the channel states.
- We analyze the minimum throughput requirement and the maximum allowable AoI for the throughput and AoI performances in the ABO-CRN, ABCs, and CRNs.
- We introduce throughput-optimal (T-O) and AoI-optimal (A-O) baseline schemes as comparisons for the throughput optimization with the AoI constraint. The simulation results show that the throughput of the ABO-CRN is close to the optimal throughput of the T-O baseline scheme, and the AoI of the ABO-CRN is close to the optimal AoI of the A-O baseline scheme.
- We evaluate the impacts of the minimum throughput requirement and maximum allowable AoI on the throughput and AoI performances of the secondary networks in the ABO-CRN, ABCs, and CRNs, and demonstrate that the ABO-CRN improves the throughput and AoI performances of the ABCs and CRNs.

Table 1. Comparison Table of the Related works.

| Scenario   | Metric                  | Limitations                                                                 |
|------------|-------------------------|-----------------------------------------------------------------------------|
| CRNs       | Throughput [14,15], Aol [26–28] | Short-term optimization [14,15,27], single ST [14,15,26–28], single metric optimization, single resource management [14,15,26,27]. |
| ABCs       | Outage probability [19], backscatter efficiency [20], throughput [21], Aol [29] | Short-term optimization, single resource management [19–21], single metric optimization [19–21,29] |
| ABO-CRNs   | Throughput [22,23], coverage probability [24] | Short-term optimization [22,23], single ST [22,23,26], single metric optimization [22–24], single resource management [22,23]. |

The remainder of this paper is organized as follows. In Section 2, we introduce the network model and operations of STs in the ABO-CRN, ABCs, and CRNs. In Section 3, we introduce the problem formulation, such as throughput and AoI definitions. In Section 4, we utilize DDPG to find the optimal policy for jointly managing time and energy of STs. In Section 5, simulation results are shown. In Section 6, we conclude the paper.

2. System Model

In this section, we first depict the structures and channel models of the ABO-CRN, ABCs, and CRNs, and then introduce the network models of the ABO-CRN, ABCs, and CRNs, respectively.

2.1. Structures and Channel Models

The ABO-CRN, ABCs, and CRNs are composed of a primary network and a secondary network. The secondary network consists of a SR and \( n + 1 \) STs, \( n \in \{0,1,\ldots\} \). In the
primary network, the primary transmitter (PT) utilizes the licensed channel to transmit data. The probability that the PT occupies the channel in each frame, denoted by $P_a$, can be obtained through the long time observation. When the PT transmits data in frame $t \in \{1, 2, \ldots, K\}$, the channel state is active, denoted by $s^a_t = 1$. When the PT does not transmit data in frame $t$, the channel state is inactive, denoted by $s^a_t = 0$.

In the secondary network, as the random distribution of the SUs in [39], the SR and STs are randomly placed within the coverage of the primary RF signals from the PT, as shown in the Figure 1a,c. Each ST is equipped with a single antenna and a rechargeable capacitor with finite capacity $E$. The SR is equipped with a single antenna and a wired energy source, hence there is no need to consider the energy supply of the SR. In order to measure the freshness of the data that the SR receives from STs, SR has capability to record the AoI of the data from each ST. Similar as [40], the SR plays the role of center controller to manage the time and energy for STs. At the very beginning of the frame, the SR senses the channel, and then provides the allocation of time and energy for STs.

In the considered scenario, frames with equal duration are successive. The frame duration is synchronized with the primary network, and without loss of generality, we normalize the frame duration as 1 [2]. As shown in Figures 1b,d, 2b and 3b,d, each frame consists of one or more slots. The duration of each slot is determined by the SR. We consider that the channel state remains unchanged in one frame, but varies in subsequent frames. When $s^a_t = 1$, the SR and STs receive stable and continuous RF signals from the PT. In frame $t$, the channel gain between the PT and ST$_i$, $i = 0, 1, \ldots, n$, is denoted by $g^a_t, i$, and that between ST$_i$ and the SR is denoted by $h^a_t$. The Rayleigh distribution [41] is used to formulate the channel gains that remain unchanged in one frame, and the channel noise is modeled as Additive White Gaussian Noise (AWGN) with variance $\delta^2$.

2.2. Network Models

2.2.1. Network Model of ABO-CRN

The operations executed by the STs in the ABO-CRN depend on the value of $s^a_t$. When $s^a_t = 1$, as shown in Figure 1a,b, STs execute AB mode transmission by TDMA scheme, and ST$_i$ harvests energy when the other STs transmit data in AB mode. We consider the scenario where the energy consumption of AB mode transmission is negligible. Therefore, no dedicated slot for energy harvesting is required by STs. When $s^a_t = 0$, as shown in Figure 1c,d, following the TDMA scheme, STs execute overlay mode transmission by consuming the energy stored in the rechargeable capacitor. With the aim to optimize the long-term average throughput with the AoI constraint, ST$_i$ may not consume all the available energy $\epsilon_{t,i}$ during frame $t$. The energy $\epsilon_{t,i}$ consumed by ST$_i$, denoted by $e_{t,i}$, is determined by the SR.

We provide a flow chart of the ABO-CRN in Figure 4. The actions in Figure 4 are executed by STs according to $s^a_t$. The SR decides the time and energy allocation of STs according to the channel information and states of each ST. The channel information includes the channel state $s^a_t$ and channel gains. The states of each ST include the available energy and the AoI of the current frame. Note that, the feedback information in Figure 4 includes two parts. The first part is the new energy state in each ST, which is the available energy of the next frame. The second part is the received reward after STs execute the actions decided by the SR.

2.2.2. Network Model of ABCs

The operations executed by the STs in the ABCs depend on the value of $s^a_t$. When $s^a_t = 1$, as shown in Figure 2, STs take turns to execute AB mode transmission by TDMA scheme. When $s^a_t = 0$, since the PT does not broadcast RF signals on the current channel, STs do not execute AB mode transmission.
Figure 1. System model and frame structure of the AB-assisted overlay CRN: (a) depicts the system model when $s_t = 1$, i.e., the channel is active. (b) depicts the time frame structure when $s_t = 1$. (c) depicts the system model when $s_t = 0$, i.e., the channel is inactive. (d) depicts the time frame structure when $s_t = 0$.

Figure 2. System model and frame structure of the ABCs: (a) depicts the system mode when $s^2_t = 1$. (b) depicts the time frame structure when $s^2_t = 1$. 
Figure 3. System model and frame structure of the CRNs: (a) depicts the system model when $s^t_i = 1$. (b) depicts the time frame structure when $s^t_i = 1$. (c) depicts the system model when $s^t_i = 0$. (d) depicts the time frame structure when $s^t_i = 0$.

Figure 4. Flow chart of the AB-assisted overlay CRN.
2.2.3. Network Model of CRNs

The operations executed by the STs in the CRNs depend on the value of $s^t_i$. When $s^t_i = 1$, as shown in Figure 3a,b, STs harvest energy. When $s^t_i = 0$, as shown in Figure 3c,d, following the TDMA scheme, STs execute overlay mode transmission by consuming the energy stored in the rechargeable capacitor. The energy $e_{t,i}$ consumed by ST$_i$ is determined by the SR.

3. Formulation and Analysis of the Problem

For the readability, we provide a parameter list in Table 2 that summarizes the main parameters and meanings.

Table 2. Parameter list.

| Parameter | Description |
|-----------|-------------|
| $n$       | The number of STs is $n + 1$ |
| $s^t_i$   | The channel state in frame $t$ |
| $P_a$     | The probability of the active channel state |
| $E$       | The capacity of rechargeable capacitor |
| $e_{t,i}$ | The allocated energy for overlay mode transmission of ST$_i$ |
| $e_{t,i}$ | The available energy of ST$_i$ in frame $t$ |
| $a_{t,i}$ | The duration of data transmission by ST$_i$ in frame $t$ |
| $T_i$     | The total throughput of secondary network in frame $t$ |
| $T_i^A$   | The throughput of ST$_i$ in frame $t$ |
| $T_i^O$   | The throughput of ST$_i$ by overlay mode transmission |
| $T_{min}$ | The minimum throughput requirement for each ST |
| $W$       | The bandwidth |
| $P$       | The transmit power of the PT |
| $g_{t,i}$ | The channel gain from the PT to ST$_i$ in frame $t$ |
| $h_{t,i}$ | The channel gain from ST$_i$ to gateway in frame $t$ |
| $\theta$  | The backscatter reflection coefficient |
| $\delta^2$ | The variance of AWGN |
| $a_{t,i}$ | The AoI of ST$_i$ in frame $t$ |
| $A_{max}$ | The maximum allowable AoI |

3.1. Throughput Definition

The total throughput $T_i$ of the secondary network in the ABO-CRN, ABCs, and CRNs in frame $t$ can be expressed as

$$T_i = \sum_{j=0}^{n} T_{i,j}, \quad (1)$$

where $T_{i,j}$ denotes the throughput of ST$_i$ in frame $t$. Due to the fact that the operations executed by STs depend on the channel state $s^t_i$, the calculation of $T_{i,j}$ depends on the value of $s^t_i$.

3.1.1. Throughput Definition of ABO-CRN

In the ABO-CRN, $T_{i,j}$ is expressed as

$$T_{i,j+1} = \begin{cases} T_{i,j}^A, & s^t_i = 1; \\ T_{i,j}^O, & s^t_i = 0. \end{cases} \quad (2)$$

When $s^t_i = 1$, according to the Shannon Theory, the throughput of ST$_i$ by AB mode transmission, denoted by $T_{i,j}^A$, is expressed as

$$T_{i,j}^A = a_{t,i} W \log_2 \left(1 + \frac{\theta P g_{t,i} h_{t,i}}{\delta^2}\right), \quad (3)$$
where $\alpha_{t,j} \in [0, 1]$ denotes the duration of data transmission by $ST_j$ through AB mode, $W$ denotes the bandwidth, $\theta \in [0, 1]$ denotes the backscatter reflection coefficient that depends on the electronic component factors, $P$ denotes the transmit power of the PT, and $g_{t,j}$ denotes the channel gain from the PT to $ST_j$, and $h_{t,j}$ denotes the channel gain from $ST_j$ to the SR. In particular, $\theta P g_{t,j}$ represents the transmit power of $ST_j$ for AB mode transmission. $ST_j$ harvests energy when the other STs transmit data in AB mode. The harvested energy of $ST_i$, denoted by $\epsilon_{t,j}^h$, is calculated as

$$\epsilon_{t,j}^h = \min \left\{ \sum_{i=0, i \neq j}^n \alpha_{t,j} P g_{t,j}, E - \epsilon_{t,j} \right\}.$$  \hspace{1cm} (4)

After energy harvesting, the available energy in $ST_j$ of frame $t + 1$, denoted by $\epsilon_{t+1,j}$, is updated as

$$\epsilon_{t+1,j} = \min \left\{ \epsilon_{t,j} + \epsilon_{t,j}^h, E \right\}.$$ \hspace{1cm} (5)

When $s_t^j = 0$, according to the Shannon Theory, the throughput of $ST_j$ by overlay mode transmission, denoted by $T_{t,j}^O$, is expressed as

$$T_{t,j}^O = \alpha_{t,j} W \log_2 \left( 1 + \frac{\epsilon_{t,j} h_{t,j}}{\alpha_{t,j} g_{t,j}} \right),$$ \hspace{1cm} (6)

where $\alpha_{t,j}, \theta, P, g_{t,j}$, and $h_{t,j}$ represent the same meaning as that in Equation (3). Since the energy consumption of AB mode transmission is negligible, the energy update is not considered in the ABCs.

3.1.2. Throughput Definition of ABCs

In the ABCs, the throughput of STs is achieved by AB mode transmission when the channel state is active. Therefore, when $s_t^j = 0$, $T_{t,j} = 0$ holds, and when $s_t^j = 1$, according to the Shannon Theory, $T_{t,j}$ is expressed as

$$T_{t,j} = T_{t,j}^A = \alpha_{t,j} W \log_2 \left( 1 + \frac{\theta P g_{t,j} h_{t,j}}{\delta^2} \right),$$ \hspace{1cm} (8)

where $\alpha_{t,j}, \theta, P, g_{t,j}$, and $h_{t,j}$ represent the same meaning as that in Equation (3). Since the energy consumption of AB mode transmission is negligible, the energy update is not considered in the ABCs.

3.1.3. Throughput Definition of CRNs

In the CRNs, the throughput of STs is achieved by overlay mode transmission when the channel is inactive. When the channel is active, STs harvest energy from the RF signal of the PT. Therefore, when $s_t^j = 1$, $T_{t,j} = 0$ holds, and $\epsilon_{t+1,j}$ in $ST_j$ of frame $t + 1$ is updated as

$$\epsilon_{t+1,j} = \min \left\{ \epsilon_{t,j} + \epsilon_{t,j}^h, E \right\}.$$ \hspace{1cm} (9)

When $s_t^j = 0$, according to the Shannon Theory, $T_{t,j}$ is expressed as

$$T_{t,j} = T_{t,j}^O = \alpha_{t,j} W \log_2 \left( 1 + \frac{\epsilon_{t,j} h_{t,j}}{\alpha_{t,j} g_{t,j}} \right).$$ \hspace{1cm} (10)

$\epsilon_{t+1,j}$ in $ST_j$ of frame $t + 1$ is updated as that in Equation (7).
3.2. Definition of AoI

AoI is a novel metric to measure the freshness of data received by the receiver. In particular, AoI is used to track the time elapsed since the time point of the latest data generation to the time point that the latest data is successfully received by the receiver [33]. We utilize the linear scheme to calculate AoI of STs, where the AoI is updated as

\[ a_{t+1,i} = \begin{cases} 1, & \lambda_{t,i} = 1; \\ a_{t,i} + 1, & \lambda_{t,i} = 0, \end{cases} \]

(11)

where \( a_{t,i} \) denotes the AoI of ST \( i \) in frame \( t \), \( \lambda_{t,i} = 1 \) indicates that the latest data of ST \( i \) is successfully received by the SR, and \( \lambda_{t,i} = 0 \) indicates that the latest data of ST \( i \) is not successfully received by the SR. With the aim to optimize the long-term average throughput of the secondary network with the AoI constraint, we set a minimum throughput requirement \( T_{\min} \) for every ST. Specifically, if the throughput of ST \( i \) during frame \( t \) is no less than \( T_{\min} \), the latest transmitted data of ST \( i \) is considered to be successfully received by the SR. Based on Equation (11) and the aforementioned analysis of \( \lambda_{t,i} \), \( \lambda_{t,i} \) is expressed as

\[ \lambda_{t,i} = \begin{cases} 1, & T_{t,i} \geq T_{\min}; \\ 0, & \text{else.} \end{cases} \]

(12)

By combining Equations (11) and (12), the update of AoI is calculated as

\[ a_{t+1,i} = \begin{cases} 1, & T_{t,i} \geq T_{\min}; \\ a_{t,i} + 1, & \text{else.} \end{cases} \]

(13)

Obviously, when \( s_i^e = 0 \) in the ABCs, STs achieve negligible throughput, hence the throughput of each ST can not exceed \( T_{\min} \), \( a_{t+1,i} = a_{t,i} + 1 \) holds. When \( s_i^e = 1 \) in the CRNs, the same conclusion holds.

3.3. Problem Formulation

The throughput optimization objective function of the ABO-CRN, ABCs, and CRNs is expressed as

Maximize \( \mathcal{T} = \lim_{K \rightarrow \infty} \frac{1}{K} \mathbb{E} \left( \sum_{t=1}^{K} T_t \right) \)

(14a)

s.t.: \( 0 \leq a_{t,i} \leq 1 \) and \( \sum_{i=0}^{n} a_{t,i} \leq 1 \),

(14b)

\( 0 \leq e_{t,i} \leq e_{t,i} \),

(14c)

\[ \lim_{K \rightarrow \infty} \frac{1}{K(n+1)} \mathbb{E} \left( \sum_{t=1}^{K} \sum_{i=0}^{n} a_{t,i} \right) \leq A_{\max}, \]

(14d)

where \( A_{\max} \) denotes the maximum allowable AoI that the secondary network tolerates, and Equation (14d) indicates that the average accumulated AoI should be smaller than \( A_{\max} \). Since the energy consumed by AB mode transmission of STs is negligible, the SR in the ABCs does not need consider the constraint in Equation (14c).

3.4. Analysis of \( T_{\min} \) and \( A_{\max} \)

In this subsection, we analyze \( T_{\min} \) and \( A_{\max} \) in the ABO-CRN, ABCs, and CRNs. The expectation of the long-term average throughput in the ABO-CRN, ABCs, and CRNs is expressed as

\[ \mathbb{E}(T_{t,i}) = \lim_{K \rightarrow \infty} \frac{1}{K(n+1)} \mathbb{E} \left( \sum_{t=1}^{K} \sum_{i=0}^{n} T_{t,i} \right). \]

(15)
According to Equations (2), (8) and (10), we have

\[
\mathbb{E}\left(\sum_{i=1}^{K} \sum_{j=0}^{n} T_{i,j}^s\right) = \begin{cases} 
\mathbb{E}\left(\sum_{i=1}^{K} \sum_{j=0}^{n} \left( P_{a} T_{i,j}^A + (1 - P_{a}) T_{i,j}^O \right)\right) & , \text{ABO-CRN;} \\
\mathbb{E}\left(\sum_{i=1}^{K} \sum_{j=0}^{n} P_{a} T_{i,j}^A\right) & , \text{ABCs;} \\
\mathbb{E}\left(\sum_{i=1}^{K} \sum_{j=0}^{n} (1 - P_{a}) T_{i,j}^O\right) & , \text{CRNs.}
\end{cases}
\]

\(T^−\) denotes the long-term average throughput of each ST whose average throughput is smaller than \(T_{\text{min}}\), \(T^+\) denotes that is no smaller than \(T_{\text{min}}\), \(a\) denotes the long-term average AoI of each ST whose average throughput is smaller than \(T_{\text{min}}\), and let \(N\) equal \(n + 1\).

**Lemma 1.** With \(\mathbb{E}(T_{i,j}) \leq T_{\text{min}}\) holds, when \(T^−\) is closer to \(T_{\text{min}}\), the tolerable value interval of \(a\) is larger. When \(a > A_{\text{max}} - 1\), the AoI constraint can not be satisfied.

**Proof.** We assume there are \(x\) STs with \(T^−\), and \(N - x\) STs with \(T^+\). We have

\[
\mathbb{E}(T_{i,j}) = \frac{x(T^−) + (N - x)T^+}{N} = \frac{x(T^− - T^+) + NT^+}{N}.
\]

(17)

Since \(\mathbb{E}(T_{i,j}) \leq T_{\text{min}}\) holds, we have

\[
\mathbb{E}(T_{i,j}) - T_{\text{min}} = \frac{x(T^− - T^+) + NT^+}{N} - T_{\text{min}} \leq 0, \text{ and } x \geq \frac{N(T^− - T^+)}{T^− - T^+}.
\]

(18)

In order to satisfy the AoI constraint, Equation (14d) is updated to \(\frac{x(a + 1) + N - x}{N}\), and we have

\[
\frac{x(a + 1) + N - x}{N} - A_{\text{max}} \leq 0, \text{ and } a \leq \frac{(A_{\text{max}} - 1)N}{x}.
\]

(19)

Bring \(x = \frac{N(T^− - T^+)}{T^− - T^+}\) into Equation (19), and we have

\[
a \leq (A_{\text{max}} - 1)\frac{T^− - T^+}{T^− - T^+}.
\]

(20)

Obviously, when \(T^−\) is closer to \(T_{\text{min}}\), the tolerable value interval of \(a\) is larger.

Since \(T^− < T_{\text{min}}\) holds, the \(\frac{T^− - T^+}{T^− - T^+} < 1\) holds, hence \(a\) can not exceed \(A_{\text{max}} - 1\). Therefore, when \(a > A_{\text{max}} - 1\), the AoI constraint can not be satisfied. The proof is completed. \(\square\)

**Lemma 2.** The lower bound of \(A_{\text{max}}\) that makes STs satisfy the Equation (14d) decreases with \(n\), and increases with the number of STs whose average throughput is smaller than \(T_{\text{min}}\).

**Proof.** We assume \(x\) is the number of STs whose average throughput is smaller than \(T_{\text{min}}\), \(a\) has been given. From Equation (19), we deduce

\[
A_{\text{max}} \geq \frac{ax}{n + 1} - 1.
\]

(21)

With the larger value of \(n\), the lower bound of \(A_{\text{max}}\) that makes STs satisfy the Equation (14d) decreases. With the larger value of \(x\), the lower bound for \(A_{\text{max}}\) increases. The proof is completed. \(\square\)

Then we compare the impacts of \(T_{\text{min}}\) and \(A_{\text{max}}\) on the ABO-CRN, ABCs, and CRNs. We discuss the impacts in some extremely cases, i.e., \(P_{a}\), the probability of \(x_i^a = 1\), is relatively small or relatively large. In the ABCs, when \(P_{a}\) is relatively small, there are few opportunities for AB mode transmission. In the CRNs, when \(P_{a}\) is relatively large, there
are few opportunities for overlay mode transmission. In these two cases, $E(T_{t,i})$ is small, and the lower bound of $T_{min}$ is low, and the tolerable value interval and $A_{max}$ is small.

Different from the ABCs and CRNs, when $s_a = 1$, the STs in the ABO-CRN execute AB mode transmission, and when $s_a = 0$, STs execute overlay mode transmission. As described in Equation (16), $E(T_{t,i})$ of the ABO-CRN is higher than that of the ABCs and CRNs. Under the same conditions of $P_a$, $T_{min}$, and $A_{max}$, the ABO-CRN achieves higher throughput while satisfying the AoI constraint.

4. Policies of Time and Energy Management

As described in the introduction, long-term optimization is more practical than the optimization of a single frame. Maximizing the throughput of a single frame with the AoI constraint may not be desirable. As a result, we consider the long-term optimization of the throughput. However, since network environmental factors, such as the channel state and channel gains, are dynamic and uncertain, it is difficult for SUs to obtain complete knowledge about the network environmental factors in advance. DRL is an excellent way to tackle the challenge. For some DRLs, such as deep Q-learning network (DQN) that is based on the value-function policy, discrete spaces are necessary. If the discrete methods are not suitable for the scenario, it may lose important information, or lead to the high space dimension. Therefore, we utilize DDPG, which deals with problems of continuous spaces, to find the optimal policy of time and energy management for throughput optimization. We define the details about DDPG in the following subsections.

4.1. Definitions of Spaces and Rewards

The SR plays the role of agent that provides decisions for STs. According to Equations (2)-(14d), the state spaces, action space, and rewards are introduced as follows.

4.1.1. State Space

The SR determines time and energy allocation of STs based on the states of STs and channel information of the current frame, including the available energy in STs, the AoI about STs, channel gains, and the channel state. Therefore, the state space contains information about energy states, AoI states, states of channel gains, and channel states. The energy-state space is represented by

$$S_E = \{(ε_{t,0}, ε_{t,1}, \ldots, ε_{t,n}); 0 \leq ε_{t,i} \leq E\}. \quad (22)$$

The AoI-state space is represented by

$$S_A = \{(a_{t,0}, a_{t,1}, \ldots, a_{t,n})\}, \quad (23)$$

where the average accumulated AoI satisfies Equation (14d). In order to reduce the dimension of the channel-gain-state space, we represent the channel gains as

$$h_{t,i} = 10^{-3}η_{h,t}^{l_i^e},$$
$$g_{t,i} = 10^{-3}η_{g,t}^{L_i^e}, \quad (24)$$

where $η_{h,t}$ denotes the path loss coefficient from the PT to STs, and $η_{g,t}$ denotes the path loss coefficient from STs to the SR, $l_i$ denotes the distance between the PT and ST$_i$, and $L_i$ denotes the distance between the SR and ST$_i$, and $e$ denotes the channel path fading exponent. Therefore, the channel-gain-state space is represented by

$$S_G = \{(η_{h,t}, η_{g,t})\}, \quad (25)$$

where $η_{h,t}$ and $η_{g,t}$ follow the Rayleigh distribution. The channel-state space is expressed as

$$S_C = \{s^d_t; s^d_t \in \{0, 1\}\}, \quad (26)$$
where \( P_a \) represents the probability of \( s^a_t = 1 \).

In summary, the state space of the ABO-CRN and of the CRNs when \( s^a_t = 0 \) is expressed as

\[
S = S_E \times S_A \times S_G \times S_C.
\]  

(27)

The state space of the ABCs when \( s^a_t = 1 \) is expressed as

\[
S = S_A \times S_G \times S_C.
\]  

(28)

Note that, when \( s^a_t = 0 \), STs in the ABCs do not execute AB mode transmission. When \( s^a_t = 1 \), STs in the CRNs only harvest energy. As a result, the SR does not need to determine actions for STs, hence we do not design state space for these two cases.

4.1.2. Action Space

In the ABO-CRN, based on the state of the current frame, the SR determines the actions that STs execute in the current frame. When \( s^a_t = 1 \), STs execute AB mode transmission, and ST\(_i\) harvests energy when the other STs transmit data in AB mode. When \( s^a_t = 0 \), STs execute overlay mode transmission. We define the action space as

\[
A = \left\{ (\alpha_{t,0}, \alpha_{t,1}, \ldots, \alpha_{t,n}, e_{t,0}, e_{t,1}, \ldots, e_{t,n}); \sum_{i=0}^{n} \alpha_{t,i} \leq 1, \ 0 \leq e_{t,i} \leq \epsilon_{t,i} \right\}
\]  

(29)

In particular, when \( s^a_t = 1 \), the SR only utilizes the time \((\alpha_{t,0}, \alpha_{t,1}, \ldots, \alpha_{t,n})\) of \( A \) to address the time management for STs, which is shown as Figure 5.

![Figure 5. The action space diagram.](image)

When \( s^a_t = 1 \), STs in the ABCs execute AB mode transmission. Since the energy consumption of AB mode transmission can be ignored, the SR in the ABCs focuses on the time allocation of STs. We define the action space of the ABCs as

\[
A = \left\{ (\alpha_{t,0}, \alpha_{t,1}, \ldots, \alpha_{t,n}); \sum_{i=0}^{n} \alpha_{t,i} \leq 1 \right\}
\]  

(30)

When \( s^a_t = 0 \), STs in the CRNs execute overlay mode transmission. The action space of the CRNs is defined as Equation (29).

4.1.3. Rewards

After the SR determines the action \( x_t \) based on the state \( s_t \) of frame \( t \), an immediate reward \( r_t(s_t, x_t) \) is obtained, where \( r_t(s_t, x_t) \) represents the evaluation of choosing \( x_t \) under \( s_t \). With the aim to optimize the throughput of the secondary networks with the AoI constraint, \( r_t(s_t, x_t) \) is defined as

\[
r_t(s_t, x_t) = \frac{1}{n+1} \left( \frac{T_t}{T_{min}} - \sum_{i=0}^{n} \rho_{t,i} \right),
\]  

(31)
where $\rho_{t,j}$ denotes the penalty that is related to the AoI of ST$_i$. Since the actions vary with respect to the value of $s^q_t$, the values of $\rho_{t,j}$ vary with respect to the value of $s^q_t$. When $s^q_t = 1$, we have

$$\rho_{t,j} = \begin{cases} a_{t,j}, & T_{i,t} \geq T_{\min}; \\ \frac{a^2_{t,j}}{\omega_{t,j}}, & \text{else.} \end{cases}$$

(32)

When $s^q_t = 0$, we have

$$\rho_{t,j} = \begin{cases} a_{t,j}, & T_{i,t} \geq T_{\min}; \\ a_{t,j}m^{\alpha_{t,j}} & \text{else,} \end{cases}$$

(33)

where $m$ is a constant that is set according to $A_{\max}$, and is used to ensure that $\rho_{t,j}$ is larger than $A_{\max}$. $\omega_{t,j}a_{t,j}$ indicates that the penalty increases with $\omega_{t,j}$ and $a_{t,j}$ when $a_{t,j}$ is larger than $A_{\max}$. The SR in the ABCs does not determine actions of the energy allocation for STs, hence we set $\frac{\omega_{t,j}}{\omega}$ in Equation (33) as $\frac{\omega_{t,j}}{\omega} = 1$.

4.2. Time and Energy Management by DDPG

DDPG utilizes the architecture of the actor-critic algorithm and the scheme of DQN. Therefore, DDPG consists of two parts, actor and critic. The actor is used to output a deterministic action, and the critic is used to output an evaluation, which fits the Q-table. Both actor and critic consist of evaluated networks and target networks. The target networks make the training process more stable, and have the same structure with the evaluated networks. The evaluated network of the actor is named as the actor network, and that of the critic is named as the critic network. The target network of the actor is named as the actor target network, and that of the critic is named as the critic target network.

These networks are expressed as parametric functions. The actor network is expressed as a function mapping $s_t$ to $x_t$,

$$x_t = \Pi(s_t|\omega),$$

(34)

where $\Pi$ denotes the policy of time and energy management, and $\omega$ denotes the weights of neural network in the actor network. The critic network is expressed as an action-value function, which maps $s_t$ and $x_t$ to a Q-value,

$$Q = Q(s_t, x_t|\mu),$$

(35)

where $\mu$ denotes the weights of the neural network in the critic network. Furthermore, the Q-value function is expressed as

$$Q = \mathbb{E}[r_t(s_t, x_t) + \gamma(Q(s_{t+1}, \Pi(s_{t+1}|\omega^+)|\mu^+))],$$

(36)

where $\omega^+$ denotes the weights of the neural network in actor target network, $\mu^+$ denotes the weights of the neural network in critic target network, and $\gamma \in [0, 1]$ denotes the discounting factor, which represents the effect of the future action choices.

In order to weaken the dependence of DDPG on hyper-parameters, the batch normalization [42] is adopted for DDPG, i.e., each layer in the neural networks of DDPG is connected to a batch normalization layer, which makes the DDPG less sensitive to the initial parameters, and prevents the unstable training process resulted from the unstable data distributions of each layer in the neural networks. The batch normalization accelerates the converge of DDPG, and efficiently avoids the gradient vanishing. Furthermore, due to the different value ranges of each factor in states, we normalize the input state of DDPG so that each factor in the state has the same value range.

Algorithm 1 finds the optimal policy for the time and energy management by DDPG. The exploration noise $N_t^\gamma$ in Algorithm 1 is used to fully explore the action space, in order to avoid being stuck in the local optimum policy. In the training process, the exploring noise decay factor $\kappa$ restricts the exploration range. The weights $\omega^+$ and $\mu^+$ of the target networks are updated by the soft replacement that increases the stability of the evaluated networks.
Algorithm 1: Finding the optimal policy for the time and energy management by DDPG.

**Input:** Training rounds $Tr$, training steps $Ts$, exploration noise $Ne$, exploration noise decay factor $\kappa$, learning rates of actor and critic $Ra$ and $Rc$, discounting factor $\gamma$, soft-update factor $\tau$, mini-batch bs.

**Output:** $x_t$, $Q_t$.

1. Empty the relay-memory pool $MP$;
2. Initializing: $\omega$ and $\mu$;
3. Then do $\omega^+ \leftarrow \omega$, and $\mu^+ \leftarrow \mu$;
4. for $i = 0$ to $Tr$ do
   5. Randomly sample a $s_t$ from $S$;
   6. for $t = 0$ to $Ts$ do
      7. Normalize the state $s_t$;
      8. Get action $x_t = \Pi(s_t|\omega) + Ne$;
      9. Clip $x_t$ to prevent it exceeding the bound of the action space;
     10. if $s_a^t = 1$ then
         11. Set each element of energy part as $e_{t,i} = 0$;
     end
    12. Broadcast $x_t$ to STs;
    13. Receive $r_t$ and $\epsilon_t + 1$;
    14. Observe $s_{t+1}$ and the channel gains of frame $t + 1$, then combine them with $\epsilon_{t+1}$ and $a_{t+1}$ as $s_{t+1}$;
    15. Store tuple $(s_t, x_t, r_t, s_{t+1})$ into $MP$;
    16. $Ne_{t+1} = Ne_t \times \kappa$;
    17. Train the actor network and critic network by DDPG;
   end
5. Simulation

In order to evaluate the performances of throughput and AoI, we compare the long-term average throughput $T$ of the ABO-CRN with the AoI constraint with two baseline schemes, throughput-optimal (T-O) scheme and AoI-optimal (A-O) scheme. The T-O baseline scheme optimizes the throughput of the secondary network, and the A-O baseline scheme optimizes the AoI of STs. Furthermore, we compare the throughput and AoI performances among the ABO-CRN, ABCs, and CRNs to evaluate the impacts of $T_{min}$ and $A_{max}$ on the throughput and AoI performances. The simulation configuration is set as follows unless otherwise specified: The transmit power of PT $P = 17$ kW, the bandwidth $W = 6$ MHz, the AWGN $\delta^2 = 10^{-3}$ $\mu$W, the energy capacity $E = 30$ $\mu$J, and backscatter reflection efficiency $\theta = 0.9$.

Figure 6 plots $T$ and AoI of the ABO-CRN, T-O baseline scheme, and A-O baseline scheme with the minimum throughput requirement $T_{min}$ under $P_a = 0.3, 0.6, 0.9$. We observe from Figure 6a that $T$ of the ABO-CRN decreases with $T_{min}$, and observe from Figure 6b that AoI of the ABO-CRN increases with $T_{min}$. For T-O baseline scheme, the throughput does not change with $T_{min}$, and the AoI increases faster than that of the ABO-CRN and A-O baseline scheme. The throughput of A-O baseline scheme decreases faster with $T_{min}$ than that of the ABO-CRN. The AoI of A-O baseline scheme increases with $T_{min}$, is close to that of the ABO-CRN, and is lower than that of T-O baseline scheme. The reasons can be explained as follows. When $T_{min}$ increases, each ST needs more throughput to reach the minimum throughput requirement. The SR in the ABO-CRN has to allocate more time and energy for the STs with high AoI and poor channel quality, and sacrifices the total throughput to satisfy the AoI constraint.
Figure 6. Throughput $\mathcal{T}$ and AoI versus $T_{\text{min}}$ under $P_a = 0.3$ (circle), 0.6 (square), 0.9 (triangle) compared with that of T-O and A-O baseline schemes, and $A_{\text{max}} = 5$: (a) describes the throughput performance. (b) describes the AoI performance.

When $P_a = 0.9$, we observe that $\mathcal{T}$ in Figure 6a decreases faster than that when $P_a = 0.3$ and 0.6, and the corresponding AoI in Figure 6b increases faster. The reason is provided as follows. When $P_a = 0.9$, due to the active channel state for the most time, STs only execute AB mode transmission for the most time. When $T_{\text{min}}$ is higher than the expected throughput achieved by STs through AB mode transmission, the AoI increases for the most frames, and the number of frames of the increased AoI becomes more with $T_{\text{min}}$. Therefore, the average AoI increases with $T_{\text{min}}$. The expected throughput achieved by STs through AB mode transmission and overlay mode transmission when $P_a = 0.3$ and 0.6 is higher than that when $P_a = 0.9$. Therefore, the average AoI when $P_a = 0.9$ increases faster with $T_{\text{min}}$ than that when $P_a = 0.3$ and 0.6. We also observe that when $T_{\text{min}}$ is small, the curves of the ABO-CRN and two baseline schemes are close. The reason is that, the throughput of three schemes meets the minimum throughput requirement, and the AoI of them satisfies the AoI constraint. In addition, Figure 6 shows that $\mathcal{T}$ of the ABO-CRN is closer to that of the T-O than that of A-O, and the AoI of the the ABO-CRN is closer to that of the A-O than that of T-O. It indicates that DDPG finds the optimal policy of time and energy management to optimize the throughput, and satisfies the AoI constraint.

Figure 7 plots $\mathcal{T}$ and AoI of the ABO-CRN, T-O baseline scheme, and A-O baseline scheme with the maximum allowable AoI constraint $A_{\text{max}}$. We observe that both $\mathcal{T}$ and AoI increase with $A_{\text{max}}$, and when $A_{\text{max}}$ is large, the throughput of the ABO-CRN and that of two baseline schemes are close. The reason is provided as follows. When $A_{\text{max}}$ increases, the limitation of the AoI constraint on throughput becomes weak. The SR allocates more time and energy to the STs with high throughput, hence the throughput increases with $A_{\text{max}}$. With the increase of $A_{\text{max}}$, STs with high AoI (but not exceed $A_{\text{max}}$) and poor channel quality is allocated less time and energy, hence AoI of these STs increases, and the average AoI of STs increases. When $A_{\text{max}}$ is large, all the three schemes satisfy the AoI constraint.

Figure 8 plots $\mathcal{T}$ and AoI in the ABO-CRN, ABCs, and CRNs with $T_{\text{min}}$, and Figure 9 plots $\mathcal{T}$ and AoI in the ABO-CRN, ABCs, and CRNs with $A_{\text{max}}$, under $P_a = 0.3, 0.6, 0.9$. It is obvious that $\mathcal{T}$ in the ABO-CRN is higher than that in the ABCs and CRNs, and AoI in the the ABO-CRN is lower than that in the ABCs and CRNs. When $P_a = 0.3$ and 0.9, the AoI in the ABCs and CRNs are high, and when $\mathcal{T}$ is large, STs in the ABCs and CRNs can not satisfy the AoI constraint. The reason is explained as follows. When $P_a = 0.3$, the channel keeps inactive for the majority part of the time. For the ABCs, AB mode transmission has a few opportunities to be executed. From Figure 6, we infer that, when $T_{\text{min}}$ is higher than the expected throughput achieved by each ST through AB mode transmission in a frame, STs in the ABCs are difficult to satisfy the AoI constraint by sacrificing the total throughput.
Therefore, the throughput of the ABCs in this case keeps nearly unchanged. When $P_a = 0.9$, the channel keeps active for the majority part of the time. For the CRNs, overlay mode transmission has few opportunities to be executed, hence the AoI of the CRNs is high, and the throughput of the CRNs keeps nearly unchanged. STs in the ABO-CRN execute AB mode transmission when the channel is active, and execute overlay mode transmission when the channel is inactive. As described in Equation (16), the expected throughput achieved by STs in the ABO-CRN is higher than that in the ABCs and CRNs. Therefore, the ABO-CRN achieves better throughput and AoI performances than that of the ABCs and CRNs.

Figure 7. Throughput $T$ and AoI versus $A_{\text{max}}$ under $P_a = 0.3$ (circle), 0.6 (square), 0.9 (triangle) compared with that of T-O and A-O baseline schemes, and $T_{\text{min}} = 700$ bps: (a) describes the throughput performance. (b) describes the AoI performance.

Figure 8. Throughput $T$ and AoI versus $T_{\text{min}}$ under $P_a = 0.3$ (circle), 0.6 (square), 0.9 (triangle) compared with that in the ABCs and CRNs, and $A_{\text{max}} = 5$: (a) describes the throughput performance. (b) describes the AoI performance.
6. Conclusions

We optimized the long-term throughput of the secondary network with the AoI constraint by jointly managing the time and energy for STs in the ABO-CRN, ABCs, and CRNs through DDPG. When the AoI constraint can not be satisfied, the impacts of time and energy allocation on the reward were investigated, and the corresponding reward functions was developed based on the channel states. We discussed the minimum throughput requirement and the maximum allowable AoI that are related to the throughput and AoI performances. We compared the throughput optimization scheme with the AoI constraint with T-O and A-O baseline schemes, and varied the minimum throughput requirement and maximum allowable AoI to evaluate the effects on the throughput and AoI performances of the secondary networks in the ABO-CRN, ABCs, and CRNs. We had following findings:

- Throughput of the ABO-CRN is close to the optimal throughput of T-O baseline scheme, and the AoI of the ABO-CRN is close to the optimal AoI of A-O baseline scheme. DDPG finds the optimal policy of time and energy management to optimize the throughput, and satisfies the AoI constraint at the same time.
- Throughput of the ABO-CRN is higher than that of A-O baseline scheme, and AoI of the ABO-CRN is lower than that of T-O baseline scheme. The observation validates the benefit of considering both throughput and AoI performances over only one metric.
- The ABO-CRN improves the throughput and AoI performances of the ABCs and CRNs. Even in extreme cases, such as the long time active channel state, the ABO-CRN obtains better throughput and AoI performances than the ABCs and CRNs.
- The lower bound of the maximum allowable AoI that makes STs satisfy the AoI constraint decreases with the total number of STs, and increases with the number of STs whose average throughput is smaller than the minimum throughput requirement.

Author Contributions: This research has been carried out through a concerted effort by seven months. Each author’s basic role has been summarizing in following: Conceptualization, X.J.; methodology, X.J.; software, X.J.; validation, K.Z., K.C., and X.L.; investigation, X.J., K.Z., K.C., and X.L.; writing—original draft preparation, X.J.; writing—review and editing, K.Z.; supervision, K.C.; funding acquisition, K.C. All authors have read and agreed to the published version of the manuscript.

Funding: National Natural Science Foundation of China, Grant Number: 61902351, 61902353, 61872322. Zhejiang Provincial Natural Science Foundation of China, Grant Number: LY21F020022, LY21F020023, LR20F020003.

Institutional Review Board Statement: Not applicable.
Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

We used the abbreviations in this paper:

- RF: Radio frequency
- CR: Cognitive radio
- CRN: CR network
- AB: Ambient backscatter
- ABC: AB communication
- AB-CRN: AB-assisted CRN
- ABO-CRN: AB-assisted overlay CRN
- ABU-CRN: AB-assisted underlay CRN
- DRL: Deep reinforcement learning
- DDPG: Deep deterministic policy gradient
- AoI: Age of information
- PU: Primary user
- PT: Primary transmitter
- PR: Primary receiver
- SU: Secondary user
- ST: Secondary transmitter
- SR: Secondary receiver

References

1. Liu, X.; Zheng, K.; Chi, K.; Zhu, Y. Cooperative spectrum sensing optimization in energy-harvesting cognitive radio networks. *IEEE Trans. Wirel. Commun.* 2020, 19, 7663–7676. [CrossRef]
2. Zheng, K.; Liu, X.; Zhu, Y.; Chi, K.; Liu, K. Total throughput maximization of cooperative cognitive radio networks with energy harvesting. *IEEE Trans. Wirel. Commun.* 2019, 19, 533–546. [CrossRef]
3. Siegel, J. E.; Kumar, S.; Sarma, S.E. The future Internet of Things: Secure, efficient, and model-based. *IEEE Internet Things J.* 2018, 5, 2386–2398. [CrossRef]
4. Zhang, S.; Kong, S.; Chi, K.; Huang, L. Energy management for secure transmission in wireless powered communication networks. *IEEE Internet Things J.* 2021, 9, 1171–1181. [CrossRef]
5. Niyato, D.; Kim, D.I.; Han, Z.; Maso, M. Wireless powered communication networks: architectures, protocol designs, and standardization. *IEEE Wirel. Commun.* 2016, 23, 8–9. [CrossRef]
6. Chi, K.; Zhu, Y.; Li, Y.; Huang, L.; Xia, M. Minimization of transmission completion time in wireless powered communication networks. *IEEE Internet Things J.* 2017, 4, 1671–1683. [CrossRef]
7. Yang, L.; Zhou, Y.J.; Zhang, C.; Zhang, X.; Yang, X.; Tan, C. Compact multiband wireless energy harvesting based battery-free body area networks sensor for mobile healthcare. *IEEE J. Electromagn. Microw. Med. Biol.* 2018, 2, 109–115. [CrossRef]
8. Chi, K.; Chen, Z.; Zheng, K.; Zhu, Y.H.; Liu, J. Energy provision minimization in wireless powered communication networks with network throughput demand: TDMA or NOMA. *IEEE Trans. Commun.* 2019, 67, 6401–6414. [CrossRef]
9. Hoang, T.M.; El Shafie, A.; da Costa, D.B.; Duong, T.; Tuan, H.; Marshall, A. Security and energy harvesting for MIMO-OFDM networks. *IEEE Trans. Commun.* 2020, 68, 2593–2606. [CrossRef]
10. Sangdeh, P.K.; Pirayesh, H.; Quadri, A.; Zeng, H. A practical spectrum sharing scheme for cognitive radio networks: Design and experiments. *IEEE/ACM Trans. Netw.* 2020, 28, 1818–1831. [CrossRef]
11. Ghosh, D.; Hanawal, M.K.; Zlatanov, N. Learning to optimize energy efficiency in energy harvesting wireless sensor networks. *IEEE Wirel. Commun. Lett.* 2020, 9, 1000–1003. [CrossRef]
12. Gu, Z.; Shen, T.; Wang, Y.; Lau, E.C.M. Efficient rendezvous for heterogeneous interference in cognitive radio networks. *IEEE Trans. Wirel. Commun.* 2020, 19, 91–105. [CrossRef]
13. Sangdeh, P.K.; Pirayesh, H.; Quadri, A.; Zeng, H. A practical spectrum sharing scheme for cognitive radio networks: Design and experiments. *IEEE/ACM Trans. Netw.* 2020, 28, 1818–1831. [CrossRef]
14. Zheng, K.; Liu, X.; Liu, X.; Zhu, Y. Hybrid overlay-underlay cognitive radio networks with energy harvesting. *IEEE Trans. Commun.* 2019, 67, 4669–4682. [CrossRef]
15. Papadopoulos, A.; Chatzidiamantis, N.D.; Georgiades, L. Network coding techniques for primary-secondary user cooperation in cognitive radio networks. *IEEE Trans. Wirel. Commun.* 2020, 19, 4195–4208. [CrossRef]
16. Rathee, G.; Jaglan, N.; Garg, S.; Choi, B.J.; Choo, K.K.R. A secure spectrum handoff mechanism in cognitive radio networks. *IEEE Trans. Cogn. Commun. Netw.* 2020, 6, 959–969. [CrossRef]
