sensors

MDPI

Article

# A Novel Dynamic Spectrum Access Framework Based on Reinforcement Learning for Cognitive Radio Sensor Networks

**Yun Lin [1,*], Chao Wang [1], Jiaxing Wang [2] and Zheng Dou [1]**

[1] College of Information and Communication Engineering, Harbin Engineering University, Harbin 150001, China; wangchao@hrbeu.edu.cn (C.W.); douzheng@hrbeu.edu.cn (Z.D.)

[2] Beijing Huawei Digital Technologies Co., Ltd., Beijing 100032, China; wangjiaxing1@huawei.com

* Correspondence: linyun@hrbeu.edu

Academic Editor: Leonhard M. Reindl

Received: 15 July 2016; Accepted: 7 October 2016; Published: 12 October 2016

**Abstract:** Cognitive radio sensor networks are one of the kinds of application where cognitive techniques can be adopted and have many potential applications, challenges and future research trends. According to the research surveys, dynamic spectrum access is an important and necessary technology for future cognitive sensor networks. Traditional methods of dynamic spectrum access are based on spectrum holes and they have some drawbacks, such as low accessibility and high interruptibility, which negatively affect the transmission performance of the sensor networks. To address this problem, in this paper a new initialization mechanism is proposed to establish a communication link and set up a sensor network without adopting spectrum holes to convey control information. Specifically, firstly a transmission channel model for analyzing the maximum accessible capacity for three different polices in a fading environment is discussed. Secondly, a hybrid spectrum access algorithm based on a reinforcement learning model is proposed for the power allocation problem of both the transmission channel and the control channel. Finally, extensive simulations have been conducted and simulation results show that this new algorithm provides a significant improvement in terms of the tradeoff between the control channel reliability and the efficiency of the transmission channel.

**Keywords:** dynamic spectrum access; control channel; power allocation; reinforcement learning

## 1. Introduction

Cognitive radio (CR) is a promising technology which can fully use the spectrum by dynamically accessing the primary network. Consequently, dynamic spectrum access technology plays a very significant role and has become a hot research topic. As illustrated in Figure 1, dynamic spectrum access strategies can be classified into three models, e.g., the dynamic exclusive use model, the open sharing model, and the hierarchical model. Among those models, the hierarchical model is a hierarchical access structure for primary users (PUs) and secondary users (SUs), and is the most promising and effective one for current spectrum access policies [1]. The basic idea of the hierarchical model is that the SUs can use the licensed spectrum of PUs, as long as they can limit any interference perceived by PUs. Furthermore, there are two models of the spectrum sharing between PUs and SUs, namely spectrum underlay and spectrum overlay.

Spectrum underlay introduces severe constraints on the transmission power of the SUs, therefore, it spreads the transmitted signals over a wide frequency band. The SUs can achieve low data rates with very low transmission power. If the PUs transmit in all the time-slots, the spectrum underlay does not need to detect and perceive the spectrum of the PUs.

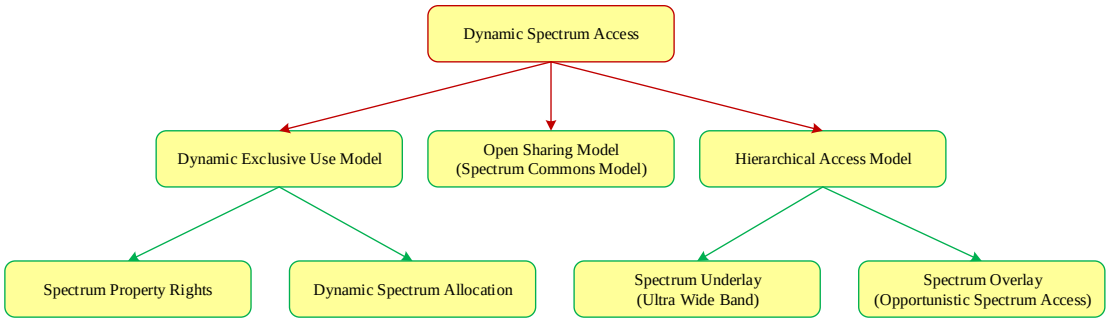


**Figure 1.** The dynamic spectrum access models.

Spectrum overlay, first presented by Mitola, can be also regarded as opportunistic spectrum access (OSA). Compared to the spectrum underlay, this model needs to detect and perceive the spectra of the PUs. It finds spatial and temporal spectrum white space for SUs to use, which is also termed as the spectrum holes (SHs). Therefore, this model does not need to obey the severe transmission power constraints of the SUs, and the SUs can achieve high date rates with high transmission power.

In most cases, the spectrum overlay and underlay models are used separately. In this paper, a hybrid spectrum access model is proposed to use both the overlay and underlay methods simultaneously to further improve the current spectrum efficiency.

The spectrum hole (SH) is a part of the licensed spectrum which is not being used by the owner during a period of time [1]. Among key technologies in CR, the design of the control channel is essential because the SUs need a control channel to coordinate and they have no licensed spectrum to carry the control information. The vulnerabilities resulting from utilizing a dedicated control channel have been well studied. Existing studies of the control channel have shown that using SHs to convey control information is only a basic approach and many shortcomings have been pointed out [2–6]. Firstly, the SUs may not have a common SH as control channel which would lead to low connectivity of the SUs. Secondly, the arrival of PU is unknown which causes interruptions in the use of the control channel.

As the SUs communicate only in the SHs, the SUs need information about those unused bands in which the PUs are inactive. Each SU should maintain a list of SHs which probably will differ from one to another. The SUs can communicate with each other if there is a common SH in their lists. Consequently, there should be a way to pass information about the lists between SUs during the initial communication.

Most of the existing MAC protocols of CR sensor networks are focused on avoiding common control channels. However, in this paper, a new method of spreading the power spectrum density in a control channel over an ultra-wide bandwidth is proposed to exploit the underused (gray) spectral regions. Like underlay spectrum sharing, the SUs can always access to the spectrum as long as the interference causing by SUs at the PU receiver can satisfactorily meet the threshold constraint [7].

According to the above analysis and considering the low power spectrum density of underlay waveforms, we propose to design a control channel to convey a small amount of control information, which is termed as SUCCH. At the same time, the spectrum overlay waveform is adopted to exchange a large amount of date, which is named as SUTCH. Our study is based on a spectrum sharing system consisting of two different waveforms. The first one is the Direct Sequence Code Division Multiple Access (DS-CDMA), which is defined as the underlay waveform used to convey control information. The second one is the Non-Contiguous Orthogonal Frequency Division Multiplexing (NC-OFDM), which is defined as the overlay waveform used to convey data information. The spectrum of NC-OFDM-based SUs is shared with the PUs which utilize DS-CDMA. Spreading Gain of DS-CDMA provides the required anti-jamming capability for the interference which may be caused by the SUs. In the meantime, based on the properties of the non-continuous power spectrum of NC-OFDM, it is more flexible for the SUs to access the SHs which are discontinuous in the frequency spectrum [8]. It is

of great significance to discuss and study this issue, since the existing DS-CDMA is anticipated to be one of the spectrum sharing applications used in the future [9].

In order to set up the hybrid spectrum access model, several questions should be answered. The first one is the procedure for network setup between two SUs. The second one is the maximum access capacity of the SUTCH with different strategies. The third one is the reliability of the SUCCH. The fourth one is the power allocation strategies of the SUs between the SUTCH and SUCCH. In the rest of this paper, the above questions will be answered in detail. Specifically, Section 2 builds application scenarios and proposes a mechanism for establishing the CR sensor networks. In Section 3, a transmission channel model for analyzing the maximum access capacity for different polices with different objectives in the fading environment will be discussed. In Section 4, the reliability of the SUCCH is analyzed, and a hybrid spectrum access algorithm based on reinforcement learning model is proposed for the power allocation problem of the SUTCH and the SUCCH. Finally, Section 5 presents our simulation results and Section 6 concludes the paper.

## 2. Application Scenarios

In this section the application scenario is described as below. As shown in Figure 2, there are four active PUs and each one is authorized to use a certain frequency band to communicate. The different types of circles represent the interference ranges of each PU, and six SUs are shown in Figure 2. In this paper there is a channel which is termed a SH and a SU that can communicate in this channel because it is a channel whose authorized PU is currently inactive or the SU is beyond the interference range of that PU.

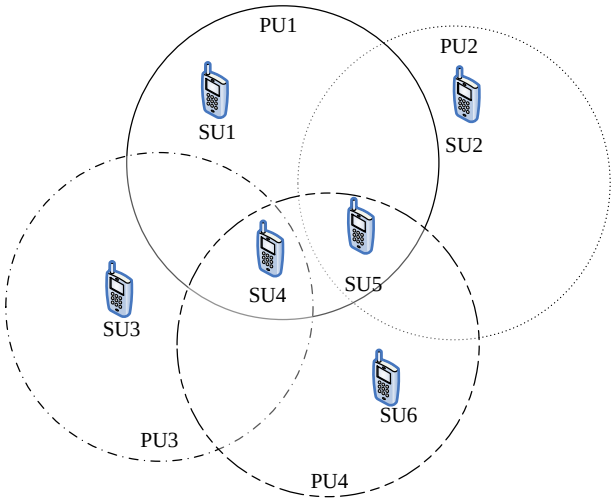


**Figure 2.** The SUs among four PUs.

A SU can establish the connection with another SU as long as they both have a shared SH in their respective lists of SHs, so it is important for a SU to identify its neighbors during the initial communication used to set up CR sensor networks. In order to fully utilize the primary spectrum and maximize the efficiency of spectrum, underlay and overlay transmissions, which exploit the white and grey spaces respectively, should be used together [1,10,11]. However, for spectrum underlay, the SUs need to transmit at low power to avoid any interference with the PUs, whereas the PUs will cause interference with SUs [12]. In consideration of the low power spectrum density of underlay waveforms, the control channel is designed to convey a small amount of control information, which is named as SUCCH, while the spectrum overlay waveform is used to exchange a large amount of data, which is named as SUTCH. Considering the perspective of a SU, the current spectrum usage is depicted in Figure 3.

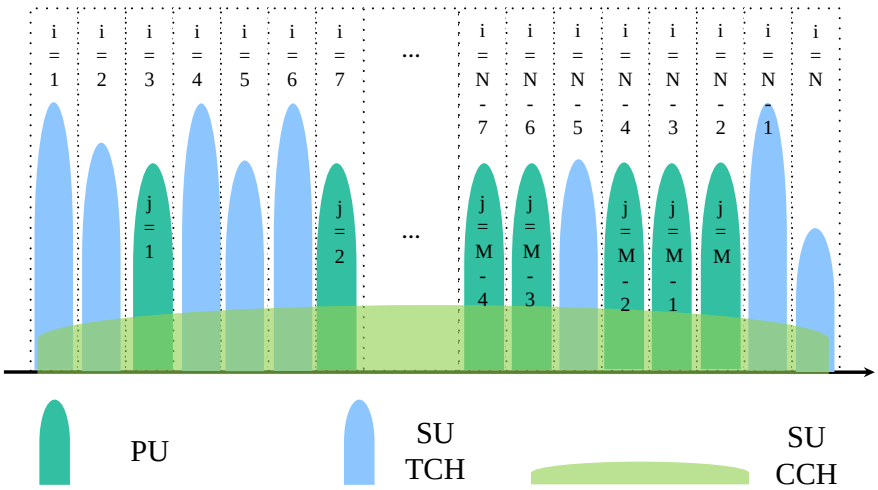


**Figure 3.** Spectral occupancy of the hybrid access.

Before explaining the protocol used to set up CR sensor networks, it is necessary to discuss the capabilities of the SUs and define some terms that will be used in the coming discussion. A SU can switch between spectra autonomously and sense the spectrum. Each SU identifies itself by using a different Orthogonal Variable Spreading Factor (OVSF) [12] over spectrum underlay. The number of the SUs in the current CR sensor networks is a priori information available to all the SUs.

The proposed protocol is firstly discussed under a distributed architecture scenario, which is also called Multi-Hop Architecture. Each SU initially starts to send beacons in different OVSF over spectrum underlay to indicate its presence. At the mean time every SU monitors the spectrum underlay by randomly selecting a form of OVSF while initially starting a timer which counts to $T_S$ seconds. If none of those beacons is captured during the $T_S$ seconds, the SU will change to another form of OVSF in the next time slot. If a beacon is received by selecting the current form of OVSF, the SUs will sent a response in the same form which is considered as the task of carrying on the negotiations. After exchanging the control information with each other, the common SH in the two SUs will start to provide service. The procedure is simply illustrated in Figure 4.

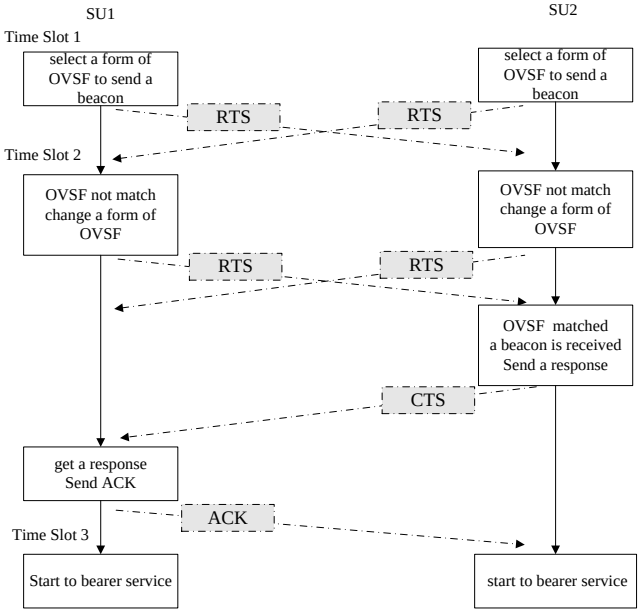


**Figure 4.** The procedure of network setup between two SUs.

In Figure 4, "Request to Send (RTS)" and "Clear to Send (CTS)" exchange messages to reserve a channel for communications in a similar manner that the IEEE 802.11 Distributed Coordination function (DCF) designs the MAC protocol [13]. RTS or CTS carries information about SUs' lists of SH and accesses SUs states.

## 3. Subchannel Selection Policies

Suppose the wireless channel is a frequency-selective Additive White Gaussian Noise (AWGN), the bandwidth is $B$ Hz, and the power spectral density is $N_0$. In this paper, it is divided into $N$ Rayleigh fading subchannels, and the subchannel coherence bandwidth is $\Delta f$ Hz. Therefore, $B = N\Delta f$. These subchannels are indexed by $i = 1, 2, \ldots, N$, and the gains of every subchannel are independent and identically distributed (i.i.d).

Active PUs use DS-CDMA technology to access the spectrum band with spreading gain $G$. According to the Central Limit Theorem, the interference process in the receiver of the SUs caused by a large number of PUs is considered a Gaussian approximation. Furthermore, according to the second-order statistics, the interference process is a white process [14]. Therefore, in each subchannel, the average interference introduced by the PUs at the receiver of the SUs is $(K-1)N_0\Delta f$, $K \geq 1$, where $K$ is a system parameter related to the characteristics of PUs network [15].

As shown in Figure 4, the SUs utilize NC-OFDM to access the SUTCH which is indexed by $j = 1, 2, \ldots, M$, $0 \leq M \leq N$. The SUs spread their SUCCH power spectrum density over an ultra-wide bandwidth to exploit the underused (gray) spectral. $Q$ is defined as the interference threshold of the PUs, which is the maximum allowable temporal interference in the receiver of the PUs caused by concurrent activity of the SUs in the same subchannel. As mentioned in Figure 4, the protocol to set up CR sensor networks is based on the time-slot structure. Therefore, in order to satisfy the interference threshold constraint, the power of the SUs accessing the SUTCH should be controlled in each time-slot.

In this paper, the structure of the accessing system is depicted in Figure 5. For subchannel $j$, the instantaneous gain between the transmitter and receiver of the SU is defined as $g_{ss}^j$, and the instantaneous gain between the transmitter of the SU and the receiver of the PU is defined as $g_{ps}^j$. Subscripts $s$ and $p$ refer to the secondary and the primary user, respectively. The $g_{ss}^j$ and $g_{ps}^j$ are assumed as the stationary and ergodic independent distributed random variables with unit-mean. Their Probability Density Functions (PDFs) are defined as $f_{ss}^j(g_{ss}^j)$ and $f_{ps}^j(g_{ps}^j)$, respectively. Channel gains $g_{ss}^j$ and $g_{ps}^j$ are i.i.d., $j = 1, 2, \ldots, M$. In this paper, we suppose the perfect Channel Side Information (CSI) pair $(g_{ss}^j, g_{ps}^j)$ can be available in the transmitters. Here, the CSI contains the probability distribution of the channel gain, as well as the actual value at a certain time-slot. Actually, the CSI pair can be estimated by a spectral coordinator or proper signaling. Note that, the result derived from this assumption is an upper-bound in the case without a perfect CSI pair.

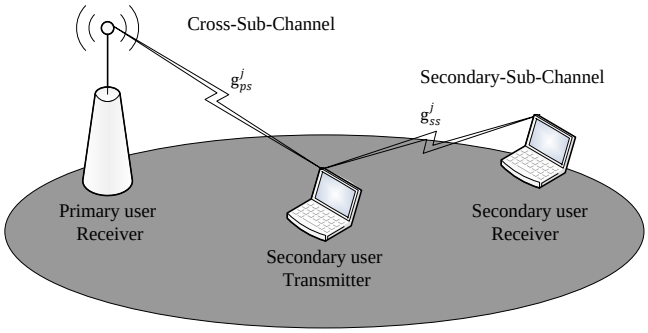


**Figure 5.** The structure of the accessing system for subchannel $j$.

In this paper, we focus on the maximum achievable spectrum capacity of SUTCH, which is studied [16,17]. Since more than one SUs will compete to access to the underused frequency band. The SUs' total available spectrum capacity is upper-bound by the case of only one SU, which is due to

the fact that SUs will impose interference on each other. Therefore, the discussion of the individual SU can also be used as the upper-bound of the total spectrum capacity of all SUs.

At a given time-slot, the power allocation policy of SUTCH is defined as $\rho_\psi$, which is based on a selection criterion $\psi(,..,)$, and set:

$$\mu_j \triangleq \psi\left(g_{ps}^j, g_{ss}^j\right) \tag{1}$$

For the observing random variables $\mu_j$, $j = 1, \ldots, M$, the selection sequence $\gamma_M$ is defined as follows:

$$\gamma_M = (\mu_{r_1}, \mu_{r_2}, \ldots, \mu_{r_M}) \triangleq \rho_\psi(\mu_1, \mu_2, \ldots, \mu_M) \tag{2}$$

The $M$-tuple selection sequence is arranged, so that its first element is the most suitable subchannel for SUTCH based on the selection criteria in Equation (1). The probability distribution function of random variable $\gamma_j$ is defined as $k_j(\gamma)$, $j = 1, \ldots, M$. It is important to note that if $j_1$, $j_2$ are entities in $\gamma_M$ and $j_1 < j_2$, then it can be considered that compared to the choice $j_2$, the SUs can get a better performance by choosing subchannel with index $j_1$.

Suppose $\psi(g_{ps}^j, g_{ss}^j)$ is constant, which means subchannels are considered equally. The SUs will randomly choose $M$ out of $N$ subchannels without any a priori information. This selection strategy is defined as the uniform subchannel selection, whereas, if the prior information of the subchannel obtained by cooperation or other techniques is $-1$, the SUs will choose the corresponding value of $\psi(g_{ps}^j, g_{ss}^j)$. This selection strategy is defined as the non-uniform selection strategy.

The transmission power of the SUTCH in the subchannel $j$ is referred to $P_{sj}$. $\boldsymbol{P}_s(P_{s1}, \ldots, P_{sM})$ is defined as the transmission power vector of SUTCH over $M$ subchannels. Suppose that SUTCH accesses to the chosen subchannel $j$ with the transmission power of $P_{sj}$, and the corresponding interference at the receiver of the PUs is $Q_j$, where:

$$Q_j = g_{ps}^j P_{sj} \tag{3}$$

Since the PUs utilize DS-CDMA with spreading gain $G$, therefore, the narrow-band interference $Q_j$ spreads over the whole bandwidth and manifests itself as an equivalent wide-band interference equal to $G^{-1}Q_j$ at the receiver of the PUs. Suppose the SUTCH transmits with the transmission power vector $\boldsymbol{P}_s(P_{s1}, P_{s2}, \ldots, P_{sM})$ in $M$ accessible subchannels. Correspondingly, an equivalent narrow-band interference vector $\boldsymbol{Q} = (Q_1, Q_2, \ldots, Q_M)$ will be imposed on the receivers of the PUs. Meanwhile, the SUCCH transmits with the transmission power vector $\boldsymbol{P}_{sc}(P_{sc1}, P_{sc2}, \ldots, P_{scN})$. Therefore, in order to comply with the interference threshold $Q$ of the PUs, the constraint function is as follows:

$$\frac{1}{G}\left(\sum_{j=1}^{M} g_{\mathrm{ss}}^j P_{sj} + \sum_{i=1}^{N} g_{ps}^i P_{sci}\right) \leq Q \tag{4}$$

In this paper, the objective is to achieve the maximum capacity of SUTCH. As discussed above, the transmitting power of SUTCH in each accessible subchannel should be optimally allocated. Meantime, the interference threshold constraint should also be considered. Consequently, according to selection policy $\rho_\psi$, for a given $Q$ and for $M$ accessible subchannels, the maximum capacity of SUTCH is defined as $C_M^\psi$, which can be obtained by the following constrained optimization problem:

$$\begin{aligned} C_M^\psi \quad = \quad & \max_{\boldsymbol{P}_s} \sum_{j=1}^{M} \Delta f \int_{g_{ps}^j g_{ss}^j} \log\left(1 + \frac{g_{ss}^j P_{sj}}{KN_0\Delta f + g_{ss}^j P_{scj}}\right) \\ & \times f_{ss}^j\left(g_{ss}^j\right) f_{ps}^j\left(g_{ps}^j\right) dg_{ss}^j dg_{ps}^j \\ s.t. \quad & \frac{1}{G}\left(\sum_{j=1}^{M} g_{ps}^j P_{sj} + \sum_{i=1}^{N} g_{ps}^i P_{sci}\right) \leq Q \\ & \sum_{j=1}^{M} P_{sj} + \sum_{i=1}^{N} P_{sci} \leq P_s \end{aligned} \tag{5}$$

where, $Q$ is the interference threshold of the PUs, which is the maximum allowable temporal interference in the receiver of the PUs caused by concurrent activity of the SUs in the same subchannel. $P_s$ $N_0$ is the power spectral density, $\Delta f$ is the subchannel coherence bandwidth. $K$ is a system parameter related to the characteristics of PUs network [15] within the range of 2–8. Equation (5) is derived from Shannon's Capacity formula with the SUs power vector $\boldsymbol{P}_s$ and $P_{sc}$. Equation (6) is the constraint function of interference threshold of the PUs and maximum transmitting power of the SUs.

Actually, in contrast to the constraint of maximum transmitting power of the SUs, the constraint function of interference threshold of the PUs is much tighter [18]. Therefore, in this paper, the constraint of maximum transmitting power of SUs is not considered. At the same time, as mentioned above, the SUCCH spreads over an ultra-wide bandwidth to exploit the underused spectrum with a very low PSD, therefore, the interference caused by SUCCH is very low. In this paper, in order to simplify the analysis, the effect of SUCCH will not be considered, and Equation (5) will be further simplified as follows:

$$\begin{aligned} C_M^{\psi} &= \max_{\boldsymbol{P}_s} \sum_{j=1}^{M} \Delta f \int\limits_{g_{ps}^j g_{ss}^j} \log\left(1+\tfrac{g_{ss}^j P_{sj}}{KN_0\Delta f}\right) \\ &\quad \times f_{ss}^j\left(g_{ss}^j\right) f_{ps}^j\left(g_{ps}^j\right) dg_{ss}^j dg_{ps}^j \\ s.t. &\quad \tfrac{1}{G}\left(\sum_{j=1}^{M} g_{ps}^j P_{sj}\right) \le Q \end{aligned} \tag{6}$$

Suppose $\psi\left(g_{ps}^j, g_{ss}^j\right) = 1$, thus the SUs will randomly choose $M$ out of $N$ subchannels without any priori information by $\rho_1$, which is a uniform subchannel selection policy. Consequently, substituting $P_{sj} = Q_j / g_{ps}^j$, $j = 1, \ldots, M$ and defining $\theta_{Q_j} \triangleq Q_j / KN_0\Delta f$ Equation (6) can be simplified as follows:

$$\begin{aligned} C_M^{\rho_1} &= \max_{\mathrm{Q}} \sum_{j=1}^{M} \Delta f \int_{v_j} \log\left(1+v_j\theta_{Q_j}\right) h_j\left(v_j\right) dv_j \\ s.t. &\quad \sum_{j=1}^{M} Q_j = GQ, \quad 0 \le Q_j \le GQ \end{aligned} \tag{7}$$

where $v_j \triangleq g_{ss}^j / g_{ps}^j$, $0 \le v_j \le \infty$, $v_j$ is the reward factor of the subchannel $j$. $\theta_{Q_j}$ is defined as the spectrum sharing load factor of the subchannel $j$.

Suppose the statistics characteristics of $\sqrt{g_{ps}^j}$, $\sqrt{g_{ss}^j}$ is i.i.d. Rayleigh random variables, $g_{ps}^j$ and $g_{ss}^j$ are exponentially distributed random variables with unit-mean, therefore, the PDF of $v_j$ can be converted into [17]:

$$\begin{aligned} h_j(v_j) &= \tfrac{d}{dv_j}\int_0^{\infty}\int_0^{g_{ps}^j v_j} e^{-g_{ps}^j} e^{-g_{ss}^j} dg_{ps}^j dg_{ss}^j \\ &= \int_0^{\infty} g_{ps}^j e^{-g_{ps}^j} e^{-g_{ps}^j \frac{g_{ss}^j}{g_{ps}^j}} dg_{ps}^j \\ &= \int_0^{\infty} g_{ps}^j e^{-g_{ps}^j(1+v_j)} dg_{ps}^j \\ &= -\tfrac{1}{1+v_j}\left\{\left[g_{ps}^j e^{-g_{ps}^j(1+v_j)}\right]\Big|_0^{\infty} - \int_0^{\infty} e^{-g_{ps}^j(1+v_j)} dg_{ps}^j\right\} \\ &= -\tfrac{1}{(1+v_j)^2}\left[e^{-g_{ps}^j(1+v_j)}\right]\Big|_0^{\infty} \\ &= \tfrac{1}{(1+v_j)^2} \quad 0 < v_j < \infty \end{aligned} \tag{8}$$

Substituting Equation (9) into Equation (7), and integrating by part, Equation (10) can be gotten as follows, which is the simplified optimization problem of $C_M^{\rho_1}$:

$$\begin{aligned} C_M^{\rho_1} = & \max_{\boldsymbol{\theta_Q}} \sum_{j=1}^{M} \Delta f \frac{\theta_{Q_j}}{\theta_{Q_j}-1} \log\left(\theta_{Q_j}\right) \\ s.t. & \sum_{j=1}^{M} \theta_{Q_j} = GN\theta_Q, \qquad 0 \le \theta_{Q_j} \le GN\theta_Q \end{aligned} \tag{9}$$

where, $\theta_Q$ is defined as the spectrum sharing load factor, and $\boldsymbol{\theta_Q} = (\theta_{Q_1}, \theta_{Q2}, \ldots, \theta_{QM})$ is defined as the spectrum sharing load vector:

$$\theta_Q \triangleq \frac{Q}{KN_0N\Delta f} = \frac{Q}{KN_0B} \tag{10}$$

Furthermore, the following pseudo linear approximation is used to get an approximate solution for Equation (10) [16]:

$$\frac{x}{x-1}\log(x) \approx -1.2015 - 0.0052x + 1.0772 \times \log\left(3.0262x + 308829\right) \tag{11}$$

Substituting Equation (12) into Equation (10), the Lagrangian function of the optimization problem Equation (10) is shown as follows [19,20]:

$$\begin{aligned} L(\theta_Q, \lambda) = & \sum_{j=1}^{M} -1.2015 + -0.0052 \times \theta_{Q_j} + 1.0772 \times \log\left(3.0262 \times \theta_{Q_j} + 3.8829\right) \\ & -\lambda\left(\sum_{j=1}^{M} \theta_{Q_j} - GN\theta_Q\right) \end{aligned} \tag{12}$$

where $\lambda$ is the Lagrangian coefficient. The derivative with respect to the $\theta_{Q_j}$ on Equation (13) is taken, and then it is equal to zero, the following formula can be obtained:

$$\theta_{Q_j}^* = \frac{1.0772}{\lambda^* + 0.0052} - \frac{3.8829}{3.0262}. \tag{13}$$

Substituting Equation (14) into Equation (10), the following formula can obtained:

$$\sum_{j=1}^{M}\left[\frac{1.0772}{\lambda^* + 0.0052} - \frac{3.8829}{3.0262}\right] = GN\theta_Q \tag{14}$$

Equivalently, Equation (16) can be derived from Equation (15):

$$\lambda^* = -0.0052 + \frac{1.0772}{\frac{GN\theta_Q}{M} + \frac{3.8829}{3.0262}}. \tag{15}$$

Eventually, substituting Equation (16) into Equation (14) gives:

$$\theta_{\theta_j}^* = \frac{GN\theta_Q}{M}, \quad j = 1, 2, \ldots, M \tag{16}$$

Note that Equation (17) suggests that for given $G$, $N$, $M$ and $\theta_Q$, the maximum capacity is achieved by dividing the total acceptable interference $GN\theta_Q$ into equal portions for $M$ accessible subchannels. Actually, it is a direct consequence of selecting $M$ out of $N$ subchannels without any prior knowledge. Furthermore, according to Equation (3) and $\theta_{Q_j} \triangleq Q_j/KN_0\Delta f$, the optimal transmitting power vector $\mathbf{P}_s^*$ can be obtained as follows:

$$\mathbf{P}_s^* = \left(\frac{1}{g_{ps}^1}\frac{GQ}{M}, \frac{1}{g_{ps}^2}\frac{GQ}{M}, \ldots, \frac{1}{g_{ps}^M}\frac{GQ}{M}\right) \tag{17}$$

Equation (18) suggests that the interference share for each accessible subchannel $j$, $\theta_{Q_j}$ is mapped to the corresponding transmission power $P_{sj}$, proportional to $1/g_{ps}^{j}$. So, if $g_{ps}^{j}$ is large, then the SUs will creates a large interference in the receivers of Pus. In this case, Equation (18) suggests a lower SUs transmission power in accessible subchannel $j$.

Equivalently, substituting Equation (18) into Equation (10), Equation (19) can be derived:

$$C_M^{\rho_1} \approx M\Delta f \frac{GN\theta_Q}{GN\theta_Q - M} \log\left(\frac{GN\theta_Q}{M}\right) \tag{18}$$

In a practical case, $Q = G^{-1}N_0B$ and $M < N$, the spectrum sharing load factor can be obtained from Equation (17) as $\theta_{\theta_j} = N/KM$, which is much higher than unity.

As mentioned above, $\rho_1$ randomly choose subchannels, which ignores the fact that it is more reasonable for the SUs to allocate higher transmission power to certain subchannels because of their corresponding CSIs, so it is essential to discuss the non-uniform selection policy for SUTCH with a prior knowledge of CSIs pair $(g_{ss}^{j}, g_{ps}^{j})$, since it will lead to a larger capacity or a smaller interference on the PUs.

Actually, an appropriate selection policy should consider the interference of the PUs receivers caused by SUs transmission. Such policy should select the lower subchannel gain of $g_{ps}^{j}$, because it will create a lower interference in the receivers of the PUs. Therefore, a lower $g_{ps}^{j}$ will give the SUs the flexibility of allocating a higher power, which will result in a higher capacity. Such a selection policy is named as SU-PU-based selection policy, which is simplified as $\rho_{ps}$. In order to implement $\rho_{ps}$, the SUs requires $g_{ps}^{j}$ during each time-slot. Therefore, a signaling channel between the receivers of the PUs and the transmitters of the SUs is required.

Similar to $\rho_{ps}$, another selection policy can be derived. It will select those subchannels which achieve the highest capacity corresponding to allocating the transmitting power of SUs. Such policy selects the subchannel with the higher $g_{ss}^{j}$, because it will create a higher power in the receivers of the SUs. Such selection is name as SU-SU-based selection policy, which is simplified as $\rho_{ss}$. In order to implement $\rho_{ss}$, the SUs requires $g_{ss}^{j}$ during each time-slot. Therefore, a signaling channel between the receivers of the SUs and the transmitters of the SUs is also required. In the following, the maximum capacity is derived with different selection policy $\rho_{ps}$ and $\rho_{ss}$.

Considering $\rho_{ps}$, the selection criteria can be assumed as follows:

$$\psi\left(g_{ps}^{j}, g_{ss}^{j}\right) = g_{ps}^{j} \tag{19}$$

Consequently, $\mu_j = g_{ps}^{j}$ and based on $\mu_j$, $j$ = 1, 2, ... , $M$, the selection sequence is defined as follows:

$$\gamma_M = (\mu_1, \mu_2, \ldots, \mu_M) \triangleq \rho_{ps}\left(\mu_1, \mu_2, \ldots, \mu_M\right) \tag{20}$$

where $\mu_1 \leq \mu_1 \leq \ldots \leq \mu_M$. Using order statistics [21], the probability distribution function of $\mu_j$, $\forall j$ is shown as follows:

$$k_j(\mu) = N_j F_\mu^{j-1}(\mu)\left[1 - F_\mu(\mu)\right]^{N-j} f_\mu(\mu) \tag{21}$$

where:

$$N_j \triangleq \frac{N!}{(j-1)!(N-j)!}, \tag{22}$$

and $f_\mu(\mu)$, $F_\mu(\mu)$ are the probability density function and probability distribution function of $\mu$. Assuming the same assumption as discussed above in Equation (9) we obtain:

$$f_\mu(\mu) = e^{-\mu}, F_\mu(\mu) = 1 - e^{-\mu}. \tag{23}$$

Equivalently:

$$k_j(\mu) = N_j(1-e^{-\mu})^{j-1}e^{-\mu(N-j+1)}. \tag{24}$$

Using a binomial expansion to replace $(1-e^{-\mu})^{j-1}$ in Equation (25) gives:

$$k_J(\mu) = N_j\sum_{l=0}^{j-1} F_l^{j-1}e^{-\mu(N-l)}, \tag{25}$$

where, $F_l^{j-1} \triangleq \begin{pmatrix} j-1 \\ l \end{pmatrix}(-1)^{j-1-l}$.

Thus, the optimization problem of maximizing the capacity of SUTCH, while satisfying the tolerable interference constraint of the PUs with selection policy $\rho_{ps}$ is shown as follows:

$$\begin{aligned} C_M^{\rho_{ps}} &= \max_{\boldsymbol{\theta_Q}} \sum_{j=1}^{M}\sum_{l=0}^{j-1} \Delta f N_j F_l^{j-1} \frac{\theta_{Q_j}\log\left[(N-l)\theta_{Q_j}\right]}{(N-l)\theta_{Q_j}-1}, \\ s.t. &\quad \sum_{j=1}^{M}\theta_{Q_j} = GN\theta_{Q_j}, \quad 0 \le \theta_{Q_j} \le GN\theta_Q. \end{aligned} \tag{26}$$

However, in practice, $M < N$, thus, $N\theta_{Q_j} \gg 1$. Therefore, Equation (27) can be approximated by Equation (28):

$$\begin{aligned} C_M^{\rho_{ps}} &\approx \max_{\boldsymbol{\theta_Q}} \sum_{j=1}^{M}\sum_{l=0}^{j-1} \Delta f \frac{N_j F_l^{j-1}}{N-l}\log\left[(N-l)\,\theta_{Q_j}\right], \\ s.t. &\quad \sum_{j=1}^{M}\theta_{Q_j} = GN\theta_{Q_j}, \quad 0 \le \theta_{Q_j} \le GN\theta_Q. \end{aligned} \tag{27}$$

The Lagrange multiplier algorithm can be used to solve the optimization problem in Equation (28) [19]:

$$\begin{aligned} L(\theta_{Q_j},\lambda) &= \sum_{j=1}^{M}\sum_{l=0}^{j-1} \frac{N_j F_l^{j-1}}{N-l}\log\left[(N-l)\,\theta_{Q_j}\right] \\ &\quad -\lambda\left(\sum_{j=1}^{M}\theta_{Q_j} - GN\theta_Q\right) \end{aligned} \tag{28}$$

where, $\lambda$ is the Lagrangian coefficient.

Taking the derivative with respect to the $\theta_{Q_j}$ on Equation (29) and setting it equal to zero gives:

$$\theta_{Q_j}^* = \frac{1}{\lambda^*}v_j, \tag{29}$$

where, $v_j \triangleq \sum_{l-0}^{j-1} N_j F_l^{j-1}/N-l$. Substituting Equation (30) into Equation (28):

$$\lambda^* = \frac{1}{GN\theta_Q}\sum_{j=1}^{M} v_j. \tag{30}$$

Substituting Equation (31) into Equation (30):

$$\theta_{Q_j}^* = GN\theta_Q \frac{v_j}{\sum_{j=1}^{M} v_j}. \tag{31}$$

Furthermore, according to Equation (3) and $\theta_{Q_j} \triangleq Q_j/KN_0\Delta f$, the optimal transmitting power vector $\mathbf{P}_s^*$ with selection policy $\rho_{ps}$ can be obtained as follows:

$$\mathbf{P}_s^* = \frac{GQ}{\sum_{j=1}^{M} v_j}\left(\frac{v_1}{g_{ps}^j}, \frac{v_2}{g_{ps}^j}, \ldots, \frac{v_M}{g_{ps}^j}\right) \tag{32}$$

Equivalently, substituting Equation (33) into Equation (28) yields the approximated maximum achievable capacity of the SUTCH with selection policy $\rho_{ps}$, which is shown in Equation (34):

$$C_M^{\rho_{ps}} \approx \sum_{j=1}^{M}\sum_{l=0}^{j-1} \frac{\Delta f N_j F_l^{j-1}}{N-l} \log\left[(N-l)\, GN\theta_Q \frac{v_j}{\sum_{j=1}^{M} v_j}\right] \tag{33}$$

Considering $\rho_{ss}$, the selection criteria can be assumed as follows:

$$\psi\left(g_{ps}^j, g_{ss}^j\right) = g_{ss}^j \tag{34}$$

Consequently, $\mu_j = g_{ss}^j$ and based on $\mu_j$, $j$ = 1, 2, … , $M$, the selection sequence is defined as follows:

$$\gamma_M = (\mu_1, \mu_2, \ldots, \mu_m) \triangleq \rho_{ss}(\mu_1, \mu_2, \ldots \mu_m). \tag{35}$$

where $\mu_1 \geq \mu_2 \geq \ldots \geq \mu_M$. Using order statistics [21], the probability distribution function of $\mu_j$, $\forall j$ is shown as follows:

$$k_j(\mu) = N_j \mathrm{F}_\mu^{N-j}(\mu)\left[1 - \mathrm{F}_\mu(\mu)\right]^{j-1} f_\mu(\mu), \tag{36}$$

Using a binomial expansion to replace $(1 - e^{-\mu})^{N-j}$ in Equation (37) one obtains:

$$k_J(\mu) = N_j \sum_{l=0}^{N-j} F_l^{N-j} e^{-\mu(l+j)}, \tag{37}$$

where $F_l^{N-j} \triangleq \begin{pmatrix} N-j \\ l \end{pmatrix}(-1)^l$.

Thus the optimization problem of maximizing the capacity of the SUTCH while satisfying the tolerable interference constraints of the PUs with selection policy $\rho_{ps}$ is shown as follows:

$$\begin{aligned} C_M^{\rho_{ss}} \quad &= \quad \max_{\theta_Q} \sum_{j=1}^{M}\sum_{l=0}^{N-j} \Delta f \frac{N_j F_l^{N-j}}{l+j} \frac{\frac{\theta_{Q_j}}{l+j}}{\frac{\theta_{Q_j}}{l+j} - 1} \log\left(\frac{\theta_{Q_j}}{l+j}\right) \\ s.t. \quad &\sum_{j=1}^{M} \theta_{Q_j} = GN\theta_{Q_j}, \qquad 0 \leq \theta_{Q_j} \leq GN\theta_Q. \end{aligned} \tag{38}$$

Utilizing the following approximation for small values of $\theta_{Q_j}/l + j$, $l$ = 0, 1, … , $N - j$ as:

$$\theta_{Q_j}^* = GN\theta_Q \frac{\chi_j^2}{\sum_{j=1}^{M} \chi_j^2} \tag{39}$$

where:

$$\chi_j \triangleq \sum_{l=0}^{N-j} N_j \frac{F_l^{N-j}}{2(l+j)^{3/2}} \tag{40}$$

Furthermore, according to Equation (3) and $\theta_{Q_j} \triangleq Q_j / KN_0\Delta f$, the optimal transmitting power vector $\mathbf{P}_s^*$ with selection policy $\rho_{ss}$ can be achieved as follows:

$$\mathbf{P}_s^* = \frac{GQ}{\sum_{j=1}^{M} \chi_j^2} \left( \frac{\chi_1^2}{g_{ss}^j}, \frac{\chi_2^2}{g_{ss}^j}, \ldots, \frac{\chi_M^2}{g_{ss}^j} \right) \tag{41}$$

Equivalently, substituting (41) into (39) yields the approximated maximum achievable capacity of the SUTCH under SU-SU-based selection policy is shown as follows:

$$C_M^{\rho_{ss}} \approx \sum_{j=1}^{M} \sum_{l=0}^{N-j} \Delta f \frac{N_j F_l^{N-j}}{(l+j)^{3/2}} \left( GN\theta_Q \frac{\chi_j^2}{\sum_{j=1}^{M} \chi_j^2} \right)^{1/2} \tag{42}$$

## 4. Reinforcement Learning for Improving Performance

In Section 3, the maximum achievable capacity of the SUTCH is analyzed. In Section 4, the reliability of the SUCCH is taken into consideration by the Bit Error Rate (BER). Suppose the signal waveform of the SUCCH is as follows:

$$\begin{cases} s_1(t) = \sqrt{\varepsilon_b} \\ s_2(t) = -\sqrt{\varepsilon_b} \end{cases} \tag{43}$$

Suppose the two signal waveforms in Equation (45) are transmitted with the same probability. Since the SUCCH spreads its power spectrum density over an ultra-wide bandwidth to exploit the underused (gray) spectral regions, the interference process caused by the PUs and the SUCCH can be considered as a Gaussian approximation. If the SUCCH transmits $s_1(t)$, after the despread-demodulation algorithm at the receiver of the SUCCH, the received signal is as follows:

$$r = \sqrt{\varepsilon_b} + \frac{1}{G_{\text{SUCCH}}} \left( n + \sum_{y=1}^{Y} \sigma_{PU} + \sum_{j=1}^{M} \sigma_{\text{SUCCH}} \right) \tag{44}$$

where $n$ is additive Gaussian white noise with mean zero, variance $N_0/2$ and $\sigma_{PU}$, $\sigma_{\text{SUCCH}}$ represent the interference caused by the PUs and the SUTCH. $G_{\text{SUCCH}}$ is the spreading gain of the SUCCH. The receiving signal of the SUCCH is compared with the threshold zero, which is as follows:

$$r \underset{s_2}{\overset{s_1}{\gtrless}} 0. \tag{45}$$

Suppose the PUs and the SUCCH are i.i.d. random processes, then two probability density functions of $r$ are given as follows:

$$\begin{aligned} p(r|s_1) &= \frac{1}{\sqrt{\frac{2\pi}{G_{\text{SUCCH}}} \left( \frac{N_0}{2} + \sum_{y=1}^{Y} \sigma_{pu}^2 + \sum_{j=1}^{M} \sigma_{\text{SUTCH}}^2 \right)}} e^{-(r-\sqrt{\varepsilon_b})^2/N_0} \\ p(r|s_2) &= \frac{1}{\sqrt{\frac{2\pi}{G_{\text{SUCCH}}} \left( \frac{N_0}{2} + \sum_{y=1}^{Y} \sigma_{pu}^2 + \sum_{j=1}^{M} \sigma_{\text{SUTCH}}^2 \right)}} e^{-(r+\sqrt{\varepsilon_b})^2/N_0} \end{aligned} \tag{46}$$

Consequently, the average error probability of the SUCCH is as follows:

$$\begin{aligned} P_e &= \tfrac{1}{2}\left[P\left(e|s_1\right) + P\left(e|s_2\right)\right] \\ &= \tfrac{1}{2}\left[\int_{-\infty}^{0} p\left(r|s_1\right)dr + \int_{0}^{+\infty} p\left(r|s_2\right)dr\right] \\ &= Q\left(\sqrt{\frac{G_{\mathrm{SUCCH}}\varepsilon_b}{\frac{N_0}{2} + \sum_{y=1}^{Y}\sigma_{PU}^2 + \sum_{j=1}^{M}\sigma_{\mathrm{SUCCH}}^2}}\right) \end{aligned} \tag{47}$$

Suppose the control information of the SUCCH consists of 8 bits. According to Figure 4, the transmitter and receiver of the SUs need to coordinate access to the spectrum three times. Therefore, the probability of successful establishment for the SUCCH can be concluded. Furthermore, the total interference caused by the SUs is divided into two parts: $Q_{\mathrm{SUTCH}}$ and $Q_{\mathrm{SUCCH}}$. $Q_{\mathrm{SUTCH}}$ represents the interference caused by the activity of the SUTCH, while $Q_{\mathrm{SUCCH}}$ represents the interference caused by the activity of the SUTCH. The loading factor $\Gamma$ is defined as the radio of $Q_{\mathrm{SUTCH}}$ and $Q_{\mathrm{SUCCH}}$, which is as follows:

$$\Gamma \triangleq \frac{Q_{\mathrm{SUCCH}}}{Q_{\mathrm{SUTCH}}}, \quad 0 < \Gamma < 1 \tag{48}$$

In consideration of the link access protocol design described above and the probability of successful establishment for SUCCH, the lower PSD of SUCCH means it may take more time to complete the setup procedure for the SUs. In other words, accessible subchannels will remain idle for a long period of time, which will lead to spectrum resource waste. However, increasing the transmitting power of the SUCCH will decrease the transmitting power of the SUTCH, because of the total interference constraint caused by the SUs is certain at a time-slot. Lower transmitting power of the SUTCH will lead to reduce the capacity of data. Therefore, it's a trade-off, which is essential to choose the appropriate transmitting power of SUTCH according to the characteristic of the activity of the PU. For this purpose, a hybrid access method based on Reinforcement Learning model is proposed to solve this problem. The most prominent feature of Reinforcement Learning model is its autonomous learning and online learning ability. By trial and error, Reinforcement Learning model can get a better strategy based on the subchannel environment.

The Cross model [22] is now widely recognized as one of the Reinforcement Learning models with memory-less characteristics, which means the learning process is a Markov Decision Process (MDP). The basic idea is to follow the rules of "Results" [23], namely, if system is rewarded by choosing a strategy, then the next period will get higher probability of choosing such strategy. On the contrary, if it is punished, the next period will reduce the probability of choosing such strategy.

Bush and Mosteller [24] introduced the Bush-Mosteller model in 1955 [25]. Afterwards, Roth and Erev improved this model and introduced the Roth-Ever model. Nowadays, as two models of reinforcement learning, both of them [26] are widely adopted. They are easy to realize and have very low computation complexity, which fit for the real-time applications. Therefore, in this paper, these two models are introduced and some necessary modifications are adopted for the application, so the model of MDP Cross and Statistical Mean are proposed.

As mentioned above, the process of connection setup is defined as the time-slotted. The optional strategies for the SUs are defined as follows:

$$\mathbf{A}_{su} = \left(\Gamma_1, \Gamma_2, \ldots, \Gamma_n, \ldots, \Gamma_{n'}, \Gamma_R\right) \tag{49}$$

where $\mathbf{A}_{su}$ is the vector of optional strategies, $R$ the number of the strategy, $n$ is the chosen strategy and $n'$ are not chosen strategies in a certain time-slot.

Consequently, during the time-slot $k$ to access the initial stage, the SUs can update the probability of choosing strategy $n$ and $n'$ by the following formula:

$$\begin{array}{l} p^n\left(k+1\right)=p^n\left(k\right)+R\left[u\left(k\right)\right]\times\left(1-p^n\right) \quad n=A_{su}\left(k\right) \\ p^{n'}\left(k+1\right)=p^{n'}\left(k\right)-R\left[u\left(k\right)\right]\times p^{n'}\left(k\right) \quad n'\neq A_{su}\left(k\right) \\ R\left[u\left(k\right)\right]=\alpha\times u\left(k\right)+\beta \end{array} \tag{50}$$

where $A_{su}(k)$ is the accessible strategy of the SUs at the time-slot $k$, which can be seen the action of MDP. $p^n(k)$ is the probability of the accessible strategy $n$ of the SUs at time-slot $k$, $p^n(k)$ is the probability of the unused strategy $n'$ of the SUs at time-slot $k$, which can be seen as the state of MDP. $u(k)$ is the reward function of the accessible performance of the SUs, which can be seen as the reward of MDP. $\alpha$ and $\beta$ are the adjustment factors, which can be used to determine the updating rate of $u(k)$. $R[u(k)]$ is defined as the monotone function of $u(k)$, which is $-1 < R[u(k)] < 1$. When the SUTCH successfully accesses idle subchannels, it obtains the reward, which is defined as follows:

$$\partial_1 I\left(k\right)C_{\mathrm{SUTCH}}\left(k\right)T\left(k\right) \tag{51}$$

where, $T(k)$ is the transmission duration of the SUs in time-slot $k$ and $\partial_1$ is a weighting factor and $I(k)$ is indicator function, which is defined as follows:

$$\begin{cases} I\left(k\right)=1 & \text{SUTCH successfully access at time-slot } k \\ I\left(k\right)=0 & \text{SUTCH fail to access at time-slot } k \end{cases} \tag{52}$$

When the SUTCH fails to access the idle subchannels, it wastes the opportunity for transmission and pays the cost, which is shown as follows:

$$-\partial_2 I\left(k\right)C_{\mathrm{SUTCH}}\left(k\right)T'\left(k\right) \tag{53}$$

where, $T'(k)$ is the access duration of the SUs and $\partial_2$ is also a weighting factor.

Equivalently:

$$u\left(k\right)=\partial_1 I\left(k\right)C_{\mathrm{SUTCH}}\left(k\right)T\left(k\right)-\partial_2 I'\left(k\right)C_{\mathrm{SUTCH}}\left(k\right)T'\left(k\right) \quad 0\leq\partial_i\leq 1, i=1,2 \tag{54}$$

In order to weaken the impact of weighting on updating the probability of the choosing strategy, Equation (52) can be further defined as follows:

$$\begin{array}{ll} p^n\left(k+1\right)=p^n\left(k\right)+\varepsilon\times\left[1-p^n\left(k\right)\right] & n=A_{su}\left(k\right),u\left(k\right)>0 \\ p^n\left(k+1\right)=p^n\left(k\right)-\varepsilon\times p^n\left(k\right) & n=A_{su}\left(k\right),u\left(k\right)<0 \\ p^{n'}\left(k+1\right)=p^{n'}\left(k\right)+\varepsilon\times\left[1-p^{n'}\left(k\right)\right] & n'\neq A_{su}\left(k\right),u\left(k\right)<0 \\ p^{n'}\left(k+1\right)=p^{n'}\left(k\right)-\varepsilon\times p^{n'}\left(k\right) & n'\neq A_{su}\left(k\right),u\left(k\right)>0 \end{array} \tag{55}$$

where, $\varepsilon = R[u(k)] = \alpha \times u(k) + \beta$. The solution to update the probability of choosing strategy is the model of MDP Cross. If the $u(k) > 0$, which means the accessible strategy $n$ is fit for the current subchannel environment. Therefore, the $p^n(k+1)$ should be increased, while the $p^{n'}(k+1)$ should be decreased. However, if the $u(k) < 0$, which means the accessible strategy $n$ is not fit for the current subchannel environment, therefore, the $p^n(k+1)$ should be decreased, while the $p^{n'}(k+1)$ should be increased.

In practice, the probability of choosing a strategy is usually not only dependent on the latest result, it also takes the "system history" into account. "System history" presents users with more information about the status of environment. In order to incorporate the "system history", the Statistical Mean is proposed, in which the reward function is modified as follows:

$$\begin{array}{l} p_{suc}^n\left(k\right)=F_{suc}^n\left(k\right)/F_{access}^n\left(k\right) \\ p_{fail}^n\left(k\right)=F_{fail}^n\left(k\right)/F_{access}^n\left(k\right) \\ u\left(k\right)=\partial_1 p_{suc}^n\left(k\right)-\partial_2 p_{fail}^n\left(k\right) \end{array} \tag{56}$$

where, $F_{suc}^{n}(k)$ represents the amount of data traffic which SUTCH has transmitted based on strategy $n$ at time-slot $k$, $F_{access}^{n}(k)$ and $F_{fail}^{n}(k)$ are the idea and wasted amount, respectively.

Therefore the probability of choosing a strategy in the Statistical Mean is shown as follows:

$$\begin{array}{ll} p^{n}(k+1) = p^{n}(k) + \varepsilon \times [1 - p^{n}(k)] & n = A_{su}(k), \forall j, j \neq n, u^{n}(k+1) > u^{j}(k+1) \\ p^{n}(k+1) = p^{n}(k) - \varepsilon \times p^{n}(k) & n = A_{su}(k), \exists j, j \neq n, u^{n}(k+1) \leq u^{j}(k+1) \\ p^{n'}(k+1) = p^{n'}(k) + \varepsilon \times \left[1 - p^{n'}(k)\right] & n' \neq A_{su}(k), u^{n}(k+1) \leq u^{n'}(k+1) \\ p^{n'}(k+1) = p^{n'}(k) - \varepsilon \times p^{n'}(k)\Big) & n' \neq A_{su}(k), u^{n}(k+1) > u^{n'}(k+1) \end{array} \quad (57)$$

## 5. Simulation Study

In this section, the achievable spectrum efficiencies with different subchannel selection policies are compared. Here, the spectrum sharing load factor is $\theta_Q = -30$ dB and the number of subchannels is $N = 40$. The mean values of random variables $g_{ps}^{j}, g_{ss}^{j}$ are denoted by $\lambda_{ps}, \lambda_{ss}$, respectively. The achieved spectrum efficiency is defined as follows:

$$C_{\rho_\psi} = C_{M}^{\rho_\psi} / M\Delta f \quad (58)$$

Here, in order to facilitate the comparison, $C_{\rho_1}$ is defined as the achieved spectrum efficiency with uniform subchannel selection, $C_{\rho_{ss}}$ is defined as the achieved spectrum efficiency with the SU-SU-based selection policy, $C_{\rho_{ps}}$ is defined as the achieved spectrum efficiency with the SU-PU-based selection policy.

In the first simulation, suppose the interference threshold is a constant and $\lambda_{ps} = \lambda_{ss}$, and the $C_{\rho_\psi}$ is analyzed by increasing $M$, which is depicted in Figure 6.

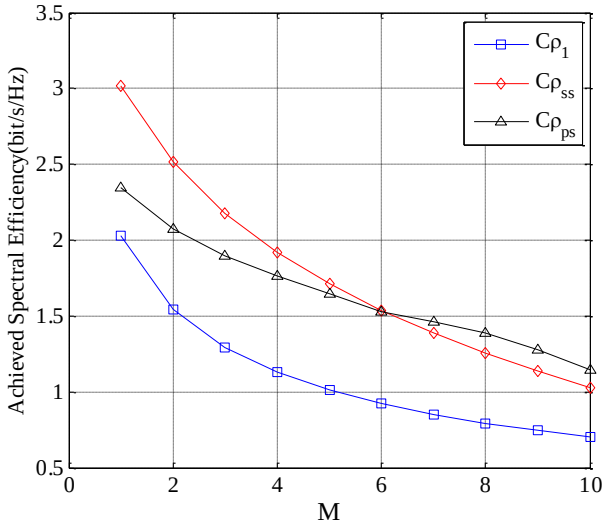


**Figure 6.** Achieved spectral efficiency of the SUTCH with three selection policies with $M$.

As depicted in Figure 6, $C_{\rho_1}$ is lower than that of $C_{\rho_{ss}}$ and $C_{\rho_{ps}}$, therefore, it indicates that $\rho_1$ has a poorer performance compared to $\rho_{ss}$ and $\rho_{ps}$. For $M = 1$, the gap between $C_{\rho_1}$ and $C_{\rho_{ss}}$ is large. However, with the increase of $M$, the gap is reduced. This result is reasonable because the tap is related to the $M/N$ ratio, and the larger $M/N$, the lower the tap is. The reason is that for a larger $M/N$, the set of $M$ subchannels accessible by $C_{\rho_1}$ and $C_{\rho_{ss}}$ probably has a large overlap.

With the increase of $M$, the rate of decrease of $\rho_{ps}$ is reduced with the slowest rate. This is mainly due to the fact that the total interference threshold of the receivers of the PUs is a constant. At the same time, $\rho_{ps}$ selects these subchannels with the lower $g_{ps}^{j}$, which enables the SUs transmitters to send the maximum transmitting power, without generating high interference on the receivers of the

Pus and satisfying the constraint of the interference threshold of the PUs. According to Figure 6, for a large number of accessible subchannels with constant interference constraint, $\rho_{ps}$ achieves a better performance.

In the second simulation, the influence of the number of subchannels $N$ is analyzed. Suppose $M = 1$, $\lambda_{ps} = \lambda_{ss}$, the $C_{\rho_\psi}$ is analyzed by increasing $N$. The result is depicted in Figure 7.

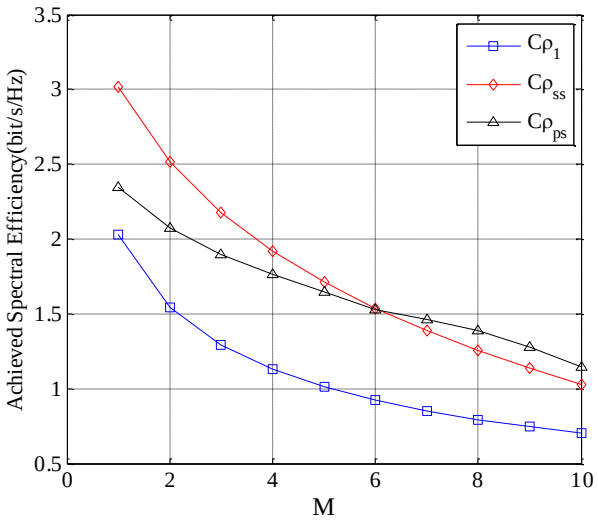


**Figure 7.** Achieved spectral efficiency of the SUTCH under three selection policies with $N$.

As seen in Figure 7, for all the different subchannel selection policies, the $C_{\rho_\psi}$ increases with the increase of $N$. This is because that the probability of selecting proper subchannels for SUTCH is increasing with $N$. Furthermore, it is interesting to find that the gap between these three selection policies also increases with the increase of $N$ and $\rho_{ss}$ outperforms the others in this simulation.

In the third simulation, both the influences of $g_{ps}$ and $g_{ss}$ are evaluated. Suppose $N = 40$, $M = 1$. The $C_{\rho_\psi}$ is analyzed with $\lambda_{ps}/\lambda_{ss}$ for different $\theta_Q$ values. The simulation result is depicted in Figure 8.

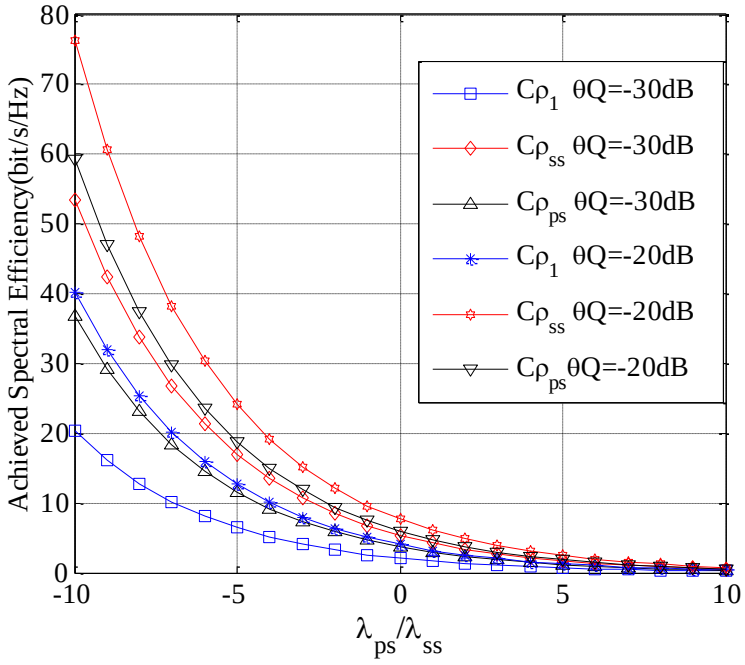


**Figure 8.** Achieved spectral efficiency of the SUTCH under three selection policies with $\lambda_{ps}/\lambda_{ss}$.

As depicted in Figure 8, it is clearly observed that the $C_{\rho_\psi}$ of the SUTCH decreases with the increase of $\lambda_{ps}/\lambda_{ss}$. Meantime, the $C_{\rho_\psi}$ of the SUTCH decreases with the decrease of $\theta_Q$. This is due to the fact that with the increase of $\lambda_{ps}/\lambda_{ss}$, the attenuation of $g_{ps}$ is decreased while that of $g_{ss}$ is increased. Consequently, the $C_{\rho_\psi}$ of SUTCH is lower with the same transmitting power. On the other hand, with the decrease of $\theta_Q$, the power allocated to each selected subchannel is bound to be reduced, which will lead to the deterioration in the $C_{\rho_\psi}$ of the SUTCH.

Compared comprehensively, the $C_{\rho_\psi}$ of the SUTCH with $\rho_1$ has the lowest value, since it just ignores any a priori knowledge of subchannel's status. However, under different conditions, the performance of the $\rho_{ss}$ and $\rho_{ps}$ are different. When the ratio of $M/N$ is small, the best subchannel selection policy is $\rho_{ss}$. However, if the ratio of $M/N$ is large, the best subchannel selection policy is $\rho_{ps}$.

In the fourth simulation, as mentioned above, in Equation (49), the BER of SUCCH is derived. Therefore, Monte Carlo Simulation is used to prove its rationality. The simulation parameters are shown in Table 1. Suppose $\sigma^2_{PU} = \sigma^2_{\text{SUTCH}}$.

**Table 1.** Simulation parameters.

| Parameter | Value |
|---|---|
| $N$ | 40 |
| Number of active PUs and SUs | [1, 40] |
| $G_{\text{SUCCH}}$ | 2048 |
| Loading factor $\Gamma$ | [1/160, 1/200, ..., 1/400] |
| Random test times for each $\Gamma$ | 750,000 |

In Figure 9, the Simulation BER is calculated by Monte Carlo Simulation Experiment, while the Theoretical BER is calculated by Equation (49). As depicted in Figure 9, the simulation BER follows the Theoretical BER very closely.

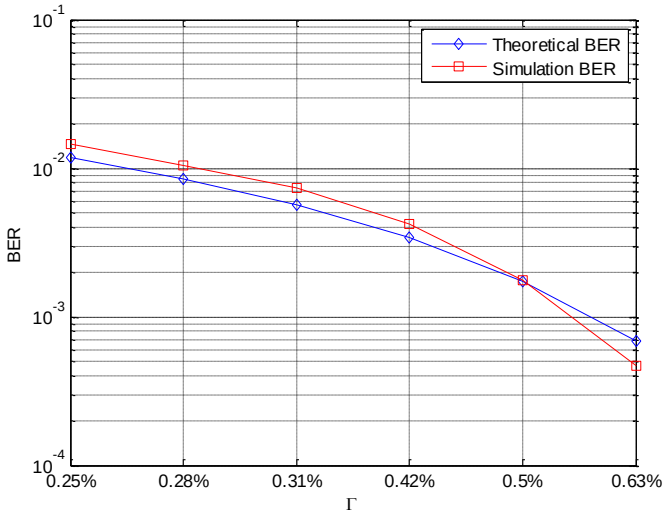


**Figure 9.** Theoretical and simulation BER versus different $\Gamma$.

As mentioned in Section 4, the trade-off problem between the reliability of the SUCCH and the efficiency of the SUTCH is discussed. Here, suppose the arrival rate of the authorized PUs accessing to the subchannels follows a Poisson distribution. Simulation parameters are shown in Table 2. Suppose $\lambda_m^j$ represents the arrival rate of the PUs in accessible subchannels.

In the fifth simulation, the achieved spectral efficiency, achieved data traffic and unused data traffic are used to compare the accessible performance of the three different selection policies. Here, achieved spectral efficiency represents the proportion between data traffic and unused data traffic.

Data traffic is the total amount of unit data traffic when the SUTCH has successfully accessed to the idle subchannel, while unused data traffic is the achievable amount of unit data traffic during the time cost in establishing the connection.

**Table 2.** Simulation parameters.

| Parameters | Values |
|---|---|
| $N$ | 40 |
| $M$ | [0, 6] |
| Number of the active PUs | 34 |
| Number of the SUs | 1 |
| $\lambda_{ps}$, $\lambda_{ss}$ | 1, 1 |
| $\lambda_m^j$, $j = 1, 2, \ldots, 6$ | [80, 160] |
| $Q$ | 0.0001 W |
| $Q_{\mathrm{SUCCH}}/Q$ | [0.01, 0.02, ..., 0.1] |
| $\theta_Q$ | −30 dB |
| $G$ | 128 |
| $e$ | 0.01 and 0.05 |
| $\partial_1$, $\partial_2$ | 0.005, 0.005 |
| $R$ | 10 |
| $P^n(k)$, $n = [1, R]$ | $[1/R]$ |
| Learning time | 100 (times of SUs access) |

In Figures 10–12, the different performances of the three strategies are shown in detail. Random strategy has the worst accessible performance, because it simply chooses the loading factor $\Gamma$ randomly without proper accessible strategies. Meanwhile, the accessible performance of MDP Cross is better than that of Statistical Mean. Furthermore, the fluctuation of performance curve of MDP Cross is lower than that of Statistical Mean. It is due to the fact that, in the simulation, suppose $\lambda_m^j$, $j = 1, 2, \ldots, 6 \in [80, 160]$ the state parameters of the accessible subchannel are changing very fast, therefore, it is a quick-changing subchannel environment. In the quick-changing subchannel environment, the history state information of subchannel environment is changing very fast. However, Statistical Mean will use a lot of history information, so the fast-changing of history information will make a bad influence on choosing the optimal allocation strategy of $\Gamma$. Therefore, the accessible strategy of MDP Cross fits better in the quick-changing subchannel environment.

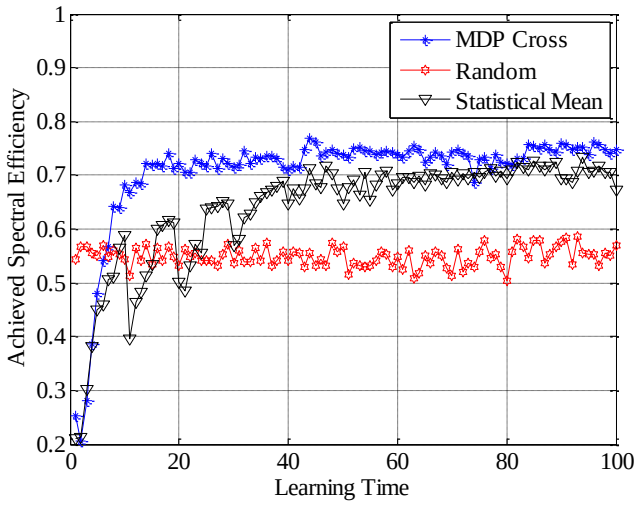


**Figure 10.** Achieved spectral efficiency of the SUTCH under three strategies with learning time.

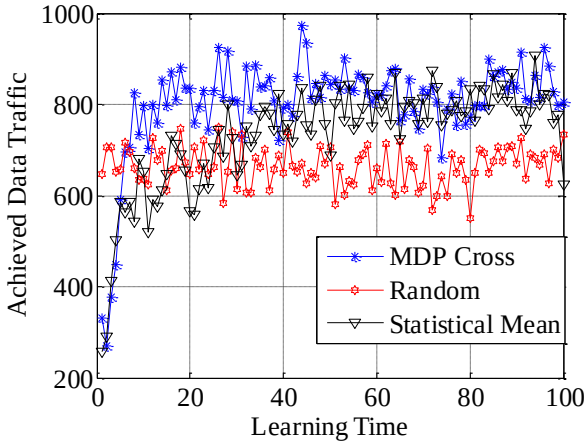


**Figure 11.** Achieved data traffic of the SUTCH under three strategies with learning time.

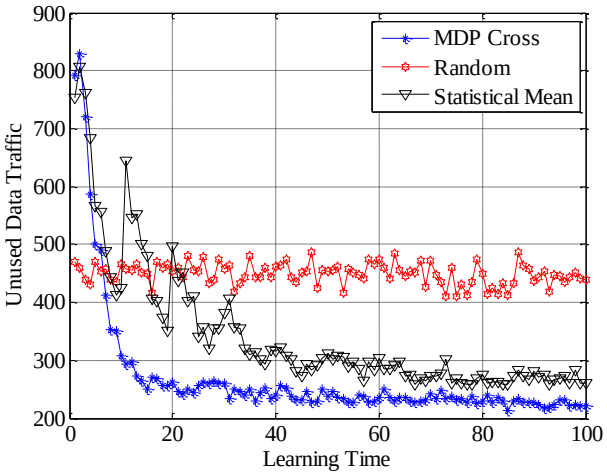


**Figure 12.** Unused data traffic of the SUTCH under three strategies with learning time.

In the sixth simulation, different performances of the three strategies under constant application scenarios are shown in Figures 13–15. Suppose $\lambda_m^j$ is defined as constant, which is shown as follows:

$$\left[\lambda_m^1, \lambda_m^2, \lambda_m^3, \lambda_m^4, \lambda_m^5, \lambda_m^6\right] = [1/90, 1/100, 1/110, 1/120, 1/130, 1/140]$$

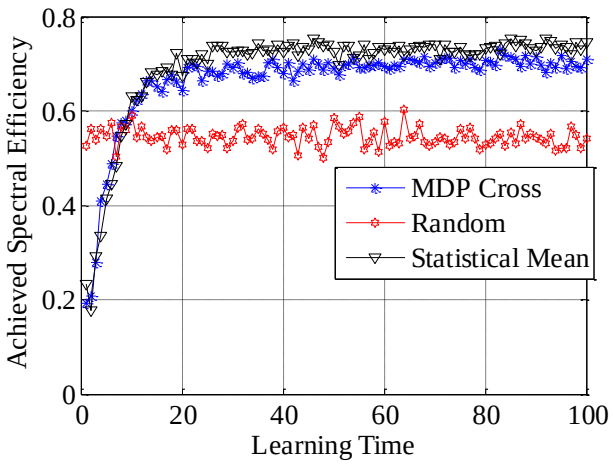


**Figure 13.** Achieved data traffic of the SUTCH under three strategies with learning time.

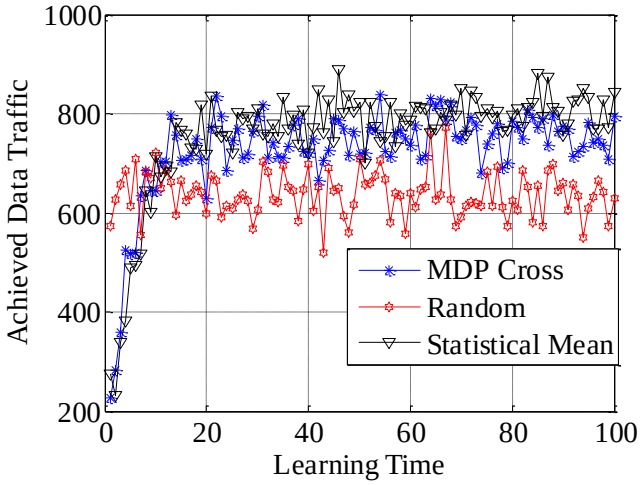


**Figure 14.** Achieved data traffic of the SUTCH under three strategies with learning time.

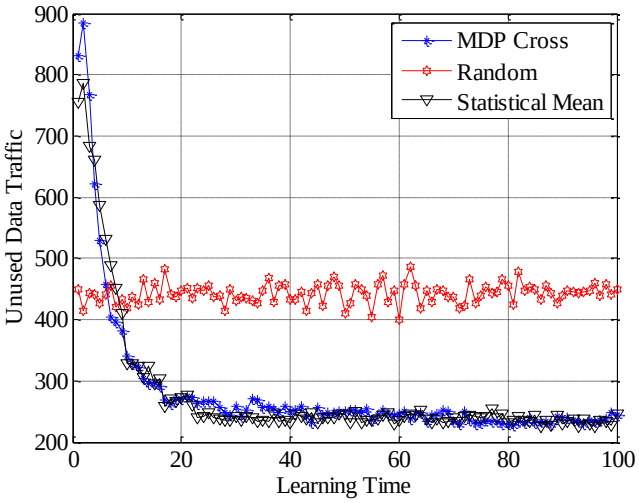


**Figure 15.** Unused data traffic of the SUTCH under three strategies with learning time.

As shown in these figures, the Random strategy still has the worst accessible performance. Meanwhile, the accessible performance of Statistical Mean is better than that of MDP Cross. Furthermore, the fluctuation of the performance curve of Statistical Mean is lower than that of MDP Cross. It is due to the fact that, in a slow-changing subchannel environment, the slow-changing of the history information will have a good influence on choosing the optimal allocation strategy of $\Gamma$. Therefore, the accessible strategy of Statistical Mean fits better in the slow-changing subchannel environment.

In addition, as shown from Figure 10 to Figure 15, both Statistical Mean and MDP Cross can learn and adapt to the subchannel environment, and converge to a stable state in a short time. Meanwhile, they have the same rate of convergence. According to the analysis in Section 4, both Statistical Mean and MDP Cross have low computation complexity. Therefore, they can be adopted in practice.

## 6. Conclusions

Dynamic spectrum access is an important and necessary technology for future cognitive sensor networks. This paper identified and discussed a new mechanism to set up CR sensor networks without using spectrum holes to convey control information. A transmission channel model was discussed for analyzing the maximum access capacity of different policies and objectives in the fading environment.

The maximum achievable capacity of the SUTCH under $\rho_1$ achieves the poorest performance, since it totally ignores any prior knowledge of the subchannel's status. When $M/N$ is small, the best policy for subchannel selection is $\rho_{ss}$. In contrast when this ratio is higher, $\rho_{ss}$ is better.

To solve the trade-off between transmitting power of SUTCH and SUCCH's capacity, a hybrid access method based on Reinforcement Learning model of MDP Cross and Statistical Mean is also proposed. Both of them outperform the Random strategy, which verified the effectiveness of the proposed methods. In addition, Statistical Mean is more suitable for slow variation application scenarios while MDP Cross performs better in fast variation scenarios.

As is well known, there are many standard structure and policy of reinforcement learning, such as Q-learning and greedy algorithm. Therefore, in the next research, the different learning function and policy should be discussed, which can make a better trade-off between the performance and computation complexity.

**Acknowledgments:** This paper is supported by the Key Development Program of Basic Research of China (JCKY2013604B001), Nation Nature Science Foundation of China (61301095), Nature Science Foundation of Heilongjiang Province of China (F201408) and the Fundamental Research Funds for the Central Universities (No. HEUCF100814 and HEUCF100816).

**Author Contributions:** Yun Lin analyzed the basic theory and wrote the paper, Chao wang designed the experiments and finish the simulation. Jiaxing wang analyzed the simulation result. Zheng Dou checked the research result.

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

## References

1. Zhao, Q.; Sadler, B.M. A Survey of Dynamic Spectrum Access: Signal Processing, Networking, and Regulatory Policy. *IEEE Signal Process. Mag.* **2007**, *24*, 79–89. [CrossRef]
2. Joshi, G.P.; Nam, S.Y.; Kim, S.W. Cognitive Radio Wireless Sensor Networks: Applications, Challenges and Research Trends. *Sensors* **2013**, *13*, 11197–11228. [CrossRef] [PubMed]
3. Liu, X. A Novel Wireless Power Transfer-Based Weighed Clustering Cooperative Spectrum Sensing Method for Cognitive Sensor Networks. *Sensors* **2015**, *15*, 27760–27782. [CrossRef] [PubMed]
4. Kondareddy, Y.R.; Agrawal, P.; Sivalingam, K. Cognitive radio network setup without a common control channel. In Proceedings of the 2008 IEEE Military Communications Conference (MILCOM 2008), San Diego, CA, USA, 16–19 November 2008; pp. 1–6.
5. Baldo, N.; Asterjadhi, A.; Zorzi, M. Dynamic spectrum access using a network coded cognitive control channel. *IEEE Trans. Wirel. Commun.* **2010**, *9*, 2575–2587. [CrossRef]
6. Cormio, C.; Chowdhury, K.R. An Energy-Efficient Spectrum-Aware Reinforcement Learning-Based Clustering Algorithm for Cognitive Radio Sensor Networks. *Sensors* **2015**, *15*, 19783–19818.
7. Khoshkholgh, M.G.; Navaie, K.; Yanikomeroglu, H. Access strategies for spectrum sharing in fading environment: Overlay, underlay and mixed. *IEEE Trans. Mob. Comput.* **2010**, *9*, 1780–1793. [CrossRef]
8. Gastpar, M. On Capacity under Receive and Spatial Spectrum Sharing Constraints. *IEEE Trans. Inf. Theory* **2007**, *53*, 471–487. [CrossRef]
9. Viterbi, A.J. *CDMA: Principles of Spread Spectrum Communication*; Addison-Wesley: Redwood City, CA, USA, 1995.
10. Chakravarthy, V.; Wu, Z.; Temple, M.; Garber, F. Novel Overlay/Underlay Cognitive Radio Waveforms Using SD-SMSE Framework to Enhance Spectrum Efficiency—Part I: Theoretical Framework and Analysis in AWGN Channel. *IEEE Trans. Commun.* **2009**, *57*, 3794–3804. [CrossRef]
11. Chakravarthy, V.; Wu, Z.; Temple, M. Novel Overlay/Underlay Cognitive Radio Waveforms Using SD-SMSE Framework to Enhance Spectrum Efficiency—Part II: Analysis in Fading Channel. *IEEE Trans. Commun.* **2010**, *58*, 1868–1876. [CrossRef]
12. Jasbi, F.; So, D.K. Hybrid Overlay/Underlay Cognitive Radio Network with MC-CDMA. *IEEE Trans. Veh. Technol.* **2016**, *65*, 2038–2047. [CrossRef]
13. Su, H.; Zhang, X. Cross-Layer Based Opportunistic MAC Protocols for QOS Provisioning over Cognitive Radio Wireless Networks. *IEEE J. Sel. Areas Commun.* **2008**, *26*, 118–129. [CrossRef]

14. Gupta, P.; Kumar, P.R. The Capacity of Wireless Networks. *IEEE Trans. Inf. Theory* **2000**, *46*, 388–404. [CrossRef]
15. Tse, D.; Viswanath, P. *Fundamentals of Wireless Communication*; Cambridge University Press: Cambridge, UK, 2004.
16. Jafar, S.A.; Srinivasa, S. Capacity Limits of Cognitive Radio with Distributed and Dynamic Spectral Activity. *IEEE J. Sel. Areas Commun.* **2007**, *25*, 529–537. [CrossRef]
17. Ghasemi, A.; Sousa, E.S. Fundamental Limits of Spectrum Sharing in Fading Environments. *IEEE Trans. Wirel. Commun.* **2007**, *6*, 649–658. [CrossRef]
18. Ross, S.M. *A First Course in Probability*; University of Southern California Press: Los Angeles, CA, USA, 2012.
19. Khoshkholgh, M.G.; Navaie, K.; Yanikomeroglu, H. Achievable Capacity in Hybrid DS-CDMA/OFDM Spectrum-Sharing. *IEEE Trans. Mob. Comput.* **2010**, *9*, 765–777. [CrossRef]
20. Boyd, S.; Vandenberghe, L. *Convex Optimization*; Cambridge University Press: Cambridge, UK, 2004.
21. Papoulis, A.; Pillai, S.U. *Probability, Random Variables, and Stochastic Processes*, 4th ed.; McGraw-Hill: New York, NY, USA, 2002.
22. Kaelbling, L.P. Reinforcement Learning: A Survey. *J. Artif. Intell. Res.* **1996**, *4*, 237–285.
23. Brodersen, R.W.; Wolisz, A.; Cabric, D.; Mishra, S.M.; Willkomm, D. *CORVUS: A Cognitive Radio Approach for Usage of Virtual Unlicensed Spectrum*; White Paper; University of Berkeley: Berkeley, CA, USA, 2004.
24. Bush, R.R.; Mosteller, F. *Stochastic Models for Learning*; John Wiley & Sons: New York, NY, USA, 1955.
25. Roth A, E.; Erev, I. Learning in Extensive Form Games: Experimental Data and Simple Dynamic Models in the Intermediate Run. *Games Econ. Behav.* **1995**, *8*, 164–212. [CrossRef]
26. Li, J.; Bai, C.; Peng, H. Review of Learning Model and Experiment Based on Learning Theory. *Sci. Technol. Manag. Res.* **2013**, *6*, 143–150.