Dynamic Bayesian Spectrum Bargaining with Non-myopic Users

Yang Yan[1], Jianwei Huang[2], Ming Zhao[1], and Jing Wang[1]

[1] State Key Laboratory on Microwave and Digital Communications,
Tsinghua National Laboratory for Information Science and Technology,
Department of Electronic Engineering, Tsinghua University, Beijing 100084, China
[2] Department of Information Engineering, CUHK, Hong Kong, China
yanyang07@mails.tsinghua.edu.cn, jwhuang@ie.cuhk.edu.hk,
{zhaoming,wangj}@mail.tsinghua.edu.cn

Abstract. In this paper, we investigate a cooperative spectrum sharing mechanism realized by a *dynamic Bayesian spectrum bargaining* between a pair of *non-myopic* primary user and secondary user. The primary user has only incomplete information of the secondary user's energy cost. We model such a bargaining process as a dynamic Bayesian game, and discuss the equilibria under all possible system parameters. Furthermore, we discuss in details the *sequential equilibrium* where the *reputation effect* plays an important role. Our analysis shows that the secondary user with a low energy cost can exploit the primary user's lack of complete information for its own benefits.

Keywords: dynamic Bayesian spectrum bargaining, cooperative spectrum sharing, sequential equilibrium, incomplete information, game theory, reputation effect.

1 Introduction

Rapid growth of today's wireless data communications leads to spectrum resource scarcity and demands more efficient resource allocation schemes. *Cooperative spectrum sharing (CSS)* is one class of mechanisms that can greatly enhance the efficiency of spectrum utilization through user cooperations. One possible CSS mechanism is to allow a licensed primary user (PU) with a poor channel link between its transmitter and receiver to improve its data rate by using a secondary user (SU) as a relay. Such interaction can increase the PU's data rate through creating cooperation diversity [1]. In fact, the benefit of cooperative communication has been well studied in the literature [2,3], and cooperative communication has already been incorporated into various wireless communication standards (*e.g.,* IEEE 802.16J standard [4]).

Different from the traditional cooperative communication technology, CSS also considers the compensation to the SU for its relay effort to create a *win-win* situation. The key issue that we study in this paper is how to determine the proper compensation with incomplete network information. A brief illustration

P. Ren et al. (Eds.): WICON 2011, LNICST 98, pp. 164–179, 2012.

of the network topology and interaction scheme is shown in Fig. 1. Detailed notations will be introduced in Section 2.1.

CSS with complete network information has been considered in [5–7]. Reference [8] considered a contract-based CSS in a *static* network. Our prior work [9] analyzed the CSS with *incomplete* information in a *dynamic* network environment. The focus of [9] is to consider the multi-stage bargaining between one PU and one SU within a *single* time slot. Our recent work [10] considered the CSS between one PU and one SU over multiple time slots, where the reputation effect happens due to incomplete information of SU's energy cost. In [10], SU is non-myopic but PU is assumed to be myopic. Our current paper extends the analysis to a more realistic where the PU is also non-myopic and wants to maximize its benefit in the long run. The analysis turns out to be much more involved compared with the one in [10].

In this paper, we analyze a bargaining-based CSS between one PU and one SU over a finite number of time slots, where both users are non-myopic and want to maximize their long-term utilities. The main results and contributions of this paper are as follows:

- *Non-myopic players:* We assume both PU and SU are non-myopic rational players, who maximize their long-term utilities. This assumption better captures the reality in dynamic spectrum bargaining.
- *Incomplete information and sequential equilibrium:* We model the SUs energy cost as the incomplete information to PU, and characterize the sequential equilibrium of the bargaining game.
- *Reputation effect:* We show that a weaker SU can take advantage of the incomplete information by establishing a strong reputation to obtain a higher long-term utility by establishing a certain reputation.

The rest of the paper is organized as follows. We introduce the system model in Section 2. In Section 3, we analyze and summarize the equilibria of the multi-slot bargaining under different system parameters. In Section 4, we focus on discussing when and how the reputation effect will affect the equilibrium. In Section 5, we discuss about the equilibrium outcome by comparing with the model under complete information. Finally, we conclude in Section 6.

2 PU-SU Cooperation and Bargaining Model

2.1 Cooperative Communication

We consider a time-slotted system with the network model as shown in Fig. 1, where one PU bargains with one SU about the spectrum allocation scheme. Here, TP and RP represent PU's transmitter and receiver, and TS and RS represent SU's transmitter and receiver. Parameter h_p, h_s, h_{ps}, and h_{sp} denote the fixed channel gains of the link TP-RP, TS-RS, TP-TS, and TS-RP, respectively. We further assume that both PU and SU know the channel gains of all links through a proper feedback mechanism. The PU and SU transmit with fixed power P_t and P_s, respectively.

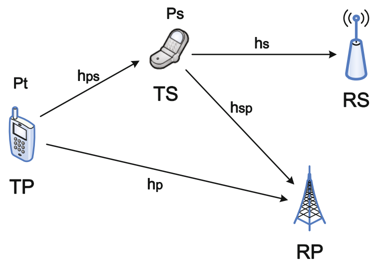

Fig. 1. PU-SU Cooperation Model

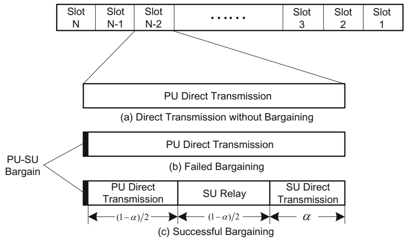

Fig. 2. The Slotted System Model with Three Possible Bargaining Results in a Single Time Slot

2.2 Dynamic Spectrum Bargaining

The bargaining process consists of N successive time slots as shown in Fig. 2. In order to facilitate later backward induction analysis, *we index time backwards, i.e., the bargaining starts with time slot N and ends in time slot 1.* Without loss of generality, we normalize each time slot length to 1. During each time slot, the PU can choose either *direct transmission only* or *bargaining with the SU.* There are three possibilities following PU's two options, which are illustrated in Fig. 2.

– Figure 2(a): If PU's direct channel gain h_p is good enough, then it will choose direct transmission only and achieves a data rate $R_{dir} = \log(1 + P_t h_p)$. In this case, The SU cannot transmit and achieves a zero utility.
– Figure 2(b): If PU believes that the SU's cooperation may be beneficial, it can offer α fraction of the time slot for SU's own transmission as remuneration for SU. If SU rejects the offer, PU proceeds with direct transmission for the remaining time.
– Figure 2(c): If SU accepts PU's offer α, then the PU and SU work in the amplified and forward (AF) relay mode. The PU achieves a data rate (per unit time) [1]

$$R_r = \log\left(1 + P_t h_p + \frac{P_t P_s h_{ps} h_{sp}}{P_t h_{ps} + P_s h_{sp} + 1}\right), \tag{1}$$

and the SU achieves a data rate (per unit time)

$$R_s = \log(1 + P_s h_s). \tag{2}$$

PU and SU will bargain over the value of α in each time slot after observing the bargaining history and anticipating the future, so as to maximize their own long-term utilities.

2.3 Incomplete Information of Energy Cost

We assume that the SU is an *energy-constrained* device (*e.g.,* wireless sensor or mobile device) with an energy cost C, which belongs to one of two types: High type C_h and Low type C_l with $(C_h > C_l)$. The SU knows its own type, but the PU does not. However, the PU has a *belief* on C's distribution in each time slot $(n = N, ..., 1)$, *i.e.,* $\Pr(C = C_h) = q_n$ and $\Pr(C = C_l) = 1 - q_n$. The belief will be updated based on the interactions between the PU and SU.

3 Multi-Slot Spectrum Bargaining Game with Non-myopic Players

In this section, we will explore how the PU and SU maximize their utilities in this dynamic Bayesian bargaining game. We assume that both PU and SU are the non-myopic players, who will maximize their total utilities in the N time slots. As the first step, we will study the single-slot bargaining game (*e.g.,* $N = 1$), which serves as a base for the study of the multi-slot case later on.

3.1 Utility Function in the Single-Slot Game

The SU's single-slot utility $U_s(\alpha)$ after accepting an offer α is

$$U_s(\alpha) = \alpha R_s - \frac{1+\alpha}{2} P_s C, \tag{3}$$

which is the difference between the SU's achievable data rate R_s (as in (2)) and energy cost. If we view C as the data rate per *watt* that the SU can get if it does not relay for PU, then $U_s(\alpha)$ is SU's *data rate increase* by accepting offer α. Note that the SU can always achieve a zero utility without participating in the cooperative communication. Given PU's offer α, it is *optimal* for SU to accept the offer if and only if $U_s(\alpha) > 0$.

The PU's single-slot utility $U_p(\alpha)$ is its *achievable data rate.* Without SU's relay, the PU can achieve a data rate R_{dir}. If PU's offer α is accepted by SU, then the PU's data rate is $\frac{1-\alpha}{2} R_r$, where R_r is given in (1). In each time slot, the PU aims to maximize its utility

$$U_p(\alpha) = \max\left\{R_{dir}, \frac{1-\alpha}{2} R_r\right\}. \tag{4}$$

3.2 Sequential Equilibrium

In this subsection, we consider the multi-slot bargaining game. This bargaining process is a dynamic Bayesian game [11], which includes the PU's and SU's dynamic decision-making and belief updates. The commonly used equilibrium concept for this dynamic Bayesian game is the *sequential equilibrium* (SE), which satisfies the following three requirements [12]:

Requirement 1. *The player taking the action must have a belief (probability distribution) about the incomplete information, reflecting what that player believes has happened so far.*

Requirement 2. *The action taken by a player must be optimal given the player's belief and the other players' subsequent strategies.*

Requirement 3. *A player's belief is determined by the Bayes' rule whenever it applies and the players' hypothesized equilibrium strategies.*

The belief in Requirement 1 is the PU's probability assessment q_n about the High type SU in time slot n, with an initial value $q_N = \eta$ in the first time slot (*i.e.,* time slot indexed as N). As the bargaining proceeds, both the PU and SU can observe all prior history actions, which might enable the PU to *update* its belief about the SU's type so that PU can accordingly make new decisions. The SU knows its own type and there is no incomplete information in the PU, thus the SU's belief is deterministic during the game.

3.3 Equilibrium Characterization

Generally, the equilibrium outcome depends crucially on both players' parameter settings, *i.e.,* the PU and SU will both influence the game result. On the one hand, an SU will only choose to cooperate and serve as a relay for PU if it can get a positive *total* utility in N time slots. It implies that the equilibrium outcome not only depends on the SU's energy cost (either C_h or C_l), but also on SU's achievable average data rate per unit power, R_s/P_s. On the other hand, the equilibrium outcome also relies on whether and how the PU decides to cooperate with the SU. This indicates that the equilibrium outcome is related to the relationship between PU's direct transmission rate R_{dir} and relay rate $\frac{1-\alpha}{2}R_r$.[1]

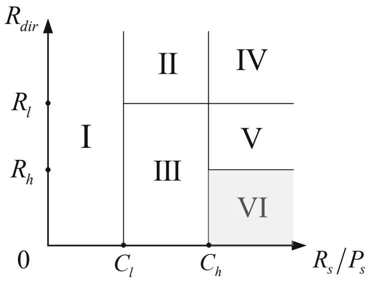

Fig. 3. Equilibrium Outcomes in Different Regions

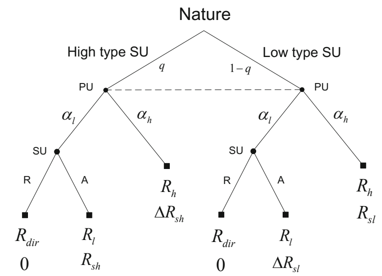

Fig. 4. Game Tree of the Single-Slot Bargaining with $R_l > R_h > R_{dir}$ and $C_l < C_h < R_s/P_s$

Next we will discuss several equilibrium contingencies based on different parameter settings of both PU and SU. Figure 3 illustrates six different cases for equilibrium discussions. Due to the page limit, we summarize the equilibria in

[1] Specifically, the relay rate has two values, *i.e.,* R_l and R_h, which will be discussed later. See [13] for details.

Regions I-V in the following theorem. We will discuss the equilibrium outcome with *reputation effect* of Region VI in next section.

Let us define

$$\alpha_h = \frac{1}{\frac{2R_s}{C_h P_s} - 1} + \varepsilon,\ \alpha_l = \frac{1}{\frac{2R_s}{C_l P_s} - 1} + \varepsilon, \tag{5}$$

where ε is an arbitrarily small positive value. Here, α_h and α_l are the PUs respective optimal offer to the High and Low type SU under the AF relay mode. For ease of discussion later on, we further define

- $R_h = \frac{1-\alpha_h}{2} R_r$: the PU's single-slot data rate if SU accepts the offer α_h.
- $R_l = \frac{1-\alpha_l}{2} R_r$: the PU's single-slot data rate if SU accepts the offer α_l.

Obviously, we have $\alpha_h > \alpha_l$ and $R_h < R_l$.

Theorem 1. *Consider a multi-slot bargaining game where the PU and SU are non-myopic players. In Regions I, II, and IV, PU always chooses direct transmission only regardless of SU's type. In Regions III and V, PU always offers α_l to SU. A High type SU rejects the offer α_l, and a Low type SU accepts the offer α_l.*

The proof of Theorem 1 is given in our online technical report [13]. In Regions I-V, we can decompose the multi-slot bargaining game into N independent single-slot bargaining game. The PU's decisions in these regions do not rely on its belief about the SU's type.

The remaining region is Region VI, where we cannot simplify the analysis of the multi-slot bargaining into N single-slot ones. We will discuss the SE of this region in next section.

4 Dynamic Bargaining with Reputation Effect

In this section, we study the SE result in Region VI. In this case, R_{dir} is small, and PU will never choose direct transmission.

4.1 Basic Analysis of the Single-Slot Bargaining Game

To attract the help from SU, PU needs to provide an offer α that makes the SU's single-slot utility $U_s(\alpha)$ *slightly* larger than zero. The PU's optimal offer to the High and Low type SU can be given in (5). For ease of discussion later on, we define two more notations,

- $\Delta R_{sh} = \left(R_s - \frac{1}{2} P_s C_h\right) \varepsilon$: the High type SU's single-slot utility if accepting α_h.
- $\Delta R_{sl} = \left(R_s - \frac{1}{2} P_s C_l\right) \varepsilon$: the Low type SU's single-slot utility if accepting α_l.

In Region VI, an SU may reject or accept the PU's offer α_h or α_l. However, no matter what happenes, PU's *expected* utility will not be worse than R_{dir} since $R_l > R_h > R_{dir}$. Therefore, PU will never choose direct transmission only without bargaining (Figure 2(a)). Thus, PU has *two* options here: offer α_h or offer α_l, with the corresponding PU's utility R_h and R_l if SU accepts the offer.

Let us first consider the game tree of the *single-slot* game in Fig. 4, where *nature* moves first and determines SU's type. PU and SU make decisions alternatively at the *non-leaf* nodes (black solid dot); each possible game result is denoted by a *leaf* node (black solid square) together with the corresponding PU utility (upper value) and SU utility (lower value). PU's belief (about SU's type) is $\Pr(C = C_h) = q$. Here, we further define

- $R_{sh} = \alpha_l R_s - \frac{1+\alpha_l}{2} P_s C_h$: the High type SU's single-slot utility if it accepts the low offer α_l.
- $R_{sl} = \alpha_h R_s - \frac{1+\alpha_h}{2} P_s C_l$: the Low type SU's single-slot utility if it accepts the high offer α_h.

When ε approaches zero in (5), we have $R_{sh} < 0$ and $R_{sl} > 0$. Thus, a High type SU will not accept a low offer, while a Low type SU has the *incentive* to accept a high offer, which leads to the reputation effect in our later analysis.

In Fig. 4, PU first decides to offer α_h or α_l. Then, the SU selects the acceptance (**A**) or rejection (**R**). If PU offers α_h, the SU will always accept regardless of its type since $\Delta R_{sh} > 0$ and $R_{sl} > 0$. Hence there is only one leaf node following the offer α_h. If PU offers α_l, a High type SU will reject as $R_{sh} < 0$, and a Low type SU will accept since $\Delta R_{sl} > 0$.

Anticipating the SU's responses, PU's *expected* utility if offering α_l is

$$U_p^{\alpha_l} = qR_{dir} + (1 - q)R_l. \tag{6}$$

PU's utility is R_h if it offers α_h. Thus, PU will offer α_l if $U_p^{\alpha_l} > R_h$, *i.e.,* $q < (R_l - R_h)/(R_l - R_{dir})$. The relationship between (6) and R_h depends on the value of q.

4.2 Sequential Equilibrium of the Multi-slot Bargaining

Now we return to the multi-slot bargaining game ($N > 1$), where the PU's belief might change over time (*i.e.,* q_n for time slot n instead of a fixed value q as in (6)) based on the game history. The SU's strategy may also change depending on the game history and its anticipation of the PU's future response. In particular, a Low type user has an incentive to reject α_l in earlier time slots even though α_l brings a positive utility for each time slot. The purpose of the Low type SU's *predation* strategy is to establish a *reputation* of a High type SU and induce the PU to offer α_h in the future, which improves the SU's total utility in N time slots. We will find the SE result based on such a prediction.

The SE of the multi-slot bargaining includes the following components: (i) the update of PU's belief q_n (*i.e.,* the probability of a High type SU) in each time slot $n = N, ..., 1$, (ii) PU's strategy (offer α_l or α_h) in each time slot n, (iii) SU's strategy (accept or reject) in each time slot n.

Theorem 2. *The sequential equilibrium of the multi-slot bargaining game with non-myopic PU and SU is given in (a) to (l), where the parameter*

$$d = \frac{R_l - R_h}{R_l - R_{dir}} \in (0, 1). \tag{7}$$

- ***PU's Belief Updates:***[2]
  (a) If $q_{n+1} = 0$*, then* $q_n = 0$*.*
  (b) If $q_{n+1} > 0$ *and SU accepts the high offer* α_h *in time slot* $n+1$*, then* $q_n = q_{n+1}$*.*
  (c) If $q_{n+1} > 0$ *and SU accepts the low offer* α_l *in time slot* $n+1$*, then* $q_n = 0$*.*
  (d) If $q_{n+1} > 0$ *and SU rejects the low offer* α_l *in time slot* $n+1$*, then* $q_n = max(d^n, q_{n+1})$*.*
- ***PU's Strategy:***
  (e) If $q_n < d^n$ *in time slot* n*, offers* α_l*.*
  (f) If $q_n > d^n$ *in time slot* n*, offers* α_h*.*
  (g) If $q_n = d^n$ *in time slot* n*, offers* α_h *with probability* $\frac{\Delta R_{sl}}{R_{sl} - \Delta R_{sl}}$ *and offers* α_l *with probability* $1 - \frac{\Delta R_{sl}}{R_{sl} - \Delta R_{sl}}$*.*[3]
- ***The High type SU's Strategy:***
  (h) Always accepts α_h *and rejects* α_l*.*
- ***The Low type SU's Strategy:***
  (i) Always accepts α_h*.*
  (j) If $n = 1$ *(the last time slot), accepts* α_l*.*
  (k) If $n > 1$ *and* $q_n \geq d^{n-1}$*, rejects* α_l*.*
  (l) If $n > 1$ *and* $q_n < d^{n-1}$*, rejects* α_l *with probability* $y_n = \frac{(1-d^{n-1})q_n}{d^{n-1}(1-q_n)}$ *and accepts* α_l *with* $1 - y_n$*.*

Proof. First, let us look at the High type SU's strategy. In the multi-slot bargaining game, we can show that there is no incentive for the High type SU to accept α_l. This is because accepting α_l leads to negative SU's utility in the current time slot, and will make PU believe that the SU is a Low type. This means that all future offers will be α_l, and thus the SU's total utility will be negative.

Next, we verify the PU's strategy and its belief update scheme. Let us first discuss PU's *limiting belief* q_n^*, which can be interpreted as a decision threshold to determine which offer (α_l or α_h) the PU should provide. It can also be viewed as the SU's *limiting threshold reputation* as the High type, above which the PU will not offer α_l.[4] We should consider the PU's utility by summing its utilities from the current time slot n to the last time slot. Define $U_{\mathbb{P}_n}(q_n)$ to be the expected utility of PU in time slot n with the belief q_n.[5] Furthermore, we use

[2] Recall that we index time backwards, and thus we will compute q_n based on q_{n+1} as time slot n is after time slot $n+1$. See Fig. 2.

[3] As ε is an arbitrary small positive, ΔR_{sl} is arbitrarily small. Therefore, the assumption $R_{sl} > 2\Delta R_{sl}$ holds.

[4] For the standard definition of "reputation", see Section 5.

[5] For clear illustrations, we interchangeably use different notations, *i.e.,* $\mathbb{P}_1$, $\mathbb{P}_2$, ..., $\mathbb{P}_N$ to mark the PU in different time slots. However, these notations all indicate the *unique* PU in the multi-slot bargaining game.

$U_{\mathbb{P}_n}(q_n|\alpha_l)$ and $U_{\mathbb{P}_n}(q_n|\alpha_h)$ to denote PU's expected utilities when offering α_l and α_h, respectively.

It is easy to analyze the last time slot ($n = 1$) since there is no other time slot following it. PU in the last time slot is indifferent between offering α_l and α_h if

$$q_1 R_{dir} + y_1(1 - q_1)R_{dir} + (1 - y_1)(1 - q_1)R_l = R_h. \tag{8}$$

The *LHS* of (8) is PU's utility if offering α_l, and the *RHS* is the utility if offering α_h. y_n denotes the probability that the Low type SU rejects α_l in time slot n. Note that for the Low type SU, accepting α_l is optimal in the single-slot game (see Fig. 4). Obviously, such a strategy also applies in the last time slot. Thus, if PU offers α_l in the last time slot indexed by $n = 1$, the Low type SU will accept and hence $y_1 = 0$ holds. Given that $y_1 = 0$, we get the limiting belief q_1^* for the case of $n = 1$,

$$q_1^* = \frac{R_l - R_h}{R_l - R_{dir}} \triangleq d. \tag{9}$$

The utility $U_{\mathbb{P}_1}(q_1|\alpha_l) = q_1 R_{dir} + (1 - q_1)R_l$ and $U_{\mathbb{P}_1}(q_1|\alpha_h) = R_h$. Further, if the actual belief $q_1 > q_1^*$, then $U_{\mathbb{P}_1}(q_1|\alpha_l) < U_{\mathbb{P}_1}(q_1|\alpha_h)$ and PU will choose to offer α_h. Otherwise PU will offer α_l. Therefore, we have $U_{\mathbb{P}_1}(q_1)$,

$$U_{\mathbb{P}_1}(q_1) = \begin{cases} q_1 R_{dir} + (1 - q_1)R_l, & \text{if } q_1 \le q_1^* = d, \\ R_h, & \text{if } q_1 > q_1^* = d. \end{cases} \tag{10}$$

Now consider the second to last time slot indexed by $n = 2$. Recall Requirement 3, which states that the belief should be consistent with strategies and satisfies the Bayes' rule whenever it applies. Such a requirement also holds when the actual belief is the limiting belief. Thus, q_1^* should satisfy the Bayes' rule and should be derived from q_2 and y_2, *i.e.,*

$$q_1^* = \text{Prob(SU is High type} \mid \text{SU rejects } \alpha_l). \tag{11}$$

Note that the Bayesian process only applies if PU provides α_l and SU rejects it. Recall that accepting α_l is an evidence of indicating that the SU belongs to the Low type. Therefore, SU's accepting α_l in time slot $n = 2$ will result in $q_1 = 0$, which contradicts the fact $q_1^* = d$. If PU provides α_h and SU accepts it, from (b) we get $q_2 = q_1^* = d$ since there is no information provided to help update PU's belief. By the Bayes' rule, we can proceed the derivation with (11) to obtain y_2 as

$$y_2 = \begin{cases} \frac{q_2(1-d)}{(1-q_2)d}, & \text{if } q_2 \le d, \\ 1, & \text{if } q_2 > d. \end{cases} \tag{12}$$

When $q_2 \le d$, $\mathbb{P}_2$'s expected utility $U_{\mathbb{P}_2}(q_2|\alpha_l)$ can be expressed as

$$q_2(R_{dir} + U_{\mathbb{P}_1}(d)) + (1 - q_2)y_2(R_{dir} + U_{\mathbb{P}_1}(d)) + (1 - q_2)(1 - y_2)(R_l + U_{\mathbb{P}_1}(0)), \tag{13}$$

if it offers α_l. When α_l is rejected by SU in the second to last period, we apply the Bayes' rule to get

$$q_1 = \frac{q_2}{q_2 + (1 - q_2) \times \left(\frac{(1-d)q_2}{d(1-q_2)}\right)} = d.$$

Therefore, the updated belief in the last time slot will be $q_1 = d$. If SU accepts α_l, then PU will update its belief as 0. Thus, we have the belief update results in (13). Substitute (12) into (13), then we have

$$U_{\mathbb{P}_2}(q_2|\alpha_l) = \frac{q_2}{d}\left[R_{dir} + U_{\mathbb{P}_1}(d)\right] + \left(1 - \frac{q_2}{d}\right)\left[R_l + U_{\mathbb{P}_1}(0)\right]. \tag{14}$$

When $q_2 > d$, $U_{\mathbb{P}_2}(q_2|\alpha_l)$ will be

$$U_{\mathbb{P}_2}(q_2|\alpha_l) = R_{dir} + U_{\mathbb{P}_1}(q_2). \tag{15}$$

Note that the updated belief in the last time slot keeps unchanged in (15). With the Bayes' rule, we have

$$q_1 = \frac{q_2}{q_2 + (1 - q_2) \times \text{Prob(SU rejects } \alpha_l \mid \text{Low type SU)}} = q_2. \tag{16}$$

Hence, we obtain the result in (15). The expected utility of PU when offering α_h is

$$U_{\mathbb{P}_2}(q_2|\alpha_h) = R_h + U_{\mathbb{P}_1}(q_2). \tag{17}$$

If PU's belief is the limiting belief, then the balancing condition must hold, *i.e.,*

$$U_{\mathbb{P}_2}(q_2|\alpha_l) = U_{\mathbb{P}_2}(q_2|\alpha_h). \tag{18}$$

However, from (15) and (17) we can observe that the limiting belief q_2^* cannot be larger than d. Recall the expression of $U_{\mathbb{P}_1}(q_1)$ in (10), we have $U_{\mathbb{P}_1}(d) = R_h = dR_{dir} + (1-d)R_l$ and $U_{\mathbb{P}_1}(0) = R_l$. Let (14) equal to (17), and we have

$$q_2^* = d \times \frac{R_l - R_h}{R_l - R_{dir}} = d^2. \tag{19}$$

Furthermore, it is easy to show that $U_{\mathbb{P}_2}(q_2|\alpha_h) > U_{\mathbb{P}_2}(q_2|\alpha_l)$ when $q_2 > q_2^*$. We thus have PU's expected utility in the second to last time slot as

$$U_{\mathbb{P}_2}(q_2) = \begin{cases} \frac{q_2}{d}\left[R_{dir} + U_{\mathbb{P}_1}(d)\right] + \left(1 - \frac{q_2}{d}\right)\left[R_l + U_{\mathbb{P}_1}(0)\right], & \text{if } q_2 \le q_2^* = d^2, \\ R_h + U_{\mathbb{P}_1}(q_2), & \text{if } q_2 > q_2^* = d^2. \end{cases} \tag{20}$$

For $n \ge 3$, we first conjecture that the limiting belief in time slot n is

$$q_n^* = d^n. \tag{21}$$

Given the above assumption, we have PU's expected utility in time slot n as

$$U_{\mathbb{P}_n}(q_n) = \begin{cases} \frac{q_n}{d^{n-1}}\left[R_{dir} + U_{\mathbb{P}_{n-1}}(d^{n-1})\right] + \left(1 - \frac{q_n}{d^{n-1}}\right)\left[R_l + U_{\mathbb{P}_{n-1}}(0)\right], & \text{if } q_n \le q_n^* = d^n, \\ R_h + U_{\mathbb{P}_{n-1}}(q_n), & \text{if } q_n > q_n^* = d^n. \end{cases} \tag{22}$$

where we define the *initial* conditions as $U_{\mathbb{P}_0}(0) = U_{\mathbb{P}_0}(1) = 0$. Obviously, (21) and (22) hold for the cases $n = 1$ and $n = 2$. Again, we apply the Bayes' rule to $q_{n-1}^* = d^{n-1}$ and get the restriction condition between q_n and y_n,

$$q_{n-1}^* = d^{n-1} = \frac{q_n}{q_n + y_n(1 - q_n)}. \tag{23}$$

From (23), we get

$$y_n = \frac{q_n(1-d^{n-1})}{(1-q_n)d^{n-1}}. \tag{24}$$

Besides, we get the *balancing strategy* y_n^* in time slot n by substituting $q_n^* = d^n$ into (24),

$$y_n^* = y_n(q_n^*) = \frac{d^n(1-d^{n-1})}{(1-d^n)d^{n-1}} = \frac{d-d^n}{1-d^n}. \tag{25}$$

For the case of time slot $n+1$, the Bayes' rule still applies to the limiting belief q_n^*, and thus we have

$$y_{n+1} = \begin{cases} \frac{q_{n+1}(1-d^n)}{(1-q_{n+1})d^n}, & \text{if } q_{n+1} \le d^n, \\ 1, & \text{if } q_{n+1} > d^n. \end{cases} \tag{26}$$

By considering (26), PU's expected utility when offering α_l in time slot $n+1$ is given

$$U_{\mathbb{P}_{n+1}}(q_{n+1}|\alpha_l) = \begin{cases} \frac{q_{n+1}}{d^n}\left[R_{dir} + U_{\mathbb{P}_n}(d^n)\right] + \left(1 - \frac{q_{n+1}}{d^n}\right)\left[R_l + U_{\mathbb{P}_n}(0)\right], & \text{if } q_{n+1} \le d^n, \\ R_{dir} + U_{\mathbb{P}_n}(q_{n+1}), & \text{if } q_{n+1} > d^n. \end{cases} \tag{27}$$

PU's expected utility when offering α_h will be

$$U_{\mathbb{P}_{n+1}}(q_{n+1}|\alpha_h) = R_h + U_{\mathbb{P}_n}(q_{n+1}). \tag{28}$$

Similarly, from (27) and (28) we can see that if the balancing condition holds, then $q_{n+1} \le d^n$ must be satisfied. According to the balancing condition, we have

$$\frac{q_{n+1}}{d^n}\left[R_{dir} + U_{\mathbb{P}_n}(d^n)\right] + \left(1 - \frac{q_{n+1}}{d^n}\right)\left[R_l + U_{\mathbb{P}_n}(0)\right] = R_h + U_{\mathbb{P}_n}(q_{n+1}). \tag{29}$$

where

$$\begin{cases} U_{\mathbb{P}_n}(d^n) = U_{\mathbb{P}_{n-1}}(d^n) + R_h, \\ U_{\mathbb{P}_n}(0) = R_l + U_{\mathbb{P}_{n-1}}(0), \\ U_{\mathbb{P}_n}(q_{n+1}) = \frac{q_{n+1}}{d^{n-1}}\left[R_{dir} + U_{\mathbb{P}_{n-1}}(d^{n-1})\right] + \left(1 - \frac{q_{n+1}}{d^{n-1}}\right)\left[R_l + U_{\mathbb{P}_{n-1}}(0)\right]. \end{cases} \tag{30}$$

Substitute (30) into (29), and we can simplify the equation to get

$$q_{n+1}^* = d^{n+1}. \tag{31}$$

Thus, the limiting belief in time slot $n+1$ conforms to our prior assumption in (21). Besides, it is easy to see that PU will offer α_l if $q_{n+1} < q_{n+1}^* = d^{n+1}$ and offer α_h if $q_{n+1} > q_{n+1}^* = d^{n+1}$.

As of now, we have solved the PU's limiting belief and its corresponding strategy when the actual belief is not the limiting belief in each time slot. At the same time, we also verify the Low type SU's strategy (y_n) and the High type SU's strategy. Since $\Delta R_{sh} > 0$ and $R_{sl} > 0$, SU always accepts α_h regardless of its type.

Next let us verify the PU's belief update scheme. If there is an offer α_h in time slot $n+1$, an SU will definitely accept it. This does not help to update the PU's belief about the SU's type. Thus, we have $q_n = q_{n+1}$ as shown in (b). Then we will verify (a) and (c) from two aspects:

- If there is an offer α_l in time slot $n+1$ and the SU accepts it, then PU immediately knows that the SU is a Low type. Therefore, PU will update its belief as $q_n = 0$.
- If $q_{n+1} = 0$, then PU knows that the SU is a Low type for time slot $n+1$ and the whole later time slots. Therefore, PU will always set its belief as zero.

From the above discussions, we complete the proof of the PU's belief update scheme in (a) and (c).

When $q_{n+1} > 0$ and SU rejects the offer α_l in time slot $n+1$, (with the Bayes' rule) q_n is determined by

$$\frac{q_{n+1} \times 1}{q_{n+1} + (1 - q_{n+1}) \times \text{Prob(SU rejects } \alpha_l \mid \text{Low type SU)}}$$

According to the Low type SU's strategy, it could be further divided into two cases:

- When $q_{n+1} \geq d^n$ in time slot $n+1$, the Low type SU will always reject α_l. Thus, we have

$$q_n = \frac{q_{n+1} \times 1}{q_{n+1} + (1 - q_{n+1}) \times 1} = q_{n+1}. \tag{32}$$

- When $q_{n+1} < d^n$ in time slot $n+1$, the Low type SU will reject α_l with probability $\frac{(1-d^n)q_{n+1}}{d^n(1-q_{n+1})}$. We have

$$q_n = \frac{q_{n+1}}{q_{n+1} + (1 - q_{n+1}) \times \left(\frac{(1-d^n)q_{n+1}}{d^n(1-q_{n+1})}\right)} = d^n. \tag{33}$$

From (32) and (33), we have

$$q_n = \max(d^n, q_{n+1}) = \max\left(q_n^*, q_{n+1}\right). \tag{34}$$

Thus, we complete the verification of (d).

Finally, let us consider PU's *mixed* strategy in (g). Note that the utility in (22), denoted as $U_{\mathbb{P}_n}(q_n^*)$ when PU's actual belief is equal to the limiting belief, is constant if the *balancing strategy* y_n^* applies. It is also subject to the limiting belief q_n^*, and independent of whether $\mathbb{P}_n$ chooses α_l or α_h, or randomizes on both alternatives. Thus we have

$$U_{\mathbb{P}_n}\left(x_n,\ y_n^*/q_n^*\right) = \textsf{constant},\ \text{for all } x_n \text{ in } X_n, \tag{35}$$

where X_n is the set of all possible $\mathbb{P}_n$'s strategies in time slot n. The mixed strategy $[x_n, 1-x_n]$ denotes PU's choosing α_h with probability x_n and choosing

α_l with probability $1 - x_n$. When $q_n = q_n^*$, from (22) we can see that the utility on the limiting belief is $R_h + U_{\mathbb{P}_{n-1}}(d^n)$, which can be *iteratively* calculated by considering PU's expected utility in time slot $n-1$, $n-2$,..., and 1. Since PU's utility in each time slot is fixed, $R_h + U_{\mathbb{P}_{n-1}}(d^n)$ is a constant for a given n. Thus it implies that all x_n in X_n are the *best response* to y_n^*.

However, when $q_n = q_n^*$ the Low type SU will only be expected to choose *the balancing strategy* y_n^*, which is the requirement of the Bayesian consistency (Requirement 3), thus y_n^*, as an equilibrium strategy, *should* maximize the Low type SU's total utility. If we choose $[x^*, 1-x^*]$ from the set X_n as $\mathbb{P}_n$'s optimal mixed strategy, then it implies that the balancing strategy y_n^* is the *best response* to x^* for the corresponding beliefs.

To solve it, we first *conjecture* that there exists a strategy for the PU in each time slot when the actual belief is equal to the limiting belief, given as follows,

$$\mathcal{X} = (x_N, ..., x_n, ..., x_1), \; x_N = \cdots = x_1 = x^* > 0 \tag{36}$$

where x_n is the PU's mixed strategy, *i.e.,* the probability of offering α_h in time slot n when $q_n = q_n^*$ and making the Low type SU *indifferent* between rejecting and accepting α_l. Since accepting α_l will result in the utility $\Delta R_{sl} > 0$, the strategy profile $\mathcal{X}$ implies the Low type SU's utility is $n\Delta R_{sl}$, irrespective of what strategy the Low type SU will choose. Therefore, any strategy will maximize the Low type SU's utility, including the equilibrium strategy y_n^*. It means that y_n^* is the best response to x^*. If we can verify that the strategy in (36) does exist and further obtain the value of x^*, then we solve x^* in PU's strategy (g). In fact, we get $x^* = \frac{\Delta R_{sl}}{R_{sl} - \Delta R_{sl}}$. Due to the page limit, we omit the proof here. See [13] for details. Therefore, we completed the proof of Theorem 2. ■

5 Reputation Effect Analysis

In noncooperative game theory, a player's "reputation" is its opponents' current beliefs about its type [11]. In the multi-slot bargaining game, the Low type SU's reputation can be viewed as the PU's belief about the SU's type in time slot n, *i.e.,* Pr(High type)=q_n. The "reputation effect" refers to the fact that a *non-myopic* Low type SU has incentive to reject α_l to sustain a reputation of the High type so as to get higher utility in the future (see (k), (l)). This is the most interesting part of our model, and we will discuss the intuitions behind it in more details and compare it to the complete information benchmark.

Intuitively, such incentive of doing so becomes higher when the bargaining process lasts longer, which means that the reputation effect is more likely to happen in long term relationships than in short ones [11]. Therefore, it is more interesting to discuss such an effect when N is sufficiently large.

Since $d \in (0, 1)$, d^N can be arbitrarily small when N is large enough, and thus condition $q_N = \eta > d^{N-1}$ can be easily satisfied, even if the initial belief η (in the first time slot N) is small. From (k), the Low type SU will reject α_l in the initial time slot. Anticipating this, the non-myopic PU will offer α_h according to (f). Thus, a Low type SU gets the high utility R_{sl} in the first time slot $n = N$.

Interestingly, such a set of strategies will last in the early several time slots. As the bargaining progresses and slot index n deceases, parameter d^n increases. If PU's belief q_n cannot increase accordingly and becomes 0 (*e.g.,* as in (c)), then PU begins to offer α_l. It indicates that the Low type SU's benefits from the reputation effect ends. To clarify the discussions, we define

$$k(\eta) = \inf\left(n : d^n < q_N = \eta\right), \tag{37}$$

Obviously, PU will offer α_h from time slot N to $k(\eta)$ according to (b), (f), (h), and (i). In time slot $k(\eta) - 1$, PU begins to offer α_l based on (e). Facing α_l, the Low type SU might reject or accept it based on (l). Since only the Low type SU might accept α_l based on (h), acceptance of α_l will reveal the Low type SU's true type to PU, after which the PU will offer α_l in each time slot till the game ends according to (a), (c), and (e).

For the purpose of comparison, let us consider the case with complete information, where the PU knows the SU's true type from the very beginning. To begin with, let us discuss the single-slot bargaining with complete information in SU's energy cost. There are two different optimal strategy profiles according to the SU's type. For the PU and Low type SU, if the PU offers α_l, the Low type SU chooses between utility 0 if it rejects and $\Delta R_{sl} > 0$ if it accepts, so definitely it will accept PU's offer. Anticipating this response, the PU chooses between R_h if it provides α_h (SU will accept it as we discussed above) and R_l if it provides α_l, and so it will offer α_l. This is the *unique* Nash equilibrium for this single-slot bargaining game with complete information.

Figure 5 (a) indicates the equilibrium discussed above. We mark the PU's and Low type SU's equilibrium decisions by the black bold line on the possible decision paths. Similarly, we can obtain the equilibrium in the case of PU and High type SU as shown in Figure 5 (b).

Next, let us consider the case that the two different single-slot games in Figure 5 are separately played a *finite* number times. The PU plays with the SU (High type or Low type) during a succession of N time slots. Take the PU and Low type SU game as an example. The analysis for the PU and High type SU case follows similarly and is omitted to due space constraint.

This is a *finitely-repeated* game with *complete* information. It allows both PU and Low type SU to perfectly observe all moves in earlier time slots. In the last time slot ($n = 1$), the SU will not reject PU's offer α_l since there are no later chances for PU to provide offers. Anticipating this, PU will definitely provide α_l in the last time slot. Backwardly, in the second to last time slot the SU will have no incentive to reject α_l because doing so is costly in the short run and will have no effect on the decisions in the last time slot. Realizing this, the PU will offer α_l to get a higher utility R_l for itself. This logic can be repeated till the first time slot ($n = N$): *in each time slot, PU offers α_l and SU always accepts.* Moreover, this is the unique *subgame perfect Nash equilibrium* (SPE) of the game.[6] We have

[6] SPE is a commonly used solution concept for dynamic game with complete information, which can effectively eliminate the *incredible threat* from the opponent. See [11] for more details.

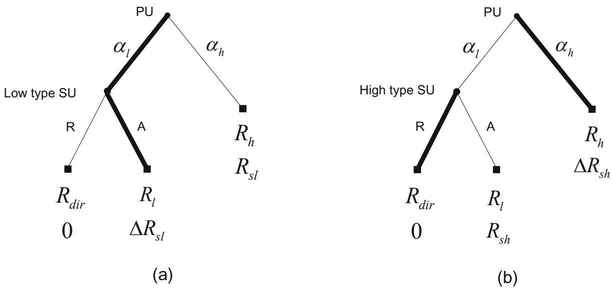

Fig. 5. Game Tree of the Single-Slot Bargaining with Complete Information: (a) PU and the Low type SU, (b) PU and the High type SU

the similar conclusion for the case of PU and High type SU, *i.e.,* PU's offering α_h and SU's rejection of α_l always occur in each time slot.

Based on the above analysis, we can see that the reputation effect does not emerge in a repeated game with complete information. The sequential equilibrium of the multi-slot (finite-repeated) bargaining game in Theorem 2 deviates from the equilibrium outcome discussed above because of *incomplete information* in SU's energy cost. In the dynamic Bayesian bargaining game, the Low type SU might convince PU that it is the High type by rejecting the offer α_l from PU and thus obtains a higher utility. This means that information incompleteness in SU's energy cost results in the reputation effect, with which the Low type SU can benefit more from the cooperative communication.

6 Conclusion

In this paper, we investigate a cooperative spectrum sharing mechanism achieved by a dynamic Bayesian spectrum bargaining between one PU and one SU. We model the bargaining as a dynamic Bayesian game, and characterize all possible equilibria under different system parameter settings. In particular, we focus on characterizing the sequential equilibrium, where the reputation effect brings higher utilities for the Low type SU. Analysis shows that the Low type SU could exploit the PUs lack of information for its own benefits, which will not happen with the complete information scenario.

Acknowledgement. This work has been supported by CSC Postgraduate Scholarship, NSF of China No.60832008, National S&T Major Project 2010ZX03003-002-02, TNList Cross-discipline Foundation, China's 973 Project 2007CB310608, PCSIRT, and National S&T Pillar Program 2008BAH30B09. This work is also supported by the General Research Funds (Project No. 412509 and 412710) established under the University Grant Committee of the Hong Kong Special Administrative Region, China.

References

1. Laneman, J., Tse, D., Wornell, G.: Cooperative Diversity in Wireless Networks: Efficient Protocols and Outrage Behavior. IEEE Transaction on Information Theory 50(12), 3062–3080 (2004)
2. Cover, T.M., El Gamaal, A.: Capacity Theorems for the relay channel. IEEE Transaction on Information Theory 25(5), 572–584 (1979)
3. Nosratinia, A., Hunter, T.E., Hedayat, A.: Cooperative communication in wireless networks. IEEE Communications Magazine 42, 74–80 (2004)
4. Genc, V., Murphy, S., Yang, Y., Murphy, J.: IEEE 802.16J Relay-Based Wireless Access Networks: An Overview. IEEE Wireless Communications Magazine 15, 56–63 (2008)
5. Wang, H., Gao, L., Gan, X., Wang, X., Hossain, E.: Cooperative Spectrum Sharing in Cognitive Radio Networks: A Game-Theoretic Approach. In: IEEE ICC, South Africa (2010)
6. Simeone, O., Stanojev, I., Savazzi, S., Bar-Ness, Y., Spagnolini, U., Pickholtz, R.: Spectrum leasing to cooperating secondary ad hoc networks. IEEE Journal on Selected Areas in Communications 26(1), 203–213 (2008)
7. Zhang, J., Zhang, Q.: Stackelberg Game for Utility-Based Cooperative Cognitive Radio Networks. In: ACM MobiHoc, New Orleans, USA (May 2009)
8. Duan, L., Gao, L., Huang, J.: Contract-Based Cooperative Spectrum Sharing. In: IEEE DySPAN, Aachen, Germany (May 2011)
9. Yan, Y., Huang, J., Zhong, X., Wang, J.: Dynamic Spectrum Negotiation with Asymmetric Information. In: GameNets, Shanghai, China (April 2011)
10. Yan, Y., Huang, J., Zhong, X., Zhao, M., Wang, J.: Sequential Bargaining in Cooperative Spectrum Sharing: Incomplete Information with Reputation Effect. Submitted to Globecom (2011)
11. Fudenberg, D., Tirole, J.: Game Theory. The MIT Press, Cambridge (1991)
12. Kreps, D.M., Wilson, R.: Sequential equilibria. Econometrica 50, 863–894 (1982)
13. Yan, Y., Huang, J., Zhong, X., Zhao, M., Wang, J.: Sequential Bargaining in Cooperative Spectrum Sharing: Incomplete Information with Reputation Effect. Technical Report, `http://home.ie.cuhk.edu.hk/jwhuang/publication/BargainReputationTechReport.pdf`
14. Kreps, D.M., Wilson, R.: Reputation and imperfect information. Journal of Economic Theory 27, 253–279 (1982)
15. Holler, M.J.: The Kreps-Wilson monopoly-entrant game and cautiously rationalizable sequential equilibria. Quality & Quantity 25, 69–83 (1991)