A cross-layer design approach on spectrum allocation and resource scheduling in cognitive PMP networks[*]

Jie CHEN[†], Min-jian ZHAO[†‡], Qiao ZHOU, Shi-ju LI

(*Department of Information Science and Electronic Engineering, Zhejiang University, Hangzhou 310027, China*)

[†]E-mail: chenjie0422@yahoo.com.cn; mjzhao@zju.edu.cn

Abstract: We propose the spectrum allocation and resource scheduling algorithms in cognitive point to multipoint (PMP) networks with rapid changes of spectrum opportunities and present a media access control (MAC) protocol based on these algorithms. The objective of spectrum allocation is to make efficient use of the spectrum while maintaining the transceiver synchronization on frequency and time in the network. The objective of resource scheduling is to guarantee the quality of service (QoS) requirements of different kinds of connections and to minimize the total energy consumption in the network as well. By sensing only a small set of possible channels in each slot based on the state transition probability of each channel, our spectrum allocation algorithm achieves high spectrum efficiency in the network. The resource scheduling problem is divided into three sub problems and we derive optimal solutions to these problems by greedy algorithm and convex optimization. The simulation results show that our algorithm can make efficient use of the spectrum and the network resources at a cost of low computational complexity.

Key words: Cognitive point to multipoint (PMP) networks, Markov process, Cross-layer design, Convex optimization
doi:10.1631/jzus.A071373   **Document code:** A   **CLC number:** TN919; TN92

INTRODUCTION

Wireless spectrum is a valuable natural resource. The allocation of spectrum is managed by the government up to now. Different wireless applications are usually allocated with different spectrums. However, with the rapid growth of wireless services, almost all usable frequencies have been occupied. Meanwhile, FCC (2002) identifies that a large portion of the licensed spectrums is unused at any given time and any location. Cognitive radio offers a possible solution to increase the spectrum efficiency (Mitola and Maguire, 1999). In cognitive networks, the cognitive user is capable of sensing the surrounding environment and making corresponding changes in certain operating parameters (e.g., carrier frequency, transmission power, modulation scheme) in real time (Haykin,

2005). And the cognitive users should not interfere with the transmission of primary (licensed) users in the same frequency band. There are two kinds of researches on avoiding interference between the cognitive users and the primary users: spectrum underlay and spectrum overlay.

The researches of spectrum underlay limit the transmission power of the cognitive users and make sure that the interference perceived by the primary users does not exceed certain threshold. So the cognitive users and primary users may transmit simultaneously in the same frequency band at the same location. Spreading transmission over an ultra wide frequency band is a good solution in this context as the transmission power can be extremely low (Lansford, 2004). The cognitive users can also use dirty-paper coding to avoid interference with the knowledge of the primary users' code information (Jovicic and Viswanath, 2006). With the assumption that the primary users will always occupy the spectrum, these approaches can sufficiently increase the spectrum

efficiency. However, each cognitive user should be aware of the interference with the primary user caused by all other cognitive users, and requires some kind of communications between the cognitive users and the primary users.

The researches of spectrum overlay divide the spectrum into several channels, seek and transmit through the channels which have not been occupied by the primary users. Spectrum sensing algorithm is the basis in this context. Cabric *et al.*(2004) introduced the energy detection and the cyclostationary feature detection algorithms. The relationship between the sample number and the signal to noise ratio (SNR) of the channel is also shown. When the spectrum opportunities (idle channels) remain stable or vary slowly (e.g., the TV band), the cognitive users can sense all the channels and allocate the spectrum based on the sensing results of the full spectrum. The allocated results can last a long time. Wang and Liu (2005) proposed a list-coloring model to maximize the spectrum efficiency. Hoang and Liang (2006) also took the power control into account and allocated the spectrum based on the graph-coloring model. Won *et al.*(2006) proposed an adaptive orthogonal frequency division multiple access (OFDMA) platform and allocated the idle sub channels for all the cognitive users. However, when the spectrum opportunities vary quickly (e.g., the 2.4 GHz unlicensed band), the cognitive user may not have enough time to sense all the channels before the spectrum opportunities change again. The synchronization between the transmitter and the receiver also becomes a big problem as the transmitter may change the working band frequently. Zhao *et al.*(2007) proposed a spectrum allocation algorithm in ad-hoc networks based on partially observable Markov decision process (POMDP). This algorithm has good performance and ensures transceiver synchronization.

In cognitive networks, the resources (e.g., transmission power, time) should also be properly scheduled to achieve the optimal network performance. A lot of researches provide scheduling algorithms based on cross-layer design method. Xiao *et al.*(2004) formulated the resource allocation problem as a convex optimization problem and solved it by Lagrangian duality. Johansson and Xiao (2006) provided a solution by a nonlinear column generation algorithm, which simplifies the calculation. However,

these researches do not consider the requirements from the application layer and do not support QoS. And each link can only support one connection in those researches.

In this paper, we consider a cognitive point to multipoint (PMP) wireless network in the environment with rapid changes of spectrum opportunities. A PMP network is a network with a base station (BS) at the center and several subscriber stations (SS) around the BS. The SS can only communicate with the BS. Both the BS and the SS are cognitive users. They should sense the spectrum and seek for the transmission opportunities. We propose a protocol to maintain transceiver synchronization among the cognitive users in the PMP networks, and make efficient use of the spectrum as well as the energy of the users. The proposed protocol can also provide QoS guarantee for all the connections in the network and ensure the fairness among different kinds of connections.

Inspired by Zhao *et al.*(2007), we first introduce the algorithm into the cognitive PMP networks to allocate the spectrum. Then we formulate an admission control problem, a rate scheduling problem and a network resource scheduling problem to schedule the network resources for each connection in the network simultaneously. We also derive efficient solutions to these problems. Finally we propose the complete MAC protocol for the cognitive PMP networks based on our spectrum allocation algorithm and resource scheduling algorithm. And the simulation results of the protocol will also be shown.

SPECTRUM ALLOCATION ALGORITHM

In this section, we first introduce the spectrum allocation algorithm which will be used in the spectrum allocation module of our MAC protocol in Section 4.

When the spectrum opportunities vary quickly, the cognitive users should frequently update the spectrum status to avoid interference with the primary users. Continuous full spectrum sensing method requires at least two antennas (one for sensing, another for transmitting) and is energy inefficient. On the other hand, if we can get the state transition probabilities of each channel beforehand, we can use this information to predict the availability of each channel.

And we shall only sense a small part of the most possible idle channels. This will reduce the hardware demands and energy cost of each cognitive user. Zhao *et al.*(2007) constructed a model based on POMDP and gave a suboptimal strategy for spectrum allocation in ad-hoc networks. Inspired by this, we introduce a model to PMP networks.

Consider N possible channels, each with bandwidth B_i (i=1, …, N). We divide the time into slots with equal length. The channel status in slot n is given by $S_1(n)$, …, $S_N(n)$, where $S_i(n)$=1 means channel i is idle and $S_i(n)$=0 otherwise. We assume that the channel status remains unchanged in each slot. For each channel, we model its status as a two-state discrete Markov process. Channel i transfers from busy state to idle state with probability α_i and stays in idle state with probability β_i. The availability of the channels at the beginning of slot n is $\boldsymbol{\Omega}$=[ω_1, …, ω_N], where ω_i is the probability that channel i is idle. The sensing result for channel i is $\Theta_i \in \{0, 1\}$. Given the knowledge of $\boldsymbol{\Omega}(n)$ at the beginning of slot n, the expected reward to be gained in slot n if channel i is chosen is $rw_i = \{\omega_i(n)\beta_i + [1-\omega_i(n)]\alpha_i\}B_i(1-\varepsilon)$, where ε is the false alarm probability of the spectrum sensing module.

In a PMP network with one BS and M SS, we divide each time slot into four parts as shown in Fig.1. At the sensing stage, each user in the network calculates the expected rewards for all the channels and chooses a set of channels (A_1) to sense by the greedy algorithm which follows:

$$rw_{i \in A_1} \geq rw_{j \in \bar{A}_1}, \quad \forall i, j. \qquad (1)$$
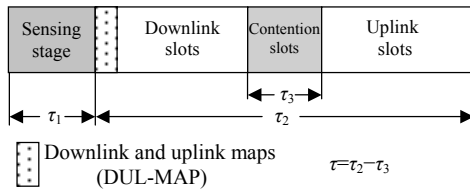

Fig.1 Time slot structure and the length of time allocated for the four parts

The number of sensed channels $|A_1|$ is related to the length of the sensing stage τ_1 and the sensing time spent on each channel which is related to the false

alarm probability ε. When the cognitive users get the sensing results Θ_{A_1}, all the users select the idle channel with the largest expected reward as the communication channel in slot n.

After the sensing stage, all the users enter the transmission stage which can be divided into downlink slots, contention slots and uplink slots. In the downlink slots, the BS transmits its data packets to the SS. An SS can register with the BS or raise a new connection request in the contention slots. And the SS can transmit to BS in the uplink slots. The transmission time of each connection is scheduled by the algorithm in the next section and the scheduling results are presented in the downlink and uplink maps (DUL_MAP) at the beginning of the downlink slots as shown in Fig.1. Let $K_i \in \{0, 1\}$ indicate whether the cognitive user is in the communication channel of the network in slot n. For the BS, if it receives the data packets from the SS in channel i in uplink slots, then K_i=1. For the SS, if they receive data packets from the BS, then K_i=1.

At the end of each slot n, all the users update the availability of each channel $\boldsymbol{\Omega}(n+1)$ by Eq.(2) with the knowledge of $\boldsymbol{\Omega}(n)$, A_1, Θ_{A_1} and K_{A_1}.

$$\boldsymbol{\Omega}(n+1) = [\omega_1(n+1), \cdots, \omega_N(n+1)], \qquad (2)$$

where

$$\omega_i(n+1) = \Pr\left[S_i(t)=1 \middle| \omega_i(n), A_1, \Theta_{A_1}, K_{A_1}\right] =$$
$$\begin{cases} 0, & \text{if } i \in A_1, \Theta_i = 0, \\ 1, & \text{if } i \in A_1, \Theta_i = 1, K_i = 1, \\ \varepsilon\{\omega_i(t)\beta_i + [1-\omega_i(t)]\alpha_i\} \big/ \{\varepsilon\{\omega_i(t)\beta_i + [1-\omega_i(t)]\alpha_i\} \\ \quad + \{\omega_i(t)(1-\beta_i) + [1-\omega_i(t)](1-\alpha_i)\}\}, \\ \quad \text{if } i \in A_1, \Theta_i = 1, K_i = 0, \\ \omega_i(t)\beta_i + [1-\omega_i(t)]\alpha_i, & \text{if } i \notin A_1. \end{cases}$$

We can see that with the same initial vector $\boldsymbol{\Omega}$ and the same transmission probabilities α=\{α_1, …, α_N\}, β=\{β_1, …, β_N\}, all the users in the network can get the same rw_i (i=1, …, N) and $\boldsymbol{\Omega}$ in each slot. So they can maintain synchronization on communication channel without any handshake after the SS registers with the BS. The details of the network protocol will be introduced in Section 4.

RESOURCE SCHEDULING ALGORITHM

After all the cognitive users in the network achieve synchronization in the communication channel in each slot, we can further consider the scheduling of the network resources to guarantee the QoS requirements of all the connections admitted to the network and minimize the total energy consumption in the network. And the resource scheduling algorithm will be used in the resource scheduling module in Section 4.

Network model

The BS can establish an uplink and a downlink with each SS. So there are M uplinks and M downlinks in the PMP network. The channel bandwidth is $W=B_{a'}$ (Hz), where a' is the selected communication channel by the spectrum allocation algorithm. The access mode of the network is time division multiple access (TDMA).

Each link can maintain several connections with different QoS requirements. We classify these connections into three classes, including constant bit rate connections (symbolized by φ), real time connections (symbolized by υ) and non-real time connections (symbolized by ψ). The constant bit rate connections support real time data streams consisting of fixed length data packets such as Voice over Internet Protocol (VoIP). The real time connections support real time data streams consisting of variable-sized data packets such as Moving Picture Experts Group (MPEG) video stream. The non-real time connections support delay tolerant data streams consisting of variable-sized data packets with minimum data rate requirements, such as File Transmission Protocol (FTP).

In slot n, $l_i(n)$ connections require to transmit data on each link i. Each connection l has a data rate requirement $\bar{a}_{il}$ (in bps) and each real time connection has a delay requirement $m_{il}\cdot\tau$, $l\in\upsilon$, where m_{il} is a positive integer. Each connection has a weight e_{il}. And $\kappa_i(n)$ is the price of transmission for unit data on link i. In this paper, we make links with worse channel conditions to have higher $\kappa_i(n)$. So the links with better channel status will transmit more data.

We assume that the network is under additive white Gaussian noise (AWGN) and shadowing propagation model, and the channel status remains stable in every slot. We use adaptive multi-level quadrature amplitude modulation (M-QAM) and assume that all the BS and SS in the network have the ability of power control. Goldsmith and Chua (1997) gave the relationship between the SNR and the bit error ratio (BER) when using M-QAM. So we can get the capacity of every link (in bps) by

$$c_i(n) = w\cdot\log_2\left\{1+\frac{-1.5p_i(n)G_i(n)}{\log(5\Gamma_i)\left[I_i(n)+w\cdot\sigma_i^2(n)\right]}\right\}, \quad (3)$$

and the throughput of each connection in slot n is

$$b_{il}(n) = t_{il}(n)\cdot c_i(n), \quad (4)$$

where for each link i, $p_i(n)$ (in watt) is the transmission power of the transmitter and is constrained by p_i^{max}, $G_i(n)$ is the channel gain, $\sigma_i^2(n)$ (in watt/Hz) is the terminal noise density, $I_i(n)$ is the interference power from other links. $G_i(n)$, $\sigma_i^2(n)$ and $I_i(n)$ can be derived from the channel estimation module in the physical layer. Γ_i is the BER requirement. $t_{il}(n)$ (in second) is the transmission time arranged for each connection l on link i in every slot n.

For all the connections in the network, $t_{il}(n)$ is limited by $\sum_i\sum_l t_{il}(n)\leq\tau$, $\forall i$ where τ is the total transmission time in each slot (see Fig.1). This can be regarded as the network resource constraint for each connection.

Problem formulation

As we want to guarantee the QoS requirements for all the connections in the network, there should be an admission control module on the BS to decide whether a new connection can be admitted to the network without harming the QoS requirements of all the admitted connections. Here we symbolize the new connections by η and admitted connections by η'. So we can formulate the admission control problem as

$$\max\sum_i\sum_{l\in\eta}e_{il}(n)\bar{a}_{il}d_{il}(n), \quad (5)$$

$$\text{s.t. } \sum_i\sum_{l\in\eta}d_{il}(n)\bar{a}_{il} \leq \mu\cdot c_{th}(n) - \sum_i\sum_{l\in\eta'}\bar{a}_{il}, \quad \forall i,$$

where $d_{il}(n)$ is the result of admission control, $d_{il}(n)=1$ means the new connection is admitted to the network and $d_{il}(n)=0$ otherwise. $\mu\cdot c_{th}(n)$ (in bps) is the thresh-

old of admission control. In this paper, we choose $c_{th}(n)$ as the total channel capacity when all the links in the network transmit at the threshold power level $\theta \cdot p_i^{max}$, $0<\theta<1$, θ is the power reservation ratio. When $\theta=1$, we define congestion threshold $b_{th}(n)$ (in bit) as the total throughput in slot n. So if the total throughput of all the links exceeds $b_{th}(n)$ in slot n, the network will congest. μ should be less than the average spectrum efficiency, which represents the ratio of the number of slots occupied by the cognitive users and the total slot number. And μ is related to the number of channels sensed in each slot.

Each admitted connection manages its own queue with length $q_{il}(n)$ (in bit). In slot n, $a_{il}(n)$ bits data enter each queue with mean rate $\bar{a}_{il}$, while $r_{il}(n)$ bits data leaving each queue, transmitting to the other nodes. Notice that $a_{il}(n)$ is constant for φ, and is stochastic for υ and ψ. So in the latter case, $\bar{a}_{il}$ is the mean value of $a_{il}(n)$ in several slots.

For υ, we set a time-stamp for the incoming $a_{il}(n)$ bits data in each slot with initial value m_{il} which is the connection's delay requirement. The time-stamp decreases one in every slot. So we should send the data out before its time-stamp is less than zero to guarantee the delay requirements. We define $rt_{il}(n)$ (in bit) as the total amount of data with zero time-stamp in slot n. For φ, $rt_{il}(n)=\bar{a}_{il}$. And for ψ, to guarantee their minimal data rate requirements, we shall also give them a large time-stamp and guarantee the average data rate in the long term.

With the knowledge of the channel status $G_i(n)$, $\sigma_i^2(n)$ and $I_i(n)$ from the physical layer, $q_{il}(n)$, $rt_{il}(n)$ from the MAC layer, the transmission requirements Γ_i, $a_{il}(n)$, $e_{il}(n)$ from the application layer, we formulate the rate allocation problem as Eqs.(6)~(10) below to guarantee the different QoS requirements for every admitted connection. And we adopt utility function $\log r_{il}$ to ensure the fairness among the different kinds of connections.

$$\max\left[\sum_i\sum_{l\in\upsilon\cup\psi}e_{il}(n)\log r_{il}(n)-\sum_i\kappa_i(n)\cdot\sum_l r_{il}(n)\right],\quad(6)$$

$$\text{s.t.}\quad r_{il}(n)\le q_{il}(n),\quad\forall i,l\in\upsilon\cup\psi,\quad(7)$$

$$r_{il}(n)\ge rt_{il}(n),\quad\forall i,l\in\upsilon\cup\psi,\quad(8)$$

$$\sum_i\sum_{l\in\upsilon\cup\psi}r_{il}(n)\le b_{th}(n)-\sum_i\sum_{l\in\varphi}r_{il}(n),\quad\forall i,\quad(9)$$

$$r_{il}(n)=\bar{a}_{il}(n)\cdot\tau,\quad\forall i,l\in\varphi.\quad(10)$$

The objective function (6) maximizes the difference between the total utility of different kinds of connections and the expense of transmission in the network. It also ensures that the most important connections which have bigger weight e_{il} can transmit more data while not starving the non-real time connections by the logarithmic utility function. The economics meaning of function (6) can be the profit of the transmission in the network. The constraint (7) ensures that the data to be sent does not exceed the queue size, so no meaningless data will be transmitted in the network. Constraint (8) ensures the real time requirements of φ and υ as well as the minimal data requirements of ψ. Otherwise, a QoS violation occurs. Eq.(9) ensures the network will not congest. Eq.(10) ensures the data rate requirements of the constant bit rate connections φ in every slot.

After the rate allocation, we can go further to schedule the transmission power $p_i(n)$ for each link as well as the transmission time $t_{il}(n)$ for each connection by the network resource scheduling problem (11)~(14) below. The objective function (11) minimizes the total energy consumption of the transmission for all the users in the network. Eq.(12) ensures that the total throughput on each link is corresponding to the transmission time and power scheduled for it. Constraints (13) and (14) ensure that the scheduled resources do not exceed their constraints in each slot.

$$\min\left(\sum_i\sum_{l\in\varphi\cup\upsilon\cup\psi}t_{il}(n)\cdot p_i(n)\right),\quad(11)$$

$$\text{s.t.}\quad\sum_l r_{il}(n)=\sum_l t_{il}(n)\cdot W\cdot\log_2[1+p_i(n)\cdot h_i(n)],\\ \forall i,l\in\varphi\cup\upsilon\cup\psi,\quad(12)$$

$$\sum_i\sum_l t_{il}\le\tau,\quad(13)$$

$$p_i\le p_i^{max},\quad\forall i,\quad(14)$$

$$h_i(n)=\frac{-1.5G_i(n)}{\log(5\Gamma_i)\left[I_i(n)+W\cdot\sigma_i^2(n)\right]}\;.$$

Solution to the problems

The admission control problem (5) can be solved by greedy algorithm. We can sort $e_{il}(n)$, and admit the connections with higher weight $e_{il}(n)$ until the total transmission rate request of the admitted connections approaches the admission threshold $\mu\cdot c_{th}(n)$.

The rate allocation problem (6)~(10) is a convex optimization problem as (6) is concave and (7)~(10)

are all linear. We can solve it by Lagrangian duality (Boyd and Vandenberghe, 2004). So we introduce Lagrangian multipliers u_{il}, v_{il}, z, and get $r_{il}(n)$ by solving the following Lagrangian dual function:

$$
\begin{aligned}
&\min f_1(u_{il}, v_{il}, z) = \\
&\sum_i \sum_l \Big[e_{il} \cdot \log r_{il}(u_{il}, v_{il}, z) - (\kappa_i + u_{il} - v_{il} + z) \cdot \quad (15) \\
&r_{il}(u_{il}, v_{il}, z) + (u_{il} q_{il} - v_{il} rt_{il}) + z \cdot \big(b_{\text{th}} - \sum_i \sum_l \bar{a}_{il} \big) \Big].
\end{aligned}
$$

$$
\text{s.t.} \quad r_{il}(u_{il}, v_{il}, z) = \frac{e_{il}}{\kappa_i + u_{il} - v_{il} + z}. \quad (16)
$$

Eq.(15) can be solved by Newton's method (Boyd and Vandenberghe, 2004) with the knowledge of q_{il}, rt_{il}, κ_i, b_{th} and $\bar{a}_{il}$ as follows:

(1) Initialize u_{il}, v_{il}, z and step size δ_1, set iteration number $I_1=0$.

(2) Update u_{il}, v_{il}, z by Eqs.(17)~(19), and increase I_1 by 1.

(3) Return to Step 2 until the convergence of f_1 in Eq.(15).

(4) Calculate $r_{il}(n)$, by Eq.(16) with the final u_{il}, v_{il} and z.

$$
u_{il}(I_1+1) = \max \left\{ \left[u_{il}(I_1) - \frac{\delta}{I_1} \cdot \left(q_{il}(I_1) - \frac{e_{il}(I_1)}{X(I_1)} \right) \right], 0 \right\}, \quad (17)
$$

$$
v_{il}(I_1+1) = \max \left\{ \left[v_{il}(I_1) - \frac{\delta}{I_1} \cdot \left(\frac{e_{il}(I_1)}{X(I_1)} - rt_{il}(I_1) \right) \right], 0 \right\}, \quad (18)
$$

$$
\begin{aligned}
z(I_1+1) = \max \Big\{ \Big[z(I_1) - (\delta/I_1) \cdot \Big(b_{\text{th}} - \sum_i \sum_l \bar{a}_{il} \\
- \sum_i \sum_l e_{il}(I_1)/X(I_1) \Big) \Big], 0 \Big\}.
\end{aligned} \quad (19)
$$

where $X(I_1)=\kappa_i+u_{il}(I_1)-v_{il}(I_1)+z(I_1)$.

The network resource scheduling problem (11)~(14) can be solved by the following formula:

$$
\sum_i \frac{\sum_l r_{il}(n)}{W \cdot \ln[1 + \lambda \cdot h_i(n)]} = \tau, \quad (20)
$$

and then we can get $p_i(n)$ and $t_{il}(n)$ by

$$
p_i(n) = \min \left\{ \left[(1 + h_i(n) \cdot \lambda)^{\ln 2} - 1 \right] / h_i(n), \ p_i^{\max} \right\}, \quad (21)
$$

$$
t_{il}(n) = \frac{r_{il}(n)}{W \cdot \log_2[1 + p_i(n) \cdot h_i(n)]}. \quad (22)
$$

Eq.(20) can also be solved by Newton's method with the knowledge of h_i, r_{il}, τ and W as follows:

(1) Initialize λ and step size δ_2, set iteration number $I_2=0$.

(2) Update λ by Eq.(24), and increase I_2 by 1.

(3) Return to Step 2 until the convergence of f_2 in Eq.(23).

(4) Get the final answer λ.

$$
f_2(\lambda(I_2)) = -\tau + \sum_i \frac{\sum_l r_{il}}{W \cdot \ln[1 + \lambda(I_2) \cdot h_i]}, \quad (23)
$$

$$
\lambda(I_2+1) = \lambda(I_2) - f_2(\lambda(I_2)) / f_2'(\lambda(I_2)), \quad (24)
$$

$$
f_2'(\lambda(I_2)) = \sum_i \frac{-\sum_l r_{il} \cdot h_i}{W \cdot \{\ln[1 + \lambda(I_2) \cdot h_i]\}^2 [1 + \lambda(I_2) \cdot h_i]}.
$$

The convergence of the Newton's method to this problem is easy to be proved as follows:

As $f_2'(\lambda)<0$ in $(0, +\infty)$, $\lim_{\lambda\to 0+} f_2(\lambda)=+\infty$ and $\lim_{\lambda\to+\infty} f_2(\lambda)=-\tau$, $f_2(\lambda)$ is decreasing monotonically in $(0, +\infty)$ and there will be only one λ^* which satisfies $f_2(\lambda^*)=0$. So λ^* is the only solution to Eq.(20) and we then prove that the sequence $\lambda(I_2)$ will converge to λ^*.

We assume that λ starts from $\lambda(0)<\lambda^*$. With Lagrange mean value theorem, we have $f_2(\lambda)-f_2(\lambda^*)= f_2'(\xi) \cdot (\lambda-\lambda^*)$, where $\lambda<\xi<\lambda^*$.

So Eq.(24) can be converted to

$$
\lambda(I_2+1) = \lambda(I_2) - (\lambda(I_2) - \lambda^*) f_2'(\xi) / f_2'(\lambda(I_2)),
$$

where $\lambda(I_2)<\xi<\lambda^*$. As

$$
f_2''(\lambda) = \sum_i \frac{\sum_l r_{il}[\ln(1 + \lambda h_i) + 2]h_i^2(1 + \lambda h_i)^{-2}}{W \ln(1 + \lambda h_i)^3} > 0,
$$

$\lambda \in (0, +\infty)$, we have $f_2'(\lambda(I_2)) \leqslant f_2'(\xi)<0$, and

$$
\begin{aligned}
\lambda(I_2+1) &= \lambda(I_2) - \frac{f_2'(\xi)}{f_2'(\lambda(I_2))}[\lambda(I_2) - \lambda^*] \\
&< \lambda(I_2) - (\lambda(I_2) - \lambda^*) = \lambda^*, \\
\lambda(I_2+1) &= \lambda(I_2) - \frac{f_2'(\xi)}{f_2'(\lambda(I_2))}[\lambda(I_2) - \lambda^*] \\
&> \lambda(I_2).
\end{aligned}
$$

So we obtain $\lambda(I_2)<\lambda(I_2+1)<\lambda^*$. We can conclude that sequence $\lambda(I_2)$ is increasing monotonically and

will converge to its upper bound λ^*. When λ starts from $\lambda(0) > \lambda^*$, the convergence can also be proved with the same method above.

Notice that in our algorithm, the convergence is achieved in every slot according to the information from all the SS. Although this may incur more calculations in each slot, which will be shown in the simulation results, the power control is stable under the assumption that the channel status will be constant in each slot.

PROTOCOL FOR PMP COGNITIVE MAC

In this section, we present the MAC protocol for the PMP cognitive networks based on the proposed spectrum allocation and resource scheduling algorithms. The protocol architecture for BS and SS is summarized in Fig.2.

We assume that τ_1, τ_2, τ_3 are all fixed as shown in Fig.1. At the beginning of the downlink slots, the BS broadcasts the DUL-MAP including the connection ID (CID), the time and power scheduling results for each CID and the availability of the channels $\mathbf{\Omega}(n)$ in current slot n.

When an SS starts to work in the PMP network, it first senses all the possible channels and seeks the DUL-MAP from the BS. When the SS gets the starting time of the contention slots from DUL-MAP, it selects a contention slot randomly and sends a registration request (REG-REQ) to BS. If the BS receives the REG-REQ, it sends back a registration reply packet (REG-REP) to the dedicated SS in the next slot. Then the SS finishes registration with the BS if it receives REG-REP successfully.

For the SS which has registered with the BS, it can select the communication channel by Eq.(1) and update $\mathbf{\Omega}$ by Eq.(2) to keep frequency synchronization with the BS. The sensing decision A_1 and the sensing results Θ_{A_1} and K_{A_1} are exchanged between the spectrum allocation module and the spectrum sensing module in each cognitive user. The time slot synchronization is guaranteed by the fixed slot length and the periodic DUL-MAP packets at the beginning of the downlink slots.

Each user can raise a new connection request (CON-REQ) including its data rate requirement $\bar{a}_{il}$ and weight e_{il} to the admission control module in the BS in the contention slots or embedded in the transmission of constant bit rate connections. The BS decides whether to admit the new connection by Eq.(5). If a connection is admitted, the BS sends back a connection reply (CON-REP) to the dedicated user with a unique CID.

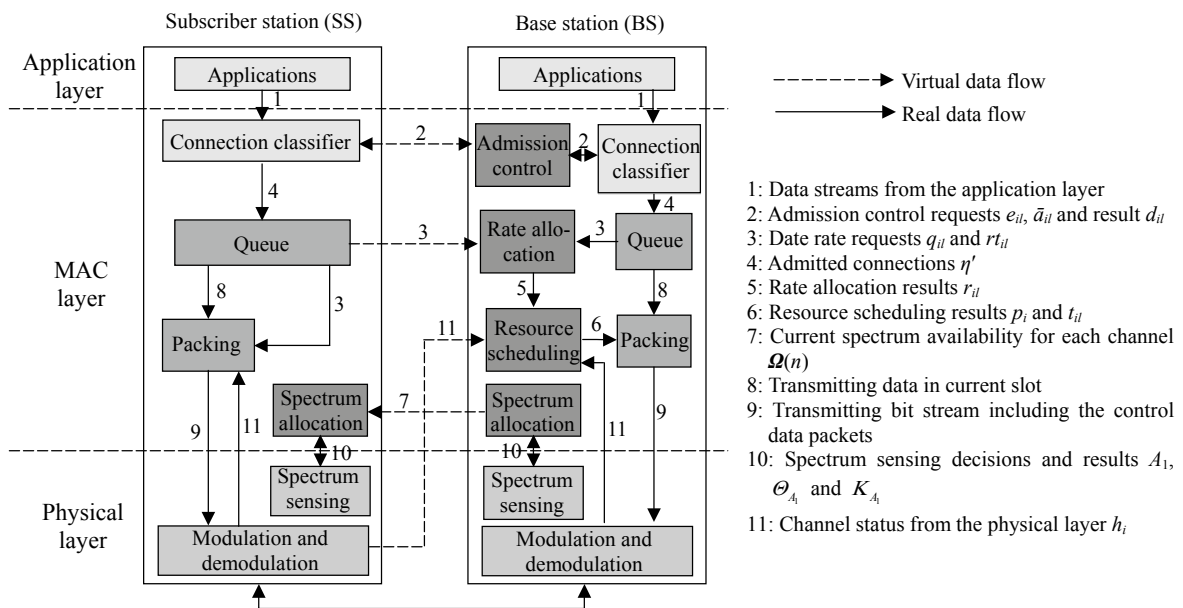Each SS maintains a management connection with the BS, which can be regarded as a constant bit

Fig.2 Protocol architecture for the subscriber station (SS) and base station (BS)

rate connection. The SS packages the data rate requirements q_{il}, rt_{il} of the admitted connections as well as the channel status h_i, and transmits the information through the management connection.

The BS decides the amount of data r_{il} transmitted in each slot for the admitted connections by the rate allocation module. This module calculates r_{il} for each user based on the data rate requirements q_{il} and rt_{il} of every admitted connection by Eqs.(6)~(10).

The rate allocation results are then sent to the resource scheduling module. And this module schedules the transmission power p_i for each user and transmission time t_{il} for each connection based on r_{il} and the channel status h_i by Eqs.(11)~(14). The scheduling results are packaged in the DUL-MAP and broadcasted to the SS at the beginning of the transmission stage.

SIMULATION RESULTS

To demonstrate the performance of the proposed algorithms, we consider a cognitive PMP network with one BS and 10 SS. The 10 SS are placed in a circle area with radius 100 m and the BS is at the center. There are 10 possible channels in the network, each with the same bandwidth of 4 MHz and the transmission probabilities α=[0.1, 0.2, 0.2, 0.4, 0.4, 0.6, 0.6, 0.8, 0.8, 0.9], β=[0.9, 0.8, 0.8, 0.6, 0.6, 0.4, 0.4, 0.2, 0.2, 0.1]. We assume the propagation model for each channel is the same, and let $G_i = M \cdot S_i \cdot (d_0/d_i)^\rho$, where $M=1$, $d_0=50$ m, $\rho=4$, d_i is the length of link i and S_i follows log-normal distribution with mean 0 and variance 6 dB. And $\sigma_i^2(n)$ is 10^{-10} W/Hz, p_i^{max} is 0.2 W, Γ_i is 10^{-6}.

The transmission time in each slot is τ=10 ms. In every slot, each SS and BS can generate different kinds of connections. For φ, $a_{il}(n)$ is constant, and for υ and ψ, $a_{il}(n)$ follows exponential distribution with mean $\bar{a}_{il}$. $\bar{a}_{il}$ distributes randomly from 100 kbps to 1 Mbps.

Two cases were simulated and for each case we simulated 20 times. In Case 1, we only sensed one channel in each slot, while three channels in Case 2. The power reservation ratio in admission control process increased 0.05 each time, from 0.05 to 1. And each simulation lasted 5000 slots. In each simulation, the traffic in the network was heavy and the sum of

data rate requests always exceeded the admission threshold $\mu \cdot c_{th}(n)$.

For each case, we got the spectrum efficiency in 5000 slots. Under each power reservation ratio, we also got the average QoS violation probability, the average iteration number in 5000 slots.

Fig.3 shows the spectrum efficiency in both cases. The spectrum efficiency is the ratio of the number of slots occupied by the cognitive users and the total slot number. We also show the performance of random algorithm which selects the sensing channels randomly. It can be seen that in Case 1, the greedy algorithm achieves a significant gain over the random algorithm with the knowledge of the transmission probability of the channel status. The spectrum efficiency increases significantly when more channels are sensed in each slot. And the greedy algorithm still achieves better performance than the random algorithm.
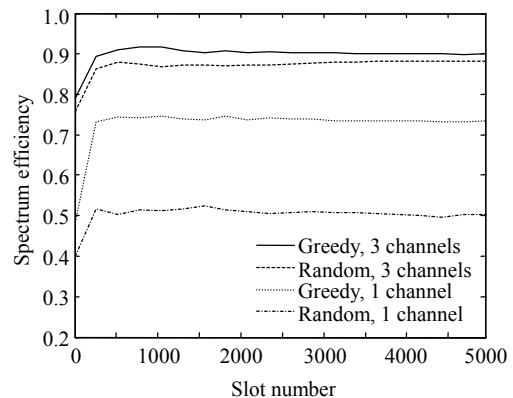

Fig.3 Comparison of the spectrum efficiency

Fig.4 shows the QoS violation probability under different power reservation ratios in each case. A QoS violation means that $r_{il}(n)$ is less than $rt_{il}(n)$ in any slot. This is possible because $a_{il \in \upsilon \cup \psi}(n)$ follows exponential distribution and is variable. It can be seen that in Case 1, when the power reservation ratio is more than 0.4, there will be QoS violations in each slot. And the QoS violation probability increases basically with the power reservation ratio. The performance in Case 2 is better than that in Case 1 because the spectrum efficiency in Case 2 is higher. So we can choose a proper power reservation ratio in the region between 0.3 and 0.6 to guarantee the QoS requirement for all connections in the network according to Fig.4.
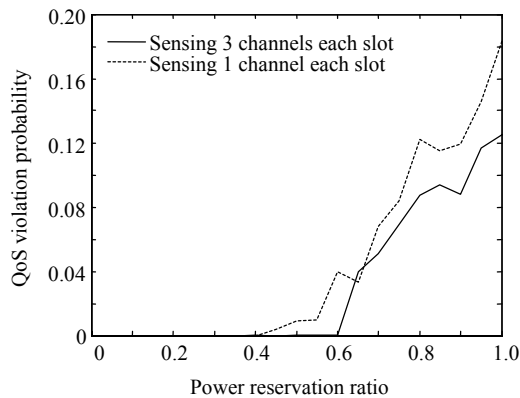
Fig.4 QoS violation probability under different power reservation ratios

Fig.5 shows the average iteration number in Case 2. We can see that I_1 in Eqs.(17)~(19) increases basically with the power reservation ratio, and the iteration number I_2 in Eqs.(23) and (24) decreases with the power reservation ratio. This is because with the increase of the power reservation ratio, the gap between $rt_{il}(n)$ and $q_{il}(n)$ increases, so I_1 increases. Meanwhile, the total rate on each link becomes more balanced, so I_2 decreases.
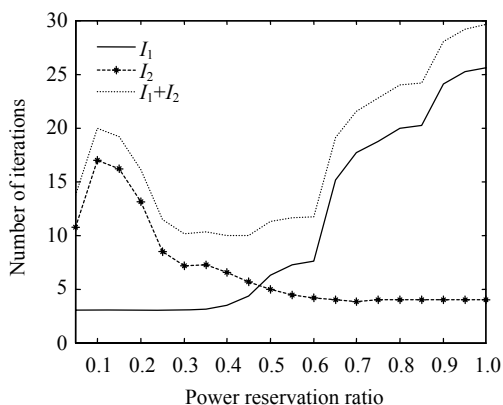

Fig.5 Iteration number under different power reservation ratios

It is easy to see that the total iteration number I_1+I_2 is quite small when the power reservation ratio is between 0.3 and 0.6, while this region can provide strong QoS guarantee as shown in Fig.4. So the computation complexity of the resource scheduling algorithm is low.

CONCLUSION

We have presented a cross-layer design approach on the spectrum allocation and resource scheduling problems in the cognitive PMP networks with rapid changes of spectrum opportunities. The spectrum allocation algorithm maintains the synchronization among all users and makes efficient use of the selected channel. The resource scheduling algorithm couples the physical layer, the MAC layer and the application layer and does optimal scheduling of transmission time and power for different kinds of connections in the network. The proposed MAC protocol based on the algorithms can manage the registration, admission control and resource allocation efficiently.

However, our work is based on the PMP networks which have a central node. For a distributed network in the same environment, efficient scheduling becomes more difficult especially when the data packets transmit through several hops. Also, for an environment with slower changes of spectrum opportunities, more technologies (such as OFDMA, multi-interface transmission) can be introduced to the cognitive radio networks. And more efficient and fast spectrum sensing algorithms should be studied. These will be our further work.

References

Boyd, S.P., Vandenberghe, L., 2004. Convex Optimization. Cambridge University Press, Cambridge, p.215-271.

Cabric, D., Mishra, S.M., Brodersen, R.W., 2004. Implementation Issues in Spectrum Sensing for Cognitive Radios. Proc. 38th Asilomar Conf. on Signals, Systems and Computers. Pacific Grove, California, USA, p.772-776. [doi:10.1109/ACSSC.2004.1399240]

Federal Communications Commission (FCC), 2002. Report of the Spectrum Efficiency Working Group.

Goldsmith, A.J., Chua, S.G., 1997. Variable-rate variable-power MQAM for fading channels. *IEEE Trans. on Commun.*, **45**(10):1218-1230. [doi:10.1109/26.634685]

Haykin, S., 2005. Cognitive radio: brain-empowered wireless communications. *IEEE J. Sel. Areas Commun.*, **23**(2): 201-220. [doi:10.1109/JSAC.2004.839380]

Hoang, A.T., Liang, Y.C., 2006. Maximizing Spectrum Utilization of Cognitive Radio Networks Using Channel Allocation and Power Control. IEEE 64th Vehicular Technology Conf., p.1-5. [doi:10.1109/VTCF.2006.257]

Johansson, M., Xiao, L., 2006. Cross-layer optimization of wireless networks using nonlinear column generation. *IEEE Trans. on Wirel. Commun.*, **5**(2):435-445.　[doi:10. 1109/TWC.2006.1611067]

Jovicic, A., Viswanath, P., 2006. Cognitive Radio: An Information-Theoretic Perspective. IEEE Int. Symp. on Information Theory, p.2413-2417.　[doi:10.1109/ISIT.2006. 262021]

Lansford, J., 2004. UWB Coexistence and Cognitive Radio. Int. Workshop on Ultra Wideband Systems, p.35-39. [doi:10.1109/UWBST.2004.1320898]

Mitola, J., Maguire, G.Q., 1999. Cognitive radio: making software radios more personal. *IEEE Pers. Commun.*, **6**(4):13-18.　[doi:10.1109/98.788210]

Wang, W., Liu, X., 2005. List-coloring Based Channel Allocation for Open Spectrum Wireless Networks. IEEE 62nd Vehicular Technology Conf., p.690-694.　[doi:10.1109/ VETECF.2005.1558001]

Won, J.Y., Shim, S.B., Kim, Y.H., Hwang, S.H., Song, M.S., Kim, C.J., 2006. An Adaptive OFDMA Platform for IEEE 802.22 Based on Cognitive Radio. Asia-Pacific Conf. on Communication, p.1-5.　[doi:10.1109/APCC. 2006.255809]

Xiao, L., Johansson, M., Boyd, S.P., 2004. Simultaneous routing and resource allocation via dual decomposition. *IEEE Trans. on Commun.*, **52**(7):1136-1144.　[doi:10. 1109/TCOMM.2004.831346]

Zhao, Q., Tong, L., Swami, A., Chen, Y., 2007. Decentralized cognitive MAC for opportunistic spectrum access in ad hoc networks: a POMDP framework. *IEEE J. Sel. Areas Commun.*, **25**(3):589-600.　[doi:10.1109/JSAC.2007.070 409]