一种基于马尔可夫决策过程的认知无线电网络传输调度方案

朱　江[①]　　徐斌阳[②]　　李少谦[①]
[①](电子科技大学通信抗干扰国家级重点实验室　成都　610054)
[②](上海贝尔阿尔卡特创新中心　上海　201206)

摘　要：该文提出了一种适用于认知无线电网络的跨层传输调度方案，即满足掉包率约束的前提下最小化平均功率消耗。此方案被建模为约束马尔可夫决策过程(MDP)。采用拉格朗日乘子法求解此 MDP，并且提出了一种黄金分割乘子搜索法。提出两种简化方法，即状态聚合以及行动集缩减来解决维灾问题。仿真结果显示简化方法对该方案的性能影响很小，且该方案的平均功耗最低。

关键词：认知无线电；马尔可夫决策过程；跨层设计；传输调度
中图分类号：TN92　　**文献标识码**：A　　**文章编号**：1009-5896(2009)08-2019-05

A Transmission and Scheduling Scheme Based on Markov Decision Process in Cognitive Radio Networks

Zhu Jiang[①]　　Xu Bin-yang[②]　　Li Shao-qian[①]
[①](*National Key Laboratory of Communications*, *University of Electronic Science and Technology of China*, *Chengdu* 610054, *China*)
[②](*Research and Innovation Center of Alcatel Shanghai Bell*, *Shanghai* 201206, *China*)

Abstract: A cross-layer transmission and scheduling scheme of average power minimization in cognitive radio networks under the constraint of packet drop probability is addressed. The scheme is formulated by constrained Markov Decision Process (MDP). Lagrangian multiplier approach is used to solve the MDP, and a golden section search method is proposed to find the multiplier. Two simplifying methods, namely, state aggregate and action set reduction are employed to cope with the curse of dimensionality. Simulation results show that simplifying methods have little influence on the performance of the scheme and average power consumption of the scheme is the lowest.

Key words: Cognitive Radio (CR); Markov Decision Process (MDP); Cross-layer design; Transmission and scheduling

1　引言

认知无线电(CR)技术是未来无线通信系统中用于缓解频谱资源紧张的主要解决方案，而跨层设计是 CR 系统的关键技术之一[1]。本文提出的跨层设计方案是在满足缓存器掉包率约束的前提下，最小化平均功率消耗。用 MDP 过程对此方案建模，考虑了频谱可用性的变化规律、上层数据的到达过程和信道衰落。用拉格朗日乘子法求解 MDP 的最优策略，并提出了一种黄金分割(GS)乘子搜索算法。考虑到最优策略的求解复杂度与状态空间和行动集的规模相关，提出了基于状态聚合和行动集缩减的近似最优策略求解方法。

2008-07-30 收到，2009-01-05 改回
国家自然科学基金(60496313)，国家 863 计划项目(2005AA123910，2007AA01Z209)和国家 973 规划项目(2009CB320405)资助课题

2　系统模型

考虑单个发送机通过 M 个窄带块衰落信道择机地向接收机发送数据的情况。上层数据包以平均速率 λ 到达长度为 L 的缓存器。各信道都采用了自适应调制(AM)。帧长为 T_f，在每帧的开始，发送机和接收机感知信道并交换感知结果(SR)，以获得信道的可用性信息。如果某信道被主用户系统占用，则 SR 为 0，否则为 1。若 SR 为 1，发送机可以通过发送导频信息获得接收机反馈的信道状态信息(CSI)。发送机可以根据缓存器的状态和所有信道的 SR 、CSI 决定如何发送数据包。若某信道的数据包被成功接收，接收机将反馈确认消息 ACK，否则反馈失败消息 NAK。没有被成功发送的数据包将被重发。

信道 $m \in \{1,\cdots,M\}$ 的 SR 可以用离散时间马尔可夫链描述[1]，其状态空间为 $\{0,1\}$，转移概率为

$p_f^m(f^m,f'^m)$。M 个信道 SR 的状态空间 $F=\{0,1\}^M$。若每个信道 SR 是独立的[1]，则转移概率为 $p_F(f,f')=\prod_{m=1}^M p_f^m\left(f^m,f'^m\right)$，且 $f,f'\in F$。

单个块衰落信道的状态可被建模为离散时间马尔可夫链(FSMC)[2]。对于合并加性高斯白噪声的瑞利信道 m，其接收信噪比(SNR)被门限值 $0=\Gamma_0<\Gamma_1\cdots\Gamma_K=\infty$ 分成若干区间。如果 $\Gamma_k<\mathrm{SNR}\le\Gamma_{k+1}$，则信道 m 的状态为 h_k。$H^m\triangleq\{h_0^m,h_1^m,\cdots,h_{K-1}^m\}$ 为信道 m 的状态空间，给定最大多普勒频移 f_d^m，平均 SNR $\overline{\gamma}_0^m$ 及 T_f 可确定信道 m 的状态转移概率 $p_{H^m}(h_k^m,h_{k+1}^m)$ [2]。M 个信道的组合状态空间为 $H=H^1\times\cdots\times H^M$，组合状态为 $h\triangleq\{h^1,\cdots,h^M\}\in H$，转移概率为 $p_H(h,h')$。每个信道的 AM 方案基于 M 元正交调幅(M-QAM)。定义传输速率空间 $U\triangleq\{u_0,u_1,\cdots,u_{U-1}\}$。其中 u_0 和 u_1 分别对应无数据传输和二进制相移键控，u_j，$j\in\{2,\cdots,U-1\}$ 对应 2^j-QAM。

q 为一帧内到达缓存器的数据包数，q 服从泊松分布且平均到达率为 $\overline{\lambda}$，各帧内 q 独立同分布。定义缓存器的状态空间 $B\triangleq\{b_0,b_1\cdots b_l\cdots b_L\}$，其中 b_l 对应为缓存器内有 l 个包。某帧开始时缓存器内有 b 个包，有 a 个包从缓存器内取出，则下帧开始时缓存器内的包数为

$$b'=\min\left\{b-\sum_{m=1}^M a^m I\{\delta_m=\mathrm{ACK}\}+q,\,L\right\} \quad (1)$$

其中 $\sum_{m=1}^M a^m=a$，a^m 为通过信道 m 发送的包数，$\delta_m\in\{\mathrm{ACK},\mathrm{NAK}\}$ 为信道 m 的反馈确认消息。$I\{x\}$ 是指示函数，如果 x 为真，函数为 1，否则为 0。每个信道的 δ_m 为独立同分布的随机变量，给定纠错编码方案，它是 BER 的函数[2]。

3 MDP 与传输调度问题

MDP 的状态空间可以定义为 $S\triangleq B\times H\times F$，转移概率为

$$p_s(s,s'/a)=p_H(h,h')p_F(f,f')p_q(q)\cdot I\left\{b'=\min\left\{b-\sum_{m=1}^M a^m I\{\delta_m=\mathrm{ACK}\}+q,\,L\right\}\right\}$$

其中 $a\in A$，$s\triangleq\{b,h,f\}$，$s'\triangleq\{b',h',f'\}$。但此定义并不符合发送机的观察结果。当 $f^m=0$，接收机不可能获得信道 m 的 CSI。将这些不可观察的状态聚合为一个宏状态[3]，得到新的状态空间 $\widehat{S}$。为每个信道 m 引入新的不可观察状态 h_{-1}^m，则 $\widehat{H}^m\triangleq H^m+\{h_{-1}^m\}$，$\widehat{H}=\widehat{H}^1\cdots\times\widehat{H}^M$，$\widehat{S}\triangleq B\times\widehat{H}$。由文献[3]可得 $\widehat{S}$ 的状态转移概率 $p_{\hat{s}}(\hat{s},\hat{s}'/a)$，$\hat{s},\hat{s}'\in\widehat{S}$。

MDP 的行动集 $A\triangleq\{a_0,a_1,\cdots,a_N\}$。当 a_i 被采纳，发送机从缓存器中取出 i 个包并通过 M 个信道发送给接收机。给定 a_i，引入如何根据信道状态将 i 个包分配到 M 个信道使得消耗的功率最小的问题。此类问题及相关算法在很多系统中都被提及(例如 OFDM，MIMO)，本文不再赘述。定义线性映射 $\varphi(u_j)\triangleq a^m$，即传输速率和 a^m 的关系。不失一般性，假设 $\varphi(u_j)=j$，可得到 $0\le a^m\le(U-1)$ 且 $N=M(U-1)$。给定比特误码率(BER)、噪声功率 WN_0，当信道 m 的状态为 $\hat{h}^m\in\widehat{H}^m$，可以确定通过信道 m 发送 a^m 个包所需的最小功率 $P(\hat{h}^m,a^m)$ [4]。当 $a^m=0$ 时，$P(\hat{h}^m,a^m)=0$；当 $a^m\ne0$ 且 $\hat{h}^m\in\{h_{-1}^m,h_0^m\}$ 时，$P(\hat{h}^m,a^m)\to\infty$。通过 M 个信道发送 a 个包所需最小功率为 $P(\hat{h},a)=\sum_{m=1}^M P(\hat{h}^m,a^{m*})$，其中 $\hat{h}\in\widehat{H}$，$a=\sum_{m=1}^M a^{m*}$，a^{m*} 为信道 m 的最优分配结果。

在 MDP 的第 n 个决策周期，即第 n 帧，如果系统状态为 $\hat{s}(n)\in\widehat{S}$，且发送机发送 $a(n)\in A$ 个包，则负收益为

$$r(\hat{s}(n),a(n))=P\left(\hat{h}(n),a(n)\right) \quad (2)$$

其中 $\hat{h}(n)$ 是 $\hat{s}(n)$ 的组合信道状态。将代价定义为缓存器的掉包率：

$$c(\hat{s}(n),a(n))=\overline{\lambda}T_f\left(1-\sum_{q=0}^{L-b(n)+a(n)-1}p_q(q)\right)-\left(L-b(n)+a(n)\right)\left(1-\sum_{q=0}^{L-b(n)+a(n)}p_q(q)\right) \quad (3)$$

其中 $b(n)$ 是状态为 $\hat{s}(n)$ 时缓存器内的包数，$p_q(q)=\exp(-\overline{\lambda}T_f)(\overline{\lambda}T_f)^q/q!$。

4 MDP 的求解问题

4.1 拉格朗日乘子法

方案设计必须满足掉包率约束 $\overline{D}$，为此引入拉格朗日乘子 ζ，将优化问题定义为

$$\hat{\pi}^*(\zeta)=\arg\min_{\hat{\pi}(\zeta)}\left\{\mathbb{E}_{\hat{\pi}(\zeta)}\left[\lim_{N\to\infty}\sum_{n=1}^N i\left(\boldsymbol{s}(n),a(n)\right)\Big/N\right]\right\} \quad (4)$$

其中 $\hat{\pi}^*(\zeta)$ 为给定 ζ 时 MDP 的最优策略，$i(\boldsymbol{s}(n),a(n))=r(\boldsymbol{s}(n),a(n))+\zeta c(\boldsymbol{s}(n),a(n))$，$\mathbb{E}_{\hat{\pi}(\zeta)}[\bullet]$ 是采纳策略 $\hat{\pi}(\zeta)$ 时的期望。可根据值迭代(VI)算法得到 $\hat{\pi}^*(\zeta)$[5]。可利用 ζ 的单调非增特性[6]及 GS 搜索算法的基本思想，实现一种基于 GS 的乘子搜索算法。由于该 MDP 只有一个约束，因此最优策略是两个次优策略的混合[6]。

4.2 维灾问题

用 VI 求解 MDP 的最优策略需要存储每个状态-行动对的 Q 值，即 $Q(s,a)$ [5]。当状态空间、行动集的规模很大时，既消耗大量内存，又减慢算法收敛速度，即所谓维灾问题[5]。

本文中 MDP 的状态数为 $(L+1)M^{K+1}$，M 过大时 MDP 不可解。为此通过状态聚合[5]减小状态空

间的规模，即状态数。

引理 1　将状态空间 $\hat{S}$ 划分为 W 个没有交集的子空间 $\hat{S}_1,\hat{S}_2\cdots\hat{S}_W$。给定 a，对于任意 $\hat{s}\in\hat{S}_w$，如果有 $Q(\hat{s},a)=\sigma(w,a)$，则基于状态聚合的 VI 能够得到最优策略[5]。

如果 W 足够小，则 MDP 可解。状态聚合的原则是：发送机无需知道每个信道的状态，只需知道落在每个状态里的信道数。将组合信道状态空间 $\widehat{H}$ 划分为若干子空间 $\widetilde{H}_1,\widetilde{H}_2,\cdots,\widetilde{H}_G$。属于子空间 $\widetilde{H}_g$ 的任意状态具有相同特点，即落在每个状态里的信道数相同。子空间数的表达式为 $G={}^{M+K}C_K$，其中 ${}^nC_k$ 是 n 个元素中取 k 个的组合函数。基于此状态聚合，重新定义状态空间为 $\widetilde{S}\triangleq B\times\widetilde{H}$，其中 $\widetilde{H}$ 的元素与 $\widehat{H}$ 的子空间一一对应。采用状态聚合得到是 MDP 的近似最优策略[5]。但某些假设条件下，此近似最优策略等价于最优策略。

定理 1　如果每个信道的状态转移概率独立且相同，且每个信道 SR 的转移概率独立且相同，则采用状态聚合得到的近似最优策略等价于最优策略。

可用数学归纳法证明该定理(略)。

定义状态 $\tilde{s}\triangleq\{b,\tilde{h}\}$，其中 $\tilde{s}\in\widetilde{S}$ 且 $\tilde{h}\in\widetilde{H}$，该状态下的行动集缩减基于规则：$A=\{a_0,\cdots,a_{\min(b,V(U-1))}\}$。其中 V 是信道状态不为 h_{-1} 和 h_0 的信道数。

引理 2　对于以最小化负收益为优化目标的 MDP，给定状态 s，若 $i(s,a')<i(s,a)$ 且 $p_s(s,s'/a)=p_s(s,s'/a')$，$\forall s'\in S$，s' 为下一个状态。则该状态下采取 a' 比采取 a 好。

可根据 VI 中迭代过程的基本定义证明该引理(略)。

定理 2　基于上述行动集缩减得到的策略等价于最优策略。

定理 2 的证明过程基于引理 2，基本思路陈述为：首先，如果某信道的状态为 h_{-1} 或 h_0，则通过该信道发送数据包不会成功，且会消耗功率。由于发送失败的包将被重发，因此无论是否发送数据包都不能改变状态的变化结果。所以当某信道的状态为 h_{-1} 或 h_0，不通过该信道发送数据包是更好的行动。其次，发送的包数不能超出缓存器内的包数。因为发送空包既要消耗不必要的能量，也无法改变状态的变化结果。

5　仿真结果与分析

先利用 VI 联合 GS 乘子搜索法求解出各策略，然后模拟出具有马尔科夫特性的系统状态。根据所得的策略，在不同状态下采取相应的行动，最终得到消耗的平均功率。每次仿真都模拟了 10^6 个决策周期。假设每个数据包中有 100 bit，系统纠错功能可以容许的最大错误为 10 bit，可得 δ_m 的概率分布[2]。对于相邻的窄带信道，假设 $f_d^m=f_d$，$\bar{\gamma}_0^m=\gamma_0$，$\forall m$，则每个信道的状态转移概率独立且相同[7]。同时可以假设每个信道 SR 的转移概率独立[1]。综合定理 1 和定理 2，当两个假设成立，采用状态聚合以及行动集缩减而得到的近似最优策略等价于最优策略。如果每个信道 SR 的转移概率不同，则近似最优策略的性能次于最优策略。为此设置两种仿真方案：方案 1 中每个信道 SR 的转移概率都等于 $\boldsymbol{P}_a$；方案 2 中每个信道 SR 的转移概率分别为 $\boldsymbol{P}_a$ 和 $\boldsymbol{P}_b$。系统参数为：$T_f=2\text{ms}$，$f_d=50\text{Hz}$，$M=2$，$\gamma_0=0\text{dB}$，$\Gamma_1=-5.41\text{dB}$，$\Gamma_2=-0.08\text{dB}$，$L=15$，$WN_0=1\text{mW}$，$\text{BER}=10^{-3}$，$U=5$，$\boldsymbol{P}_a=\begin{matrix}&0&1\\0&0.88&0.12\\1&0.04&0.96\end{matrix}$，$\boldsymbol{P}_b=\begin{matrix}&0&1\\0&0.96&0.04\\1&0.13&0.87\end{matrix}$。

图 1 显示了不同 $\bar{D}$ 的平均功率。平均功率随 $\bar{D}$ 增大而减小，这是因为较大的 $\bar{D}$ 使得发送机可以选择在较好的信道条件下发送较少的数据包。图 1 给出方案 1，方案 2 在给定不同 $\bar{\lambda}$(包/帧)下最优策略 OP-1, OP-2 和近似最优策略 AOP-1, AOP-2 的结果。图 2 显示了不同 $\bar{\lambda}$ 的平均功率。平均功率随 $\bar{\lambda}$ 增大而增大，这是因为当 $\bar{\lambda}$ 增大时，为维持给定的 $\bar{D}$，即使在信道条件较差的情况下发送机也可能要增加数据包的发送量。图 2 给出了方案 1，方案 2 在给定不同 $\bar{D}$ 下 OP-1, OP-2 和 AOP-1, AOP-2 的结果。综合图 1 和图 2 可以看到：方案 1 中，由于子空间中每个子状态的 Q 值完全相同，AOP-1 和 OP-1 完全相同，相同的策略在经历同一次仿真后所得到的平均功率完全相同；方案 2 中，由于子空间中每个子状态的 Q 值不完全相同，AOP-2 和 OP-2 有差异，此差异使得两种策略在经历同一次仿真后所得到的平均功率有区别(区别非常小，无法直接从图上观察出来)。据前文的分析，聚合前系统共有 $|\hat{S}|=(L+1)(K+1)^M=256$ 种状态，聚合后减少为 $|\tilde{S}|=(L+1)^{M+K}C_K=160$ 种状态。结合行动集缩减，需存储的 $Q(\tilde{s},a)$ 数为 $\sum_{\tilde{s}\in\widetilde{S}}(1+\min(b,V(U-1)))=652$，而未进行状态聚合及行动集缩减时，需存储的 $Q(\tilde{s},a)$ 数为 $\sum_{\hat{s}\in\hat{S}}\big(1+M(U-1)\big)=2304$，其存空间压缩率为 71.70%。所以本文提出的状态聚合以

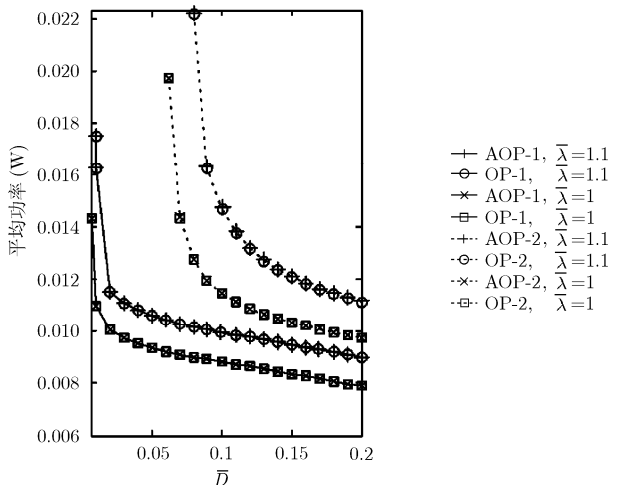

图 1 不同 $\overline{D}$ 的平均功率

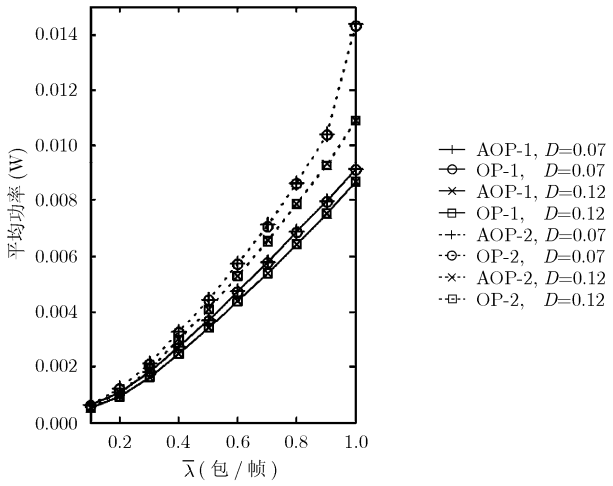

图 2 不同 $\overline{\lambda}$ 的平均功率

储及行动集缩减能够减小 MDP 求解所需的存储空间，加速 VI 的收敛，而为此带来的性能下降却很小。

图 3 和图 4 给出了根据本文所提出的调度方案而得到的策略(APP)和其它两种策略的性能比较，且设定每个信道 SR 的转移概率都等于 $\boldsymbol{P}_a$。启发式策略(HP)是策略 π_b 和 π_d 的概率混合，且混合概率分别为 p_a 和 $1-p_a$。其中 π_b 可表示为 $a=\min(b, V(U-1))$，$\forall \tilde{s}\in\tilde{S}$；$\pi_d$ 可表示为 $a=a_0$，$\forall \tilde{s}\in\tilde{S}$。$p_a$ 由 $\overline{D}$ 决定：$p_a\overline{D}(\pi_b)+(1-p_a)\overline{D}(\pi_d)=\overline{D}$。文献[8]给出了一种不同的能耗优化标准：每消耗单位功率时，成功传输的数据包数，即功效。其优化目标是最大化平均功效，由此得到的策略被称为平均功效策略(AEP)。给定 $\overline{\lambda}=1.1$，从图 3 中可以看出，当 $\overline{D}$ 很小，即为 0.01085 时，3 种策略的性能相同。这是由于为满足很小的 $\overline{D}$，3 种策略都等价于 π_b。由于 HP 是 π_b 和 π_d 由 $\overline{D}$ 决定的线性概率混合，所以其平均功率是 $\overline{D}$ 的线性减函数。给定 $\overline{D}=1.2$，从图 4 中可以看出，当 $\overline{\lambda}=0$ 时，3 种策略的平均功率都为 0，因为此时 3 种策略都退化为 π_d。综合图 3 和图 4，APP 获得了最好的平均功率性能，这是因为它是以最小化平均功率消耗作为优化目标，理论上可以获得最优值[5]。而 AEP 没有将最小化平均功率作为优化目标，尽管它也是一种能耗优化标准，但与 APP 的优化目标并不等价，所以其平均功率性能次之。HP 的平均功率是 $\overline{D}$ 的线性减函数，因此性能最差。另一方面，HP 不需通过 VI 以及 GS 乘子搜索法来求解，所以求解最简单，而另外两种策略的求解复杂度相同。

6 结束语

本文将 CR 网络的跨层传输调度问题建模为 MDP，并提出了相应的求解方法。在未来的工作中，将把系统扩展到参数未知的情况下，并利用增强学习算法求解策略[5]。还将考虑信道的 SR 和 CSI 有错误和时延的情况。

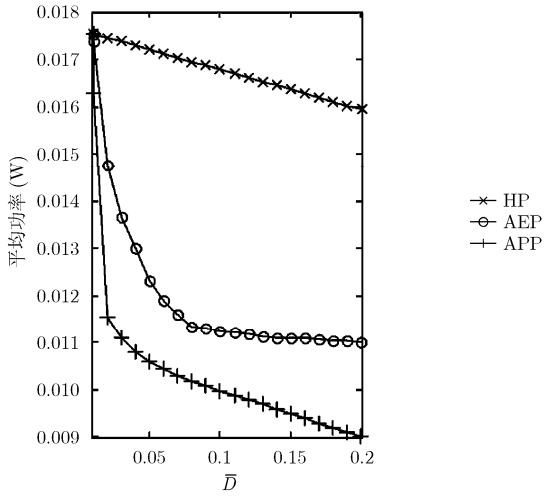

图 3 不同 $\overline{D}$ 下 3 种策略的平均功率

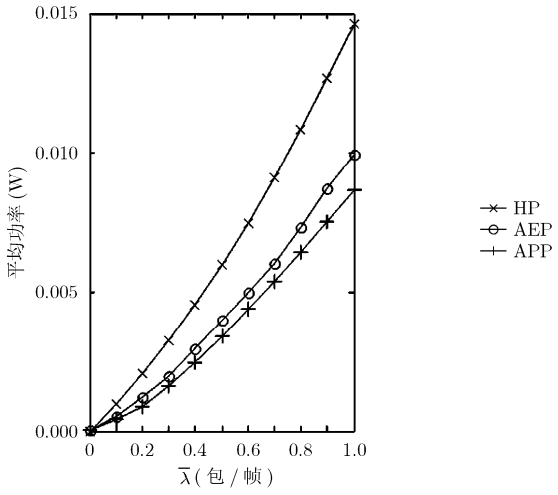

图 4 不同 $\overline{\lambda}$ 下 3 种策略的平均功率

参 考 文 献

[1] Hossain E and Bhargava V. Cognitive Wireless Communication Networks [M]. First Edition, New York: Springer, 2007: 1–301.

[2] Djonin D V, *et al*.. Joint rate and power adaptation for type-I hybrid ARQ systems over correlated fading channels under different buffer cost constraints [J]. *IEEE Transactions. on Wireless Communications*, 2008, 57(1): 421–435.

[3] Bolch G, *et al*.. Queueing Networks and Markov Chains: Modeling and Performance Evaluation with Computer Science Applications [M]. Second Edition, New York: John Wiley & Sons, 2006: 185–206.

[4] Chung Seong Taek and Goldsmith A. Degrees of freedom in adaptive modulation: A unified view [J]. *IEEE Transactions. on Communications*, 2001, 49(9): 1561–1571.

[5] Chang H S, *et al*.. Simulation-based Algorithms for Markov Decision Processes [M]. First Edition, London: Springer-Verlag, 2007: 9–167.

[6] Beutle F J and Ross K W. Optimal policies for controlled markov chains with a constraint [J]. *Journal of Mathematical Analysis and Application*, 1985, 112(1): 236–252.

[7] Hossain M J, *et al*.. Delay limited optimal and suboptimal power and bit loading algorithms for OFDM systems over correlated fading [C]. IEEE GLOBECOM, St. Louis, USA, Dec. 1–2, 2005: 3448–3453.

[8] Pandana C and Liu K J R. Near-optimal reinforcement learning framework for energy-aware sensor communications [J]. *IEEE Transactions. on Wireless Communications*, 2005, 23(4): 788–797.

朱　江：　男，1977 年生，博士生，研究方向为跨层设计.

徐斌阳：　男，1977 年生，博士，研究方向为无线资源管理.

李少谦：　男，1957 年生，教授，博士生导师，研究方向为移动与无线通信系统.