

# Power Control for Femtocell Two-Tier Networks Based on Distributed $Q$ -Learning

Haiqin Ning, Peisheng Pan

Nanjing University of Posts and Telecommunications, College of Communication and Information Technology, Nanjing  
Email: ninghaiqin2010@163.com, panps@njupt.edu.cn

Received: Jan. 8<sup>th</sup>, 2013; revised: Jan. 11<sup>th</sup>, 2013; accepted: Jan. 25<sup>th</sup>, 2013

**Abstract:** With the development of the next generation wireless communications technology, the concept of Femtocell came into being. Femtocell is a kind of home base station which can expand coverage indoor, improve system performance and decrease macrocell communication overhead. However, the existing co-channel interference of the network between macrocells and femtocells can dramatically degrade the overall performance of the network. For this reason, we did a research and wrote the paper. First we analyze a two-tier network of distributed  $Q$ -learning algorithm in Femtocell and Macrocell downlink power control, then simulate and analyze the independent  $Q$ -learning algorithm and cooperation  $Q$ -learning algorithm based on that. In addition, we optimize the cooperation  $Q$ -learning algorithm so that it can ensure that the value of return function is positive in the case of satisfying the base station transmit power and capacity performance, and is negative in other cases. This improvement can indeed increase the efficiency and fairness of the system capacity after the confirmation of matlab simulation.

**Keywords:** Power Control; Femtocell; Interference; Independent  $Q$ -Learning; Cooperative  $Q$ -Learning

## 基于分布式 $Q$ 学习的 Femtocell 两层网络功率控制方法

宁海芹, 潘沛生

南京邮电大学通信与信息工程学院, 南京  
Email: ninghaiqin2010@163.com, panps@njupt.edu.cn

收稿日期: 2013 年 1 月 8 日; 修回日期: 2013 年 1 月 11 日; 录用日期: 2013 年 1 月 25 日

**摘要:** 随着下一代无线通信技术的发展, Femtocell 的概念应运而生。Femtocell 是一种可以改善室内覆盖的家庭基站, 其主要致力于增加室内覆盖, 改善系统性能, 减少 Macrocell 网络通信负担。然而, 在 Femtocell 与 Macrocell 共存的双层网络中存在同信道干扰问题, 这会降低系统性能。本文首先分析了分布式  $Q$  学习算法在 Femtocell 与 Macrocell 构成的两层网络下行链路功率控制中的应用, 在此基础上仿真并分析了独立  $Q$  学习 (Independent  $Q$ -learning, IQL) 算法和合作  $Q$  学习 (Cooperation  $Q$ -learning, CQL) 算法, 并对 CQL 进行了改进, 保证在满足基站发射功率和容量性能的情况下回报函数值是正值, 其他情况下是负值, 经 matlab 仿真验证, 这可提高系统容量效率和公平性。

**关键词:** 功率控制; Femtocell; 干扰; IQL; CQL

### 1. 引言

研究表明<sup>[1]</sup>无线通信网络中, 超过 50% 的语音业务和超过 70% 的数据业务发生在室内, 室内覆盖的好坏直接影响着用户对无线通信服务的满意度。由此

Femtocell 作为一种有效地解决方法应运而生。Femtocell 主要用于扩大室内覆盖, 增加系统容量, 减少宏网络负担。Femtocell 是一种低功耗设备, 其发射功率低, 可以降低辐射; 使用简单, 即插即用。由于

两层网络中要保证宏网络用户能正常通信, 所以两层网络的干扰控制就显得尤为重要。在无线通信网络中, 发射功率是极为重要的资源, 如果能有效地控制发射功率就可以极大的提高系统容量。功率控制<sup>[2]</sup>的目的就是减少同信道干扰, 在保证系统 QoS 的前提下通过对家庭基站(Femtocell Base Station, FBS)发射功率的调整来降低家庭基站对宏基站(Macrocell Base Station, MBS)的干扰, 保证 Macrocell 用户(Macrocell User, MUE)正常通信, 同时保证 Femtocell 用户(Femtocell User, FUE)可以正常通信。

在 Macrocell 和 Femtocell 构成的两层网络<sup>[3]</sup>中, 下行链路 FBS 的发射功率控制符合马尔科夫决策过程, 适合使用分布式  $Q$  学习功率控制方法。而  $Q$  学习<sup>[4]</sup>是一种无模型算法, FBS 可以通过学习找到最佳的学习策略, 这对于本文要研究的两层网络功率控制来说是很重要的, 比如在文献[5]中使用了加强学习的方法来控制由 Femtocell 引入的对 Macrocell 基站以及用户产生的干扰; 在文献[6]中, 使用  $Q$  学习算法来进行功率控制;  $Q$  学习允许智能体(agent)在学习过程中选择动作, 而不需要集中控制, 这使得  $Q$  学习非常适合有多个 Femtocell 随机分布的系统中, 也就是多智能体的  $Q$  学习系统。在文献[7,8]中分别研究了基于分布式  $Q$  学习的 IQL 和 CQL 在多智能体系统中的应用。在文献[9,10]中应用 IQL 对数字电视系统中存在多个智能体的网络进行了功率控制的研究。在[11]中也将 IQL 应用到了含有 Femtocell 的两层网络的功率控制研究当中。但是这些文献研究中大多只针对保证主用户的 QoS 而忽略了对次用户 QoS 的保证。

在本文中, 结合 Femtocell 与 Macrocell 两层网络下行链路的实际通信模型分析了一种新的基于  $Q$  学习的分布式功率控制算法。在这一算法中, 分别分析并仿真了 IQL 和 CQL 应用于两层网络下行链路功率控制时对系统性能的影响, 并且对 CQL 算法进行了改进, 改进后的 CQL 算法在保证系统容量和功率都满足的情况下回报函数是正值, 其他不符合要求的情况是负值。

## 2. 系统模型和算法分析

### 2.1. 系统模型

本文主要分析 Femtocell 与 Macrocell 组成的两层

网络下行链路中对 FBS 发射功率的控制, 忽略 Femtocell 与 Femtocell 之间的干扰, 同时为描述方便, 假设只有一个 Macrocell 网络, 并且 Macrocell 与 Femtocell 共享频谱资源, 采用正交频分多址接入 (OFDMA), MBS 和 FBS 在同一时隙内在相同的子载波上传输。在 Macrocell 网络中随机分布 4~15 个 Femtocell 网络, 每个 Femtocell 网络中有 1 个 FUE, FUE 采用闭合式接入方式, 即只允许授权用户使用 Femtocell 网络。如图 1 所示:

Macrocell 的传输功率要满足  $\sum_{n=1}^{N_{sub}} P_0^{(n)} \leq P_{max}^m$ , 其中

$N_{sub}$  是子载波数量,  $P_0^{(n)}$  表示 MBS 在子载波  $n$  上的传输功率,  $P_{max}^m$  表示 MBS 的最大传输功率; Femtocell 的传输功率要满足  $\sum_{n=1}^{N_{sub}} P_i^{(n)} \leq P_{max}^f$ , 其中  $P_i^{(n)}$  表示第  $i$  个 FBS 在子载波  $n$  上的传输功率,  $P_{max}^f$  表示 FBS 的最大传输功率。

### 2.2. 分布式 $Q$ 学习算法 IQL 和 CQL

分布式  $Q$  学习的过程可表示为: 1) 认知用户观察当前状态  $s_n$ ; 2) 选择并执行一个动作  $a_n$ ; 3) 观察随后的状态  $v_n$ ; 4) 收到一个即时的回报  $r_n$ ; 5) 通过学习系数  $\alpha$  调整  $Q$  函数值并更新  $Q$  表。

$r_n$  是系统在状态  $s_n$  时执行动作  $a_n$  后得到的回报;  $\alpha$  是系统学习系数,  $0 \leq \alpha < 1$ ;  $Q_n(s, a)$  是折扣回报的期望值, 假设  $Q_0(s, a)$  已经给定,

$$Q_n(s, a) = \begin{cases} \text{if } s = s_n, a = a_n \\ (1 - \alpha)Q_{n-1}(s, a) + \alpha[r_n + \gamma V_{n-1}(v_n)] \\ Q_{n-1}(s, a), & \text{otherwise} \end{cases} \quad (1)$$

其中  $V_{n-1}(v) \equiv \max_b \{Q_{n-1}(v, b)\}$ 。

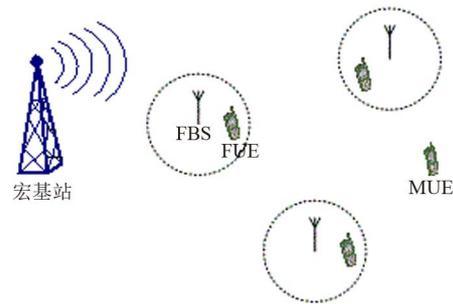


Figure 1. System model  
图 1. 系统模型

在独立学习 IQL 中, 每个 Femtocell 独立的进行学习, 也就是说把其他 Femtocell 当做环境的一部分。对于第  $i$  个 Femtocell 来说, 就如同系统中只有单个 Femtocell 的情况, 其动作的选择为:

$$a_i = \arg \max_{a \in A} Q_i(s_i, a) \quad (2)$$

其中  $A$  是 Femtocell 可选择的动作的集合, 其环境状态集合表示为  $S$ 。在 IQL 中回报函数是与参加学习的所有 Femtocell 的动作选择序列有关, 即其  $Q$  函数为:

$$Q_i(s_i, a_i) = (1-\alpha)Q_i(s_i, a_i) + \alpha(\gamma_i(s_i, \mathbf{a})) + \gamma_{b_i \in A}^{\max} Q_i(v_i, b_i) \quad (3)$$

在多个 Femtocell 的系统中, 独立学习并不是最好的方式, 因为 Femtocell 彼此之间会互相影响, 而 CQL 就可以解决这些问题。

CQL 就是 Femtocell 之间分享环境信息, 每个 Femtocell 都会将自身的当前状态分享给所有的 Femtocell。CQL 算法中第  $i$  个 Femtocell 的动作选择依据为:  $a_i = \arg \max_a \left( \sum_{1 \leq j \leq N} Q_j(s_j, a) \right)$ , 分布式  $Q$  学习算法 IQL 和 CQL 中的智能体、状态、动作以及回报函数的选取如下:

智能体: 智能体为 FBS, 即  $FBS_i, \forall 1 \leq i \leq N$ , 其中  $N$  为两层网络中 Femtocell 的数量。

状态: 对于第  $i$  个在子载波  $n$  上传输的 Femtocell 来说, 在时刻  $t$  定义其状态为:  $s_t^{i,n} = \{I_t^n, P_t^i\}$ , 其中  $I_t^n \in \{0, 1\}$ ,  $0$  表示 Macrocell 系统容量效率大于设定的目标容量效率值  $\tau^0$ ,  $1$  表示小于  $\tau^0$ , 即:

$$I_t^n = \begin{cases} 0, & C_0^{(n)} \geq \tau^0 \\ 1, & C_0^{(n)} < \tau^0 \end{cases} \quad (4)$$

式(4)中  $C_0^{(n)}$  是 MBS 接收到的其用户在子载波  $n$  上传输的容量效率,  $\tau^0$  是在保证 Macrocell 的 QoS 的情况下的目标容量效率。

$P_t^i$  表示第  $i$  个 Femtocell 在时间  $t$  传输的总功率:

$$P_t^i = \begin{cases} 0, & \sum_{n=0}^{N_{sub}} P_t^{i,n} < (P_{max}^f - 15) \\ 1, & (P_{max}^f - 15) \leq \sum_{n=0}^{N_{sub}} P_t^{i,n} \leq P_{max}^f \\ 2, & \sum_{n=0}^{N_{sub}} P_t^{i,n} > P_{max}^f \end{cases} \quad (5)$$

其中  $p_t^{i,n}$  表示在时刻  $t$  在子载波  $n$  上地  $i$  个 Femtocell 的传输功率。

动作: 动作集合就是每个 FBS 的可选动作的集合, 本文中动作集合是 FBS 的发射功率, 其范围为  $-20$  到  $15$  dBm, 其发射功率的间隔步长为  $2$  dBm, 在  $Q$  学习的动作选择中为平衡探索与利用的关系, 引入动作选择系数  $\varepsilon$ , 若  $a$  为当前选择的动作, 则

$$a = \begin{cases} a_i', & \text{rand} \geq \varepsilon \\ \text{随机选择动作}, & \text{rand} < \varepsilon \end{cases}$$

其中 rand 是  $0$  到  $1$  之间的随机数,  $a_i'$  为第  $i$  个 FBS 在时间  $t$  时的动作, 即发射功率。

回报函数 R1: 在时间  $t$ , 在子载波  $n$  上传输的第  $i$  个 Femtocell 的回报函数为:

$$r_t^{i,n} = \begin{cases} e^{-\left(C_0^{(n)} - \tau^0\right)^2}, & \sum_{n=0}^{N_{sub}} P_t^{i,n} \leq P_{max}^f \\ -1, & \text{其他} \end{cases} \quad (6)$$

### 2.3. 改进的分布式 $Q$ 学习算法 IQL 和 CQL

在独立学习 IQL 中, 每个 Femtocell 独立的进行学习, 也就是说把其他 Femtocell 当做环境的一部分。对于第  $i$  个 Femtocell 来说, 就如同系统中只有单个 Femtocell 的情况, 其动作的选择为:

$$a_i = \arg \max_{a \in A} Q_i(s_i, a) \quad (7)$$

其中  $A$  是 Femtocell 可选择的动作的集合, 其环境状态集合表示为  $S$ 。在 IQL 中回报函数是与参加学习的所有 Femtocell 的动作选择序列有关, 即其  $Q$  函数为:

$$Q_i(s_i, a_i) = (1-\alpha)Q_i(s_i, a_i) + \alpha(\gamma_i(s_i, \mathbf{a})) + \gamma_{b_i \in A}^{\max} Q_i(v_i, b_i) \quad (8)$$

在多个 Femtocell 的系统中, 独立学习并不是最好的方式, 因为 Femtocell 彼此之间会互相影响, 而 CQL 就可以解决这些问题。

CQL 就是 Femtocell 之间分享环境信息, 每个 Femtocell 都会将自身的当前状态分享给所有的 Femtocell。本文中对分布式  $Q$  学习算法 IQL 和 CQL 进行了改进, 改进的回报函数为 R2:

$$r_t^{i,n} = \begin{cases} e^{-\left(C_0^{(n)} - \tau^0\right)^2} - e^{-C_t^{(n)}}, & I_t^n = 0, P_t^i = 0 \text{ 或 } P_t^i = 1 \\ -1, & \text{其他} \end{cases} \quad (9)$$

原算法中的回报函数  $R1$  只保证在总的 FBS 发射功率满足系统要求时回报函数是正值，但是并不能保证 Macrocell 系统容量效率也同时满足系统要求，并且，在原算法的回报函数  $R1$  中并不能体现 FBS 系统容量效率在  $Q$  学习过程中的作用。而在改进的 CQL 回报函数  $R2$  中，首先限制在  $I_i^n = 0$  且  $P_i^j = 0$  或者  $P_i^j = 1$  时回报函数为正值，也就是只有系统同时满足 FBS 总的发射功率小于系统要求的 FBS 最大发射功率且 Macrocell 系统容量效率也在设定值以上时回报函数才是正值，其他不满足系统条件要求的一律要给出负回报值；其次引入了总的 Femtocell 容量效率作为回报函数的一个参数，使得在  $Q$  学习中能更好的提高 Femtocell 系统的容量效率以及公平性。

CQL 算法中第  $i$  个 Femtocell 的动作选择依据为：

$$a_i = \arg \max_a \left( \sum_{1 \leq j \leq N} Q_j(s_j, a) \right) \quad (10)$$

此动作选择的依据主要有以下两点：

1)  $Q$  函数即回报函数的期望值，当环境状态处于  $s$  时，选择使得  $Q$  函数最大的动作。例如，在状态  $s$  下，有两个可选动作  $a1, a2$ ，对应应有  $Q(s, a1), Q(s, a2)$ ，假如  $Q(s, a1) > Q(s, a2)$ ，这时 Femtocell 就会选择执行  $a1$ ；

2) 对于总的系统的  $Q$  函数，定义为系统中每个 Femtocell 产生的  $Q$  函数值之和，即：

$Q(s, a) = \sum_{j=1}^N Q_j(s_j, a_j)$ ，由于对于每一个 Femtocell 来说，其动作选择均会选择能使其  $Q$  函数最大的那个动作，因此整个系统的  $Q$  函数值也是最大。

而基于以上分析，若跟据上述依据来选择动作，所有的 Femtocell 都会选择同一个动作，例如，有两个 Femtocell 分别为 Femtocell1 和 Femtocell2，每个 Femtocell 都只有一种环境状态  $s$ ，可选的动作有 3 个，分别为  $a1, a2$  和  $a3$ 。其  $Q$  函数值分别为：

$$Q_1(s, a1) = 2, Q_1(s, a2) = 3, Q_1(s, a3) = 4,$$

$$Q_2(s, a1) = 5, Q_2(s, a2) = 7, Q_2(s, a3) = 5.5,$$

每个 Femtocell 单独选择动作时就相当于 IQL，对于 Femtocell1 来说，比较上述  $Q$  函数值，选择执行动作  $a3$  对应的  $Q$  函数值最大；同理，对于 Femtocell2 来说，选择执行动作  $a2$  对应的  $Q$  函数值最大。但是在 CQL 中，两个 Femtocell 都会选择执行动作  $a2$ ，因为

在合作学习中动作选择的依据是使得所有的 Femtocell 的  $Q$  函数值之和最大，此时

$$Q_1(s, a2) + Q_2(s, a2) = 10 \text{ 是最大的。}$$

### 3. 仿真分析

仿真中用到的系统参量值为表 1：

本文以 Jain 公平系数<sup>[12]</sup>和系统容量来分析系统性能。Jain 系数定义为：

$$f(x_1, x_2, \dots, x_n) = \frac{\left( \sum_{i=1}^n x_i \right)^2}{n \sum_{i=1}^n x_i^2} \quad (11)$$

其中  $0 \leq f(x_1, x_2, \dots, x_n) \leq 1$ 。  $f(x_1, x_2, \dots, x_n)$  越接近 1 表明系统中各个 Femtocell 的容量效率越相似，公平性越好；当  $f(x_1, x_2, \dots, x_n) = 1$  时，表示所有 Femtocell 都达到了相同的容量效率。

系统容量效率即每个子载波上的传输容量，其单位为比特/秒/Hz (bits/s/Hz)。MBS 接收到的 MUE 在子载波  $n$  上的容量效率为：

$$C_0^{(n)} = \log_2 \left( 1 + \frac{h_{00}^{(n)} P_0^{(n)}}{\sum_{i=1}^N h_{i0}^{(n)} P_i^{(n)} + \sigma^2} \right) \quad (12)$$

其中， $h_{00}^{(n)}$  表示 MBS 到 MUE 在子载波  $n$  上的信道增益， $h_{i0}^{(n)}$  表示第  $i$  个 FBS 到 MUE 在子载波  $n$  上的信道增益， $\sigma^2$  表示噪声功率。

第  $i$  个 FBS 接收到的 FUE 在子载波  $n$  上的容量效率为：

Table 1. Parameters of the simulation system [10]

表 1. 系统仿真参数<sup>[10]</sup>

参数	数值	物理意义
$N_{sub}$	6	子载波数目
$P_{max}^f$	15 dBm	Femtocell 最大传输功率
$P_{max}^m$	43 dBm	Macrocell 最大传输功率
$\tau^0$	6 bits/sec/Hz	设定容量效率
$\sigma^2$	$10^{-7}$ Watt	噪声功率
$\gamma$	0.9	折扣系数
$\alpha$	0.5	学习系数
$\varepsilon$	0.1	随机系数

$$C_i^{(n)} = \log_2 \left( 1 + \frac{h_{ii}^{(n)} P_i^{(n)}}{\sum_{j=1, j \neq i}^N h_{ji}^{(n)} P_j^{(n)} + h_{0i}^{(n)} P_0^{(n)} + \sigma^2} \right) \quad (13)$$

其中,  $h_{ii}^{(n)}$  表示第  $i$  个 FBS 到其覆盖的 FUE 在子载波  $n$  上的信道增益,  $h_{ji}^{(n)}$  表示第  $j$  个 FBS 到第  $i$  个 FBS 覆盖的 FUE 在子载波  $n$  上的信道增益。

在子载波  $n$  上发射端与接收端之间的信道增益用其路径损耗来计算, 定义为:  $h_{ij}^{(n)} = d_{ij}^{(-k)}$ , 其中  $d_{ij}$  是发射端  $i$  与接收端  $j$  之间的实际距离,  $k$  是路径损耗系数, 在仿真中取  $k=2$ , MBS、FBS 及其各自覆盖范围内的移动用户随机分布, 各个发射端与接收端之间距离的选取如下: MBS 与其覆盖范围内的 MUE 之间的最大距离为 1000 米; MBS 与 FUE 之间的最大距离为 800 米; FBS 与其所覆盖的 FUE 之间的最大距离为 80 米; FBS 与其他 Femtocell 覆盖范围内的 FUE 之间的最大距离为 300 米; FBS 与 Macrocell 覆盖的 MUE 之间的最大距离为 800 米。

本文中对比分析了分布式  $Q$  学习算法 IQL 和 CQL 算法中采用已有的回报函数  $R1$  和采用改进的回报函数  $R2$  时, Macrocell 的容量效率、总的 Femtocell 容量效率和 Femtocell 公平性随着 IQL 和 CQL 中回报函数的改进而得到的改善。

图 2 为两层网络系统中有 4 个 Femtocell, 仿真中设定的  $\tau^0 = 6$ , 并且 Femtocell 采用回报函数  $R1$  时 IQL 算法和 CQL 算法下 Macrocell 的容量效率, 从图 2 中仿真曲线可以看出两种算法都可以提高 Macrocell 系统的容量效率, 但在收敛状态时, CQL 收敛于 9.3bits/sec/Hz, 相对于 IQL 收敛于 7.2bits/sec/Hz 来说, CQL 算法对于系统容量效率的改善效果更好。

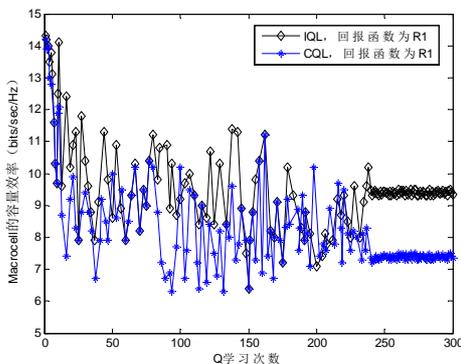


Figure 2. Macrocell capacity  
图 2. Macrocell 的容量效率

图 3 为 IQL 和 CQL 分别采用回报函数  $R1$  和改进的回报函数  $R2$  时, 总的 Femtocell 容量效率随着网络中 FBS 数量的增加而变化的仿真曲线。由图 3 分析可知, 在 IQL 和 CQL 算法中采用改进的回报函数  $R2$  后, 相比原 IQL 和 CQL 采用回报函数  $R1$  时总的 Femtocell 容量效率都得到很大提高, 但是 CQL 在采用改进的回报函数  $R2$  时比 IQL 在采用改进的回报函数  $R2$  时对总的 Femtocell 容量效率的提高更高, 比如在两层网络中有 10 个 Femtocell 时, IQL 在采用改进的回报函数  $R2$  时的总的 Femtocell 容量效率大约为 7.7bits/s/Hz, 而 CQL 在采用改进的回报函数  $R2$  时的总的 Femtocell 容量效率大约为 10bits/s/Hz。

图 4 中是 IQL 与 CQL 在分别采用回报函数  $R1$  和改进的回报函数  $R2$  时系统的 Femtocell 容量效率的公平性, 由图 4 可以看出, 在 IQL 中采用回报函数  $R1$  时系统中 Femtocell 容量效率差别很大, 公平性较差, 在采用改进的回报函数  $R2$  时, Femtocell 容量效率比采用回报函数  $R1$  时略微公平, 但是随着 Femtocell 数量的增加, 其公平性的改善效果变差, 基本与采用回报函数  $R1$  时相同; 在 CQL 中采用改进的回报函数  $R2$  时, 其系统公平性比较好, 即在同一时刻达到相同容量效率的 Femtocell 占系统中总的 Femtocell 数量的比例能基本维持在一个确定的水平值, 从图 4 中可以看出在 0.98 左右。

#### 4. 结束语

本文以基于  $Q$  学习的分布式功率控制算法 IQL 和 CQL 对 Femtocell 和 Macrocell 两层网络中的干扰进行分析, 并对  $Q$  学习算法 IQL 和 CQL 的回报函数

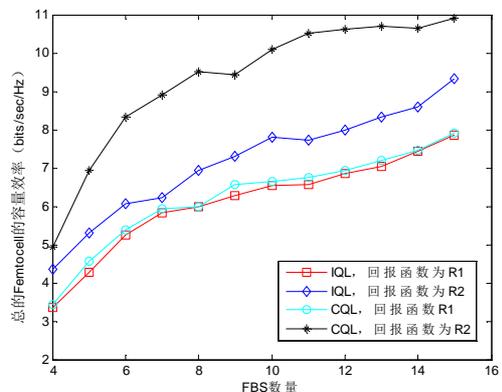


Figure 3. Total femtocell capacity  
图 3. 总的 Femtocell 容量效率

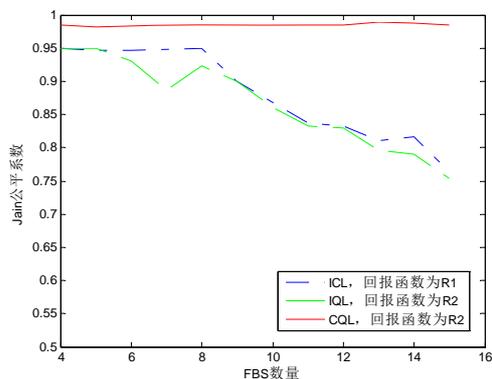


Figure 4. Jain's fairness  
图 4. Jain 公平系数

进行了改进，仿真并验证了原有分布式  $Q$  学习算法 IQL 和 CQL 以及回报函数改进后的分布式  $Q$  学习算法 IQL 和 CQL 在两层网络功率控制中的应用。

### 参考文献 (References)

- [1] J. Zhang, G. de la Roche. Femtocells: Technologies and deployment. Hoboken: John Wiley and Sons, Ltd., 2010.
- [2] S. A. Grandhi, J. Zander. Constrained power control in cellular radio systems. 1994 IEEE 44th Vehicular Technology Conference, 1994, 2: 824-828.
- [3] C. J. C. H. Watkins. Learning from delayed rewards. PhD Thesis, Cambridge: University of Cambridge, 1989.
- [4] S. Mahadevan, J. Connell. Automatic programming of behavior-based robots using reinforcement learning. Proceedings of the 1991 National Conference on AI, 1991: 768-773.
- [5] R. S. Sutton, A. G. Barto. Reinforcement learning: An introduction. Cambridge: MIT Press, 1998.
- [6] C. J. C. H. Watkins, P. Dayan. Technical note  $Q$ -learning. Journal of Machine Learning, 1992, 8(3-4): 279-292.
- [7] J. R. Kok, N. Vlassis. Collaborative multiagent reinforcement learning by payoff propagation. The Journal of Machine Learning Research, 2006, 7: 1789-1828. <http://portal.acm.org/citation.cfm?id=1248547.124861>
- [8] M. Ahmadabadi, M. Asadpour. Expertness based cooperative  $Q$ -learning. Systems, Man, and Cybernetics, Part B: IEEE Transactions on Cybernetics, 2002, 32(1): 66-76.
- [9] A. Galindo-Serrano, L. Giupponi. Distributed  $Q$ -learning for aggregated interference control in cognitive radio networks. IEEE Transactions on Vehicular Technology, 2010, 59(4): 1823-1834.
- [10] A. Galindo, L. Giupponi. Decentralized  $Q$ -learning for aggregated interference control in completely and partially observable cognitive radio networks. 2010 7th IEEE Proceedings of the Consumer Communications and Networking Conference (CCNC), Las Vegas, 2010: 1-6.
- [11] A. Galindo-Serrano, L. Giupponi and M. Dohler. Cognition and docation in OFDMA-based Femtocell networks. 2010 IEEE Global Telecommunications Conference, Miami, 2010: 1-6.
- [12] R. Jain, D.-M. Chiu and W. Hawe. A quantitative measure of fairness and discrimination for resource allocation in shared computer systems. Maynard: Eastern Research Laboratory, Digital Equipment Corporation, 1998.