

多用户 MIMO-OFDMA 系统混合业务的跨层分组调度算法

雷海鹏^{①②} 李小强^② 王亚峰^① 杨大成^①

^①(北京邮电大学电信工程学院 北京 100876)

^②(北京三星通信技术研究有限公司 北京 100125)

摘要: 该文提出了一种针对多输入多输出-正交频分多址系统混合业务的跨层分组调度算法。该算法基于物理层的信道状况和媒体接入控制子层用户的排队时延, 分别为实时业务用户和非实时业务用户的排队时延构建相应的效用函数, 以保障不同业务的服务需求。仿真结果表明: 该算法不但能保证实时业务的时延需要, 而且能够获得很好的吞吐量, 并满足公平性要求。

关键词: 多输入多输出; 正交频分多址; 分组调度; 效用函数

中图分类号: TN929.5

文献标识码: A

文章编号: 1009-5896(2008)07-1535-05

A Cross-Layer Packet Scheduling Algorithm for Mixed Services in Multi-user MIMO-OFDMA Systems

Lei Hai-peng^{①②} Li Xiao-qiang^② Wang Ya-feng^① Yang Da-cheng^①

^①(School of Telecom Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China)

^②(Beijing Samsung Telecom R & D Center, Beijing 100125, China)

Abstract: A cross-layer packet scheduling algorithm is proposed for multi-user MIMO-OFDMA systems with the mixed services, such as Real-Time (RT) and Non-Real-Time (NRT) services. The proposed scheduling algorithm is based on the channel state information of physical layer and user queuing delay of Medium Access Control (MAC) sub-layer, and constructs the corresponding utility functions for different types of services. Simulation results show that this algorithm can not only guarantee the delay requirements of RT service but also obtain good system throughput and satisfy the fairness criterion.

Key words: MIMO; OFDMA; Packet scheduling; Utility function

1 引言

多输入多输出-正交频分多址(MIMO-OFDMA)是3GPP Long Term Evolution (LTE)的核心技术之一。MIMO技术在不增加带宽的条件下可以成倍地提高系统容量和频谱效率, OFDMA基于OFDM,继承了OFDM抗符号间干扰和频率选择性衰落的特性。和时分复用、频分复用相比, MIMO-OFDMA系统能提供空-时-频三维资源给用户。

无线衰落信道上的分组调度是一个很重要的研究方向。现有的支持无线系统数据业务的分组调度算法有很多,其中具有代表性的有如下几种: Proportional Fair (PF)算法可以最大化用户的长期吞吐量,并可以在一定程度上满足用户之间的公平性^[1]。然而,该算法并没有考虑实时业务数据包的时延,因此不支持时延敏感型业务。Modified-Largest Weighted Delay First(M-LWDF)算法充分考虑了信道条件和排队状况,可以满足实时业务用户的时延需要,针对每种类型业务,为其分配相应的权值^[2]。EXPonential Proportional Fair (EXP/PF)算法也可以满足实时业务用

户的时延需要,其算法结构类似于M-LWDF^[3]。这两种算法主要解决的是单载波CDMA系统中时间资源分配的问题,且为了保障实时业务用户的服务质量(QoS),牺牲了非实时业务,所以公平性较差。而且,在MIMO-OFDMA系统中,空-时-频三维资源特性使得调度更加复杂。

针对包含实时和非实时业务的MIMO-OFDMA系统,本文提出了一种基于物理层信道状况和媒体接入控制(MAC)子层用户排队时延的跨层分组调度算法。该算法为不同类型业务的排队时延构建效用函数,能满足不同业务的时延需要。最后,基于LTE系统级仿真,评估了所提调度算法的性能。

2 MIMO-OFDMA 系统模型

如图1所示,本文考虑了一个具有 N_T 根发射天线, N_R 根接收天线和 M 个子载波的MIMO-OFDMA下行分组数据蜂窝系统。在满足 $N_R \geq N_T$ 的条件下,采用Bell-Labs Layered Space-Time(BLAST)技术可以获得很高的数据传输速率。

在基站端,基站为每个用户建立了一个数据队列。到达

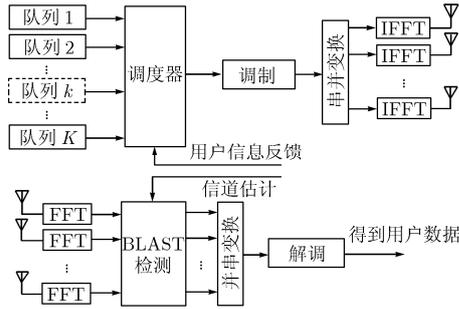


图1 下行链路基站发送端和用户接收端框图

的每个用户的数据包以先入先出(FIFO)的方式存放在相应的数据队列中。在每一时隙，用户都要测量其在各个子载波上的信道状况，通过专用上行反馈信道报告给基站；在另一方面，用户信息，如队列长度、QoS要求、排队时延，也被送入到基站调度器去竞争系统资源。根据调度算法，基站将系统资源分配给优先级最高的用户。然后，将用户的信息比特首先进行编码调制，然后分成 N_T 路数据流后进行IFFT变换，再由 N_T 根发射天线同时发射出去。在接收端，通过FFT和BLAST检测，解调出用户数据。

将基站的发射功率平均分配在所有的子载波和发射天线上，从而，接收信号可以表示为

$$\mathbf{y}_{k,m} = \sqrt{\frac{P}{N_T M}} \mathbf{H}_{k,m} \mathbf{X}_{k,m} + \mathbf{n}_{k,m} \quad (1)$$

其中 $\mathbf{y}_{k,m} = (y_{k,m,1}, \dots, y_{k,m,N_R})^T$ 表示用户 k 在子载波 m 上的接收矢量； P 为基站端总的发射功率； $\mathbf{H}_{k,m} = \begin{pmatrix} h_{k,m}^{1,1} & \dots & h_{k,m}^{1,N_T} \\ \vdots & \ddots & \vdots \\ h_{k,m}^{N_R,1} & \dots & h_{k,m}^{N_R,N_T} \end{pmatrix}$ 是一个 $N_R \times N_T$ 维的复矩阵，其元素

$h_{k,m}^{i,j}$ 表示用户 k 在子载波 m 上从第 j 个发射天线到第 i 个接收天线的信道冲激响应，该矩阵可以通过信道估计在接收端获得； $\mathbf{X}_{k,m}$ 是一个 $N_R \times 1$ 维的列向量，表示用户 k 在子载波 m 上的发送信号向量； $\mathbf{n}_{k,m}$ 是一个 $N_R \times 1$ 维的高斯白噪声列向量，其服从均值为0方差矩阵为 $\sigma_{k,m}^2 \mathbf{I}$ 的高斯分布。

在MIMO-OFDMA系统中，设 B 为OFDM子载波带宽，所以，采用MIMO空间复用后，用户 k 在子载波 m 上可获得的瞬时MIMO信道容量 $r_{k,m}$ 可以表示为

$$r_{k,m} = B \log_2 \left(\det \left(\mathbf{I}_{N_R} + \frac{\rho_{k,m}}{N_T} \mathbf{H}_{k,m} \mathbf{H}_{k,m}^H \right) \right) \quad (2)$$

其中 $\rho_{k,m}$ 是用户 k 在子载波 m 上的信噪比^[4,5]。

3 基于效用函数的调度算法

无线通信系统是资源受限的，如何利用有限的资源满足日益增长的用户需求，已经成为亟需解决的问题。在无线资源分配中，需要在吞吐量、公平性和QoS之间达到有效的折中。

本文通过使用微观经济学中的效用函数理论来评估用户对无线通信系统的满意程度。效用函数的主要思想是将系统资源(如带宽、功率等)或者性能指标(如数据速率、时延等)映射成相应的效用值，以用来优化整个系统。代表了用户对系统的满意度，效用函数在无线资源管理以及保障用户QoS要求方面起着非常关键的作用。

3.1 优化目标

本文用 $U_k(d_k)$ 来表示用户 k 的效用函数，即用 $U_k(d_k)$ 代表用户 k 对无线系统的满意度， d_k 为用户 k 的排队时延。因为随着时延的增加，用户的满意度下降，该函数应该是递减的；而且，对于特定的最大允许时延，越靠近该时延限，用户的满意度下降越快，而远离该时延的地方，用户并不怎么太在意。所以，它的导数也是递减的，且为负值。所以 $U_k(d_k)$ 应该是一个非负的单调递减的凸函数。本文的优化目标就是最大化总的效用函数值，也就是说要让所有用户的满意度之和最大。

本文给出如下的优化目标函数：

$$\max \sum_{k=1}^K U_k(d_k) \quad (3)$$

$$\text{over } d_k \geq 0, 1 \leq k \leq K \quad (4)$$

其中 K 是小区中激活用户的数目。

3.2 排队时延

用 $d_k(n)$ 表示用户 k 在第 n 个时隙的排队时延，用 $r_{k,m}(n)$ 表示用户 k 第 n 个时隙在子载波 m 上可获得的瞬时MIMO信道容量。 $r_k(n)$ 和 $R_k(n)$ 分别表示用户 k 在第 n 个时隙可获得的总的瞬时速率和短(short-term)平均速率。所以， $r_k(n)$ 可表示为用户 k 第 n 个时隙在所占子载波上瞬时速率之和。

$$r_k(n) = \sum_{m=1}^M r_{k,m}(n) \delta_{k,m}(n) \quad (5)$$

其中 $\delta_{k,m}(n)$ 是一个表示子载波分配情况的变量，如果第 m 个子载波在第 n 个时隙分配给了用户 k ，则 $\delta_{k,m}(n) = 1$ ；反之，则 $\delta_{k,m}(n) = 0$ 。为了避免小区内干扰，在每个时隙，每个子载波对应的MIMO信道只能唯一地分配给一个用户，所以有如下等式成立， $\sum_{k=1}^K \delta_{k,m}(n) = 1$ 。如果用户 k 在第 n 个时隙没有分到任何子载波，则 $r_k(n) = 0$ 。

用 $a_k(n)$ 表示第 n 个时隙内到达的用户 k 数据包的数据量，假设时隙长度为 T_s ，则用户 k 排队时延的递归方程可以表示成^[6]

$$d_k(n+1) = d_k(n) + \frac{a_k(n) - r_k(n)T_s}{R_k(n)} \quad (6)$$

在每个时隙的末尾更新所有用户的短平均速率，

$$R_k(n+1) = \left(1 - \frac{1}{T_c} \right) R_k(n) + \frac{1}{T_c} r_k(n) \quad (7)$$

其中 T_c 是一个较大的时间窗口，和PF算法中的 T 含义相同^[1]。选择合适的 T_c ，可以达到吞吐量和公平性之间的折中。

在本文仿真中,取 $T_c = 1000T_s$ [1]。

因为每个时隙的长度足够短,可做如下推导:

$$\begin{aligned} & \sum_{k=1}^K U_k(d_k(n+1)) - \sum_{k=1}^K U_k(d_k(n)) \\ & \approx \sum_{k=1}^K \frac{\partial U_k}{\partial d_k} \Big|_{d_k=d_k(n)} (d_k(n+1) - d_k(n)) \\ & = \sum_{k=1}^K \frac{\partial U_k(d_k(n))}{\partial d_k(n)} \frac{a_k(n) - r_k(n)T_s}{R_k(n)} \\ & = \sum_{k=1}^K \frac{\partial U_k(d_k(n))}{\partial d_k(n)} \frac{a_k(n)}{R_k(n)} - \sum_{k=1}^K \frac{\partial U_k(d_k(n))}{\partial d_k(n)} \frac{r_k(n)T_s}{R_k(n)} \quad (8) \end{aligned}$$

考虑到在第 n 个时隙, $d_k(n)$ 是已知的, 且是定值, 所以式(3)的优化目标可以转为式(8)。因为 $U_k(d_k)$ 单调递减的, 所以式(8)可以写为

$$\begin{aligned} & \sum_{k=1}^K U_k(d_k(n+1)) - \sum_{k=1}^K U_k(d_k(n)) \\ & = \sum_{k=1}^K \frac{\partial U_k(d_k(n))}{\partial d_k(n)} \frac{a_k(n)}{R_k(n)} + \sum_{k=1}^K \left| \frac{\partial U_k(d_k(n))}{\partial d_k(n)} \right| \frac{r_k(n)T_s}{R_k(n)} \quad (9) \end{aligned}$$

在第 n 个时隙, 式(9)的第 1 项是已知的, 故优化目标函数可以进一步简写为

$$\max \sum_{k=1}^K \left| \frac{\partial U_k}{\partial d_k} \right| \frac{r_k(n)}{R_k(n)} \quad (10)$$

对式(10)做线性优化, 设 $D_k^*(n)$ 是第 n 个时隙调度器分配给用户 k 的最优的子载波集合, 则当 $\forall m^* \in D_k^*(n)$ 时有如下结论成立 [7]

$$\left| \frac{\partial U_k}{\partial d_k} \right| \frac{r_{k,m^*}(n)}{R_k(n)} \geq \left| \frac{\partial U_h}{\partial d_h} \right| \frac{r_{h,m^*}(n)}{R_h(n)} \quad (11)$$

其中 h 是用户集合中除 k 之外的任一用户, 即 h 满足条件 $1 \leq h \leq K, h \neq k$ 。所以, 最优的子载波调度算法可以表示为下式

$$\hat{k} = \arg \max_{1 \leq k \leq K} \left| \frac{\partial U_k}{\partial d_k} \right| \frac{r_{k,m}(n)}{R_k(n)} \quad (12)$$

即在每个时隙, 能最大化目标函数的用户将获得该子载波。

3.3 效用函数的选择

如 3.1 节中所述, 以用户排队时延为指标的效用函数应该是一个严格单调的非负的减函数, 且上凸。以 s 型函数 $f(t) = 1 - (1 + e^{-\alpha(t-\beta)})^{-1}, \alpha > 0, \beta > 0$ 为例, 作为实时业务用户的效用函数。其中, α 和 β 分别决定函数的斜率和坡度中点的位置。一般地, β 的取值为实时业务用户的 QoS 时延, α 越大, 函数的坡度就越陡峭, 函数在坡度中点的导数的绝对值就越大。一个 s 型函数如图 2 所示。

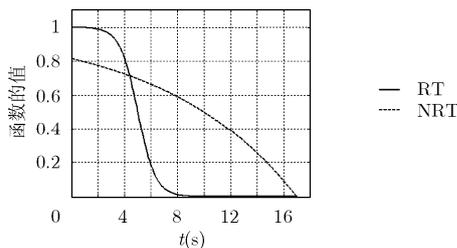


图 2 实时业务和非实时业务随用户排队时延变化的效用函数曲线

为实时业务用户选择 s 型函数, 主要有以下几个方面的考虑: (1) 在 s 型函数出现下降趋势前, 一般都比较平坦, 所以其导数绝对值较小, 那么在这段时间, 非实时业务用户可以优先获得调度; (2) 当实时业务用户排队时延接近 QoS 限时, 此时 s 型函数开始下降, 下降的剧烈程度则由 α 决定, α 越大, 下降越剧烈, 其导数的绝对值也就越大, 使实时业务用户获得更高的调度权值, 以减小实时业务用户的丢包率; (3) 由于 s 型函数的特点, 当用户时延超过 QoS 限时, 效用值并不是立即下降到 0, 而是有一个逐渐下降的过程, 这样可以使实时业务用户继续参与资源的竞争, 以避免因为时延超过 QoS 限而直接丢弃数据包, 从而能够进一步减小用户的丢包率。

对非实时业务用户来说, 他们对时延的要求不像实时业务用户那么严格, 但是长时间接受不到服务的话也会导致非实时业务用户满意度下降。所以, 非实时业务的时延一般为“软时延”, 随着时延的增大, 其效用函数值也会逐渐下降。因此, 本文以函数 $f(t) = 1 - ce^{a(t-b)}, a > 0, b > 0, c > 0$ 作为非实时业务用户的效用函数。其中, a, c 和 b 分别决定函数的斜率、幅度和时延要求。图 2 所示为其一例。

3.4 调度算法

综上所述, 本文提出一种针对混合业务的分组调度算法, 其描述如下:

$$\hat{k} = \arg \max_{1 \leq k \leq K} W_{k,m}(n) \quad (13)$$

其中

$$W_{k,m}(n) = \begin{cases} \left| \frac{\partial U_k^{(1)}}{\partial d_k} \right| \frac{r_{k,m}(n)}{R_k(n)} w, & k \in \text{RT} \\ \left| \frac{\partial U_k^{(2)}}{\partial d_k} \right| \frac{r_{k,m}(n)}{R_k(n)}, & k \in \text{NRT} \end{cases} \quad (14)$$

w 是一个加权因子, 可根据实际需要调节实时业务用户的优先级, $U_k^{(1)}(d_k)$ 和 $U_k^{(2)}(d_k)$ 分别为实时业务和非实时业务用户关于排队时延的效用函数, 可根据实际的业务需求改变参数数值或者选择不同的效用函数。

根据式(13), 每个用户在子载波 m 上的优先级都可以计算出来, 从中选出具有最大优先级的用户 \hat{k} , 将子载波 m 分配给用户 \hat{k} 。然后, 根据相同的调度算法, 再分配下一个子载波, 直到所有的子载波分配完毕为止。分配到子载波的用户就可以在其对应的 MIMO 信道上传输数据比特。

3.5 公平性准则

在多用户系统中, 除了资源利用率外, 还需要考虑用户之间的公平性。调度算法的公平性一般以公平性准则来衡量 [8], 该准则用各用户归一化吞吐量的累积分布函数 (CDF, Cumulative Distribution Function) 曲线来表示。吞吐量一般用单位时间内正确传输的数据量来表示, 将所有用户的吞吐量相对于平均吞吐量做归一化。如果用户 k 的吞吐量为 $T_{\text{put}}(k)$, 相对于所有用户平均吞吐量的归一化吞吐量

$\tilde{T}_{\text{put}}(k)$ 为

$$\tilde{T}_{\text{put}}(k) = \frac{T_{\text{put}}(k)}{\text{avg}_{1 \leq k' \leq K}(T_{\text{put}}(k'))} \quad (15)$$

公平性准则由表1的3个点表示。该准则实质上是限制了低吞吐量用户占总用户数的比例，比如低于0.1倍平均吞吐量的用户数不能超过总用户数的10%。按照该准则，所有满足公平性要求的调度算法，其CDF曲线一定在这3点连成的直线的右侧，否则就是违反了公平性准则^[8]。

表1 公平性准则

相对于平均吞吐量的归一化吞吐量	CDF
0.1	0.1
0.2	0.2
0.5	0.5

4 系统仿真及结果

在LTE系统仿真中，采用了快照加时间驱动的仿真方法，用户在小区中均匀分布，基站采用全向发射天线。每个用户都受到快衰落和大尺度衰落的影响。仿真采用了常见的4种信道，如表2所示。在每一时隙，调度器参考用户的排队时延和上报的信噪比，计算出优先级，并将系统资源分配给具有最高优先级的用户。仿真环境中可以产生两种业务，实时视频和FTP。其中，实时视频业务用影片中的一段视频数据来模拟，非实时业务则采用满buffer FTP业务。LTE仿真平台用C++语言编写，并在相应的开发环境下编译运行通过。LTE系统级仿真的主要参数如表3所示^[9]。

表2 仿真中所使用的信道模型

信道模型	移动速率(km/h)
信道1	3
信道2	3
信道3	30
信道4	120

图3为4种算法的吞吐量性能比较。本文提出的使用效用函数的调度算法(即图3中的Utility)表现出了较高的吞吐量性能。这是因为所提算法同时考虑了物理层信道状况和MAC层排队时延，当实时业务远离时延限时赋予非实时业务更大的优先权，而接近时延限的实时业务数据则权值更大一些。由于所提算法可充分利用多用户分集增益且可使实时业务丢包率维持在较低水平，所以其具有较高的吞吐量，这是因为被丢弃的数据包是无益于系统吞吐量的。

如图4所示，所提的算法在丢包率指标上要优于其它3种算法，这是因为使用了基于排队时延的效用函数之后，充分考虑到了实时业务用户对时延的要求，当用户排队时延接近时延限时，根据式(13)，急剧下降的效用值有利于该用户

表3 系统级仿真参数

载波频率	2 GHz	基站发射功率	43dBm
系统带宽	5 MHz	空时衰落模型	SCME
采样频率	7.68 MHz	MIMO天线配置	2发2收
相邻子载波间隔	15 kHz	穿透损耗	20dB
FFT点数	512	HARQ	Chase combining
可用的子载波数	300	最大队列延迟	50ms
OFDM符号长度	71.43μs	帧速率	25frame/s
时隙长度	0.5ms	平均视频数据速率	53.352kbps
每小区用户数	20	仿真时长	100s
用户分布	均匀分布	α, β	1.5, 50ms
业务分布	50% RT Video + 50% NRT FTP	a, b	0.1, 150ms,
小区半径	1000m	c	0.5
路径损耗模型	$PL = 128.1 + 37.6 \lg(R), R \text{ in km}$	w	1

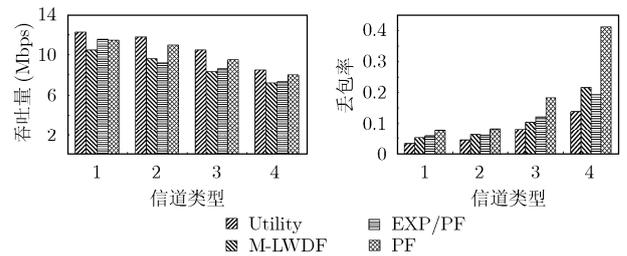


图3 4种算法分别在4种信道上的吞吐量性能

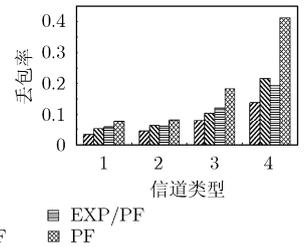


图4 4种算法分别在4种信道上的丢包率

优先得到调度，从而尽量避免产生丢包。即使用户时延超过时延限，其效用值并不是立即下降到0，而是有一个逐渐下降的过程，这样可以使实时业务用户继续参与资源的竞争，以避免因为时延超过门限而直接丢弃数据包，从而能够进一步减小用户的丢包率。从图4可以看出，使用了效用函数的调度算法丢包率要明显低于其它算法，这也有助于其获得较高的吞吐量。

实时业务用户对时延要求苛刻，所以应优先保障实时业务的调度，以期降低丢包率，从而导致非实时业务用户平均速率下降，影响整体的公平性，但只要满足公平性准则，也就是在公平性准则曲线的右侧即可。图5为4种算法在第1种信道下的公平性曲线，虚线即为公平性准则曲线。可以看出，所提算法公平性仅比PF算法稍差一些，但比其余两种算法要好。这是因为PF算法在调度中并不区分业务类型，所以获得了很好的公平性。在保障实时业务时延需要的同

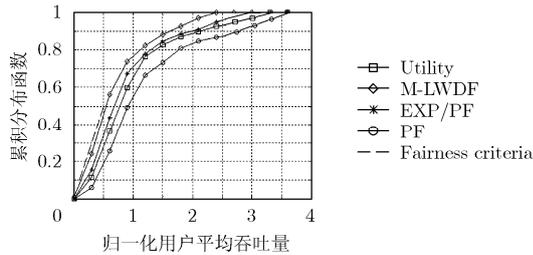


图 5 4 种算法的归一化吞吐量 CDF 曲线

时,所提算法也充分考虑了非实时业务对时延的需要,并为非实时业务设计了时延效用函数,所以使得算法也兼顾了非实时业务,满足公平性要求。对其它几种信道模型,也能得到相同的结论。

5 结束语

本文提出了应用在 MIMO-OFDMA 系统混合业务中的一种跨层分组调度算法。该算法基于物理层信道状况和 MAC 层用户排队时延。在该算法中,根据实时业务和非实时业务对时延的敏感程度不同,分别为其设计了效用函数,并在 LTE 系统中做了仿真。仿真结果表明:该算法不但在保证实时业务用户时延要求方面好于 PF, M-LWDF 和 EXP 等算法,还能获得很好的吞吐量,并能满足公平性要求。

参考文献

- [1] Kaneko M, Popovski P, and Dahl J. Proportional fairness in multi-carrier system: upper bound and approximation algorithms [J]. *IEEE Communications Letters*, 2006, 10(6): 462-464.
- [2] Andrews M, Kumaran K, and Ramanan K, *et al.* Providing quality of service over a shared wireless link [J]. *IEEE Communications Magazine*, 2001, 39(2): 150-154.

- [3] Rhee Jong-Hun, Holtzman J M, and Kim Dong-Ku. Performance analysis of the adaptive EXP/PF channel scheduler in an AMC/TDM system [J]. *IEEE Communications Letters*, 2004, 8(8): 497-499.
- [4] 盛延敏, 奚宏生, 王子磊, 等. 一种用于 MIMO 系统的快速天线选择算法[J]. *电子与信息学报*, 2006, 28(9): 1703-1705. Sheng Yan-min, Xi Hong-sheng, and Wang Zi-lei, *et al.* A fast antenna selection algorithm in MIMO systems [J]. *Journal of Electronics & Information Technology*, 2006, 28(9): 1703-1705.
- [5] Kang Ming and Alouini M-S. Capacity of correlated MIMO rayleigh channels [J]. *IEEE Trans. on Wireless Communications*, 2006, 5(1): 143-155.
- [6] Hosein P. Scheduling of VoIP traffic over a time-shared wireless packet data channel [C]. ICPWC 2005, India, Jan. 2005: 38-41.
- [7] Song Guocong and Li Ye(Geoffrey). Utility-based joint physical-MAC layer optimization in OFDM[C]. *GlobeCom 2002*, Taiwan, Nov. 2002: 671-675.
- [8] 3gpp2. C.R1002-0. CDMA2000 evaluation methodology Revision 0. <http://www.3gpp2.org>, Dec, 2004.
- [9] 3GPP LTE RAN1, TR25.814, Physical Layer Aspects for Evolved UTRA (Release 7) [S], Sep. 2006.

- 雷海鹏: 男, 1980 年生, 博士生, 研究方向为 MIMO-OFDMA 系统级仿真及无线资源管理。
- 李小强: 男, 1971 年生, 首席研究员, 研究方向为 3GPP/SAE 技术研究, HSDPA/HSUPA TDD 技术等。
- 王亚峰: 男, 1974 年生, 副教授, 硕士生导师, 研究方向为移动通信系统无线资源管理及协同中继。
- 杨大成: 男, 1951 年生, 教授, 博士生导师, 研究方向为无线通信理论与关键技术等。