

## 摘要

无线分组调度对无线网络中的业务进行管理和调度, 决定何时发起传输及传输速率, 在保证业务的服务质量方面起了重要的作用。现代无线网络中, 各种新的业务类型相继出现, 同时物理层单载波和多载波的传输技术也在不断发展, 这些都对无线分组调度技术提出了新的挑战。

本文研究单载波和多载波系统中, 针对尽力而为业务和非实时业务的无线分组调度技术, 提出了几种无线分组调度算法。

在单载波系统中, 针对尽力而为业务, 提出渐近吞吐量最大化算法, 采用一种以线性函数为渐近线的效益函数作为算法的优化目标, 在满足 3GPP2 的公平性准则的前提下, 克服了比例公平算法造成较大吞吐量损失的缺陷。

在单载波系统中, 针对非实时业务, 提出 S-效益最优化算法, 将效益函数写成 S 函数与对数函数相乘的形式, 解决了效益函数在最小速率限制处的定义和斜率大小问题。

在多载波系统中, 针对尽力而为业务, 提出改进多载波比例公平调度算法, 采用全搜索的方法寻找用户在各子载波上最大可达速率与该用户已达速率比值的最大值, 提高了系统效益与系统吞吐量。

在多载波系统中, 针对非实时业务, 提出多载波有最小/最大速率限制时的比例公平调度算法, 引入子载波上最小/最大速率限制的概念, 并给每个用户一个令牌计数器, 用令牌数反映速率限制的要求, 用子信道状况和已达速率的比值反映比例公平的要求, 并在每个子载波被分配后都更新各用户令牌数。仿真结果表明, 该算法在容纳用户数、系统效益等方面性能优于现有其他常用算法, 并且能与传统的无速率限制的比例公平调度完全兼容。

在多载波系统中, 针对非实时业务, 提出多载波有最小/最大速率限制时的最大化吞吐量调度算法, 用令牌反映速率限制的要求, 用子信道状况反映最大化吞吐量的要求。仿真结果验证了该算法的有效性。

关键字: 分组调度, QoS, 单载波, 多载波

# ABSTRACT

Wireless packet scheduling manages services in wireless networks, schedules the transmission duration and determines the transmission rate. It plays a key role in QoS provisioning. In modern wireless networks, new service types are constantly emerging, and physical layer transmission techniques (both single-carrier and multi-carrier) are constantly evolving, all these raise new challenges to wireless packet scheduling techniques.

This thesis focuses on wireless packet scheduling for single-carrier and multi-carrier systems. Several new wireless packet scheduling algorithms are proposed for Best-Effort (BE) services and Non-Real-Time (NRT) services.

For BE services in single-carrier systems, we proposed an Asymptotic Throughput Maximization Algorithm (ATMA). This algorithm employs a new utility function which is asymptotic to linear function. It can solve the large throughput loss problem in the traditional proportional fair scheduling, without violating the 3GPP2 fairness criteria.

For NRT services in single-carrier systems, we proposed an S-Utility Optimization (S-UO) algorithm, which employs a utility function in the form of multiplication of S-function and logarithm function. It gives the utility function a proper definition and a slope near the minimum rate constraint point.

For BE services in multi-carrier systems, we proposed a Modified Multi-Carrier Proportional Fairness (Modified-MC-PF) scheduling algorithm. This algorithm performs a full-search to find the maximum ratio of each user's maximum supportable rate to his currently achieved rate. It can increase the system utility and system throughput.

For NRT services in multi-carrier systems, we proposed a Multi-Carrier Proportional Fairness with Minimum/Maximum Rate constraints (MC-PFMR) scheduling algorithm. This algorithm introduces the concept of minimum/maximum rate constraints on single sub-carrier and assigns each user a token counter. The token counter is the key mechanism to guarantee rate constraints and the ratio of sub-carrier channel gain to user's achieved rate is employed to guarantee proportional fairness. Simulation results show that this algorithm outperforms other existing algorithms in terms of maximum allowable number of users, system utility and system throughput. Besides, it is compatible with traditional proportional fair scheduling without rate constraints.

For NRT services in multi-carrier systems, we also proposed a Multi-Carrier Maximum Throughput with Minimum/Maximum Rate constraints (MC-MTMR) scheduling algorithm. The token counter is the key mechanism to guarantee rate constraints and the sub-carrier channel gain is used to guarantee maximum throughput. Simulation results verify the effectiveness of the algorithm.

Keywords: packet scheduling, QoS, single-carrier, multi-carrier

# 第一章 绪论

移动通信系统是指建立交换信息的两端中至少有一方可以处于移动状态的网络系统<sup>[1]</sup>，它是当今通信领域内最为活跃和发展最为迅速的领域之一，对人们的生活方式产生了巨大的影响。本章首先回顾了移动通信的发展历程，然后介绍了本文涉及的几个重要概念，最后介绍了本文的主要贡献和结构安排。

## 1.1 移动通信的发展

六十年代，Bell 实验室提出了蜂窝通信的概念，使得移动通信技术在理论上开始了大发展。随着射频电路技术、数字通信技术、大规模集成电路技术的发展，移动通信逐渐成熟起来，并给人们的生活带来了极大的影响。模拟蜂窝移动通信技术的发明是移动通信的里程碑，它采用频分多址方式，被称为第一代移动通信。八十年代末，随着以大规模集成电路为基础的数字信号处理技术和数字通信技术的发展和应用，产生了以数字语音编码，信道编码和时分多址(Time Division Multiple Access, TDMA)等新技术为基础的第二代移动通信。以欧洲的全球通信系统(Globe System for Mobile Communications, GSM)为例，它采用 GMSK 数字调制方式，采用语音压缩方法和 TDMA 接入方式，使得每个语音的信道占用带宽降为模拟系统的四分之一左右，并且由于超大规模集成电路技术的应用，移动终端的体积和功耗大大降低。现在大规模普及的就是第二代移动通信产品。

为满足迅速增长的移动用户群和多样化的移动业务需求，90 年代国际电联提出了第三代移动通信(3<sup>rd</sup> Generation, 3G)系统的概念，主要以码分多址<sup>[2]</sup>(Code Division Multiple Access, CDMA)为核心技术，提供高质量、高保密性、低成本的话音、分组数据和多媒体业务服务。在第三代移动通信标准的制定过程中，由于经济、政治等多种因素存在，最终未完成统一的标准，而是多种制式共存的局面，包括欧洲的 W-CDMA、美国的 cdma2000 和我国的 TD-SCDMA 等制式。

由于 3G 系统的核心网还没有完全脱离第二代移动通信系统的核心网结构，所以普遍认为 3G 系统仅仅是一个从窄带向未来移动通信系统过渡的阶段。目前，人们已经把目光越来越多地投向 B3G(Beyond 3G)的移动通信系统，该系统可以容纳庞大的用户数，改善现有通信质量，达到高速数据传输的要求。从技术层面来看，3G 系统主要是以 CDMA 为核心技术，而在 3G 以后的移动通信系统中，正交频分复用<sup>[3]</sup>(Orthogonal Frequency Division Multiplexing, OFDM)技术最受瞩目。

OFDM 已有近 40 年的历史, 第一个实际应用是军用的无线高频通信链路。经过多年的发展, 该技术在通信的各个领域都得到广泛的应用, 例如非对称数字用户环路(Asymmetric Digital Subscriber Line, ADSL), 高清晰度电视(High Definition Television, HDTV), 无线局域网(Wireless Local Area Network, WLAN)等。OFDM 技术可以有效地消除信号多径传播所造成的符号间干扰。OFDM 还易于结合空时编码、分集、干扰抑制以及智能天线等技术, 极大地提高了物理层信息传输的可靠性。如果再结合自适应调制、自适应编码以及动态子载波分配、动态比特分配等技术, 其性能可以进一步得到提高。

基于 OFDM 技术, 存在多种多址方式, 其中包括正交频分多址(Orthogonal Frequency Multiple Access, OFDMA), OFDM-TDMA 以及 OFDM-FDMA 等等。其中 OFDMA 系统因为实现了更小的资源分配粒度, 从而利用不同用户在不同子载波上信道响应各不相同的特点实现多用户分集(Multi-user Diversity), 近年来得到了广泛的研究。采用 OFDMA 方式的最重要的通信协议就是 IEEE802.16 系列协议<sup>[4]</sup>, 又称为无线城域网(Wireless Metropolitan Area Network, WMAN)空中接口标准, 它给出了 2-66GHz 频率范围内无线接入系统的底层标准(物理层和媒体接入控制层), 给出了系统的设计、配置和频率使用的解决方案。IEEE802.16 系列协议的物理层有几种方案, 包括单载波方案(又称为 WMAN SCa), OFDM+TDMA 方案(又称为 WMAN-OFDM)和 OFDMA 方案(又称为 WMAN-OFDMA)等, 目前研究最多的是 IEEE802.16e OFDMA 物理层协议。

## 1.2 QoS 的基本概念

QoS(Quality of Service, 服务质量)是网络在传输数据流时要求满足的一系列服务请求<sup>[5]</sup>, 具体可以量化为最小/最大速率限制, 最大允许时延, 时延抖动等性能指标。根据对时延要求的不同, 网络中的业务类型可以分为实时(Real-Time, RT)业务和非实时(Non-Real-Time, NRT)业务两大类。实时业务要求服务时延小于某个设定参数, 常见实时业务包括语音、流媒体等, 最主要的 QoS 参数为最大允许时延、最小/最大速率限制等。因为实时业务对时延的要求较高, 通常又被称为时延敏感型业务(delay-sensitive service)。非实时业务在时延上并没有严格的要求, 但是常常要求传输速率满足一定的限制要求, 常见的非实时业务包括 HTTP, FTP 等, 最主要的 QoS 参数为最小/最大速率限制。因为非实时业务对速率的要求较高, 通常又被称为速率敏感型业务(rate-sensitive service)。

RT 业务和 NRT 业务的最主要 QoS 要求可以分别表示为:

$$\text{RT 业务: } W_i > T_i \quad (1.1)$$

其中  $W_i$  表示包延时, 参数  $T_i$  表示最大允许时延;

$$\text{NRT 业务: } R_i^{\min} \leq R_i \leq R_i^{\max} \quad (1.2)$$

其中  $R_i$  表示用户的实际速率,  $R_i^{\min}$  和  $R_i^{\max}$  表示最小/最大速率限制。

除此之外还有一个业务类型: 尽力而为(Best-Effort, BE)业务, 既没有时延要求, 也没有最小速率要求。

表 1.1 总结了以上提到的业务类型及相应的 QoS 要求。

表 1.1 常见业务类型及其 QoS 要求

业务类型	主要 QoS 要求	应用(常用 QoS 参数)
实时业务 (RT 业务)	$W_i > T_i$	语音业务(最大允许时延, 固定传输速率); 流媒体业务(最大允许时延, 最小/最大速率限制);
非实时业务 (NRT 业务)	$R_i^{\min} \leq R_i \leq R_i^{\max}$	FTP 业务(最小/最大速率限制);
尽力而为业务 (BE 业务)	无	E-mail;

### 1.3 无线分组调度

QoS 的保证主要依靠通信系统中的分组调度技术完成。本节首先简单介绍有线网络中的分组调度技术。无线分组调度面临的情况要复杂的多, 1.3.2 节将对其主要的三个研究方向做简单的总结, 并指明现在的研究重点。

#### 1.3.1 有线网络中的分组调度

有线网络在提供给用户更多服务内容的同时还必须保证用户的服务质量。常见分组调度算法包括加权轮询<sup>[6]</sup>(Weighted Round Robin, WRR)算法以及广义处理器共享<sup>[7][8]</sup>(Generalized Processor Sharing, GPS)相关的公平排队(Fair Queuing, FQ)算法。公平排队算法可以保证各个连接的基本带宽以及对时延的要求, 是目前有线网络中的调度算法研究的热点。

GPS 假设分组无限可分, 是一种理想的排队系统, 它既可保证连接的最小带宽, 又能绝对公平地在各个连接之间分配带宽。考虑如下带宽分配问题: 假设有  $I$  个用户分享一个速率为  $C$  的单向链路,  $i \in \{1, 2, \dots, I\}$ , 每个用户都根据其 QoS 要求, 计算出一个权重  $r_i$ , 满足如下条件:

$$\sum_i r_i \leq 1 \quad (1.3)$$

GPS 假设其中分组无限可分,任意两个有分组在等待的用户在任意一段时间得到的服务量与他们的权重成正比。也就是说,在任意时刻所有用户均得到相应的服务。这样, GPS 对每个连接的保证带宽为

$$g_i = \frac{r_i}{\sum_i r_i} \times C \geq r_i \cdot C \quad (1.4)$$

即可以保证每个用户的最小带宽,分组的最大时延只与用户本身的参数有关,与其它用户的流量特性无关。

GPS 是一种想象中的排队方法,现实中分组网络接受服务的只能是一个分组。分组化的 GPS(Packet GPS)按照分组在 GPS 系统中离开的顺序依次对分组进行服务,又被称为 WFQ(Weighted Fair Queuing)。

### 1.3.2 无线网络中的分组调度

将有线网络的分组调度技术应用到无线网络中时,最主要考虑是无线环境的特殊性:无线信道的容量是动态变化的,信道状态具有位置依赖性,信道误码率高且具有突发性,并且由于带宽有限存在多用户对信道的竞争,而移动用户并不知道竞争状态等问题。无线网络中的分组调度近年来主要有三个研究方向:

(1) 仍然以有线网络的 GPS 模型为参考,将信道分为 clean 和 dirty 两种状态, clean 意味着信道两端均检测到信道无误码,而 dirty 意味着任一端检测到信道误码。定义无差错服务(error free service)为理想状态下(即 clean channel)用户得到的服务,并依此将队列分为超前,滞后和同步三种类型,在固定时刻实际接受无差错服务超过其预定的无差错服务的流为超前流(lead),而实际接受无差错服务少于其预定的无差错服务的流为滞后流(lag),如果流既不超前也不滞后,则称之为同步流(syn)。当滞后队列的信道从 dirty 变成 clean 时,根据一定的补偿模型对其服务量进行补偿。

(2) 完全应用有线网络的分组调度算法,不考虑无线环境,分组的正确传输由上层的 ARQ(Automatic Repeat reQuest)保证。有线网络中由于信道条件稳定,传输错误概率低,所以可以利用上层的 ARQ 来保证正确传输,但无线信道的传输错误概率要远高于有线信道,如果完全依赖上层的 ARQ 则重传概率非常大,造成了无线资源的浪费。为了解决这个问题,研究人员提出将 ARQ 放在下层,并结合自适应调制编码等技术,提出混合 ARQ(Hybrid-ARQ, HARQ)技术。

(3) 将无线信道条件因素带入权重因子的计算中,使得权重因子不仅是 QoS 参数的函数,也是信道条件的函数。

1999 年以前关于方向(1)的研究很多, 如 IWFQ(Idealized Wireless Fair Queuing), SBFA(Server Based Fairness Approach), WFS(Wireless Fair Service)算法等<sup>[9]</sup>。实际上这些算法都存在结构复杂等缺点。2000 年之后方向(2)和(3)成为研究热点。本文的主要研究方向是(3), 即将 QoS 参数带入权重函数的方法。

## 1.4 本文贡献与结构安排

本文研究单载波和多载波系统中, 针对 BE 业务和 NRT 业务的无线分组调度技术, 主要贡献和文章结构安排如下。

### 1.4.1 本文主要贡献

- (1) 在单载波系统中, 提出了一种针对 BE 业务的渐近吞吐量最大化算法(Asymptotic Throughput Maximization Algorithm, ATMA), 目标是满足公平性准则的同时, 提高系统吞吐量。为了克服比例公平调度造成较大的吞吐量损失的缺陷, 该算法采用了一种新的、以线性函数为渐近线的效益函数, 称为渐近线性效益函数, 算法的目标是最大化各用户渐近线性效益之和。仿真结果表明, 参数合适的 ATMA 算法满足 3GPP2 的公平性准则, 并且系统吞吐量要明显高于传统的比例公平算法。
- (2) 在单载波系统中, 提出了一种针对 NRT 业务的 S-效益最优化(Sigmoid-Utility Optimization, S-UO)算法, 目标是提供最小速率保证的同时, 提高系统效益与系统吞吐量。为了克服传统 Penalty 算法效益函数在最小速率限制点处无定义, 且给速率稍大于最小速率限制的用户过多的优先权的缺陷, 该算法利用 S 曲线形状类似于阶跃函数的特性, 将效益函数写成 S 函数与对数函数相乘的形式, 称为 S-效益(Sigmoid-Utility)函数, 算法优化目标就变为最大化各用户 S-效益之和。与传统的 Penalty 算法相比, S-UO 算法所采用的效益函数更符合实际情况。仿真结果表明, 该算法在系统效益和系统吞吐量两方面都优于传统的 Penalty 算法。
- (3) 在多载波系统中, 提出了一种适用于 OFDMA 系统的、针对 BE 业务的改进多载波比例公平调度(Modified Multi-Carrier Proportional Fairness, Modified-MC-PF)算法, 采用全搜索的方法寻找用户在各子载波上最大可达速率与该用户已达速率比值的最大值, 将最大值对应的子载波分配给对应的用户。与其他算法相比, 此算法能够提高系统效益与系统吞吐量。
- (4) 在多载波系统中, 提出了一种适用于 OFDMA 系统、针对 NRT 业务的多载

波有最小/最大速率限制时的比例公平调度(Multi-Carrier Proportional Fairness with Minimum/Maximum Rate constraints, MC-PFMR)算法,并给出了几种变形算法,这些算法的目标都是在满足最小/最大速率限制的条件下尽可能达到用户速率间的比例公平。MC-PFMR 利用 OFDMA 系统的各用户在各子载波上信道响应各不相同的特点,引入子载波上最小/最大速率限制的概念,并给每个用户一个令牌计数器。调度器在每一时隙开始时刻依次将各个子载波分配给用户,用令牌数反映速率限制的要求,用子信道状况和已达速率的比值反映比例公平的要求,并在每个子载波被分配后都更新各用户已达速率和令牌数。仿真结果表明,此算法在容纳用户数、系统效益等方面性能优于现有其他常用算法,并且能与传统的无速率限制的比例公平调度完全兼容。

- (5) 在多载波系统中,提出了一种适用于 OFDMA 系统、针对 NRT 业务的多载波有最小/最大速率限制时的最大化吞吐量调度(Multi-Carrier Maximum Throughput with Minimum/Maximum Rate constraints, MC-MTMR)算法,目标是在满足最小/最大速率限制的条件下最大化系统吞吐量。算法引入子载波上最小/最大速率限制的概念,采用令牌计数器的方法,用令牌反映速率限制的要求,用子信道状况反映最大化吞吐量的要求,并在每个子载波被分配后都更新各用户令牌数。仿真结果验证了 MC-MTMR 算法的有效性。

## 1.4.2 本文结构安排

第二章讨论单载波系统中的无线分组调度技术。由于单载波无线分组调度技术已经发展得很成熟,多载波系统的调度算法常常以单载波系统的研究成果为基础,所以第二章首先详细介绍了针对不同业务类型的各种算法的目标、步骤和性能,包括针对 BE 业务的最大化吞吐量调度,资源公平的轮询调度,速率公平的最小-最大公平调度,和比例公平调度等,以及针对 NRT 业务的只有最小速率限制时的 Penalty 算法,有最小/最大速率限制时的比例公平调度和最大化吞吐量调度等。在第二章的最后,提出了针对 BE 业务的 ATMA 算法和针对 NRT 业务的 S-UO 算法,并且给出了仿真结果。

第三章讨论多载波系统中的无线分组调度技术,研究对象为 OFDMA 系统。首先介绍了 OFDM 技术和 OFDMA 系统。在 OFDMA 系统中,资源分配问题的复杂度要远远大于单载波系统,因为不仅牵涉到时隙的分配,还牵涉到子载波和功率的分配。在第三章的开始,先讨论了 OFDMA 系统中的功率分配问题,说明在各个子载波上均分功率方案的理论依据,而此时分组调度问题将得到大大的简化。基于以上认识,第三章第二部分讨论了功率均分时 OFDMA 系统模型,并总



结了已有的调度算法，包括针对 BE 业务的最大化吞吐量调度算法和三种比例公平调度方案。第三部分提出了针对 BE 业务的改进多载波比例公平调度算法，通过仿真分析了算法的性能。第四部分提出了针对 NRT 业务的 MC-PFMR 算法，讨论了 MC-PFMR 算法的几种变形算法，并用仿真实验详细分析了 MC-PFMR 算法的性能。第五部分提出了针对 NRT 业务的 MC-MTMR 算法及其变形算法，并给出了仿真结果。

第四章对全文进行总结，小结了本文研究工作的意义，以及上面提到的几个算法的目标、特点和性能。最后指出了下一阶段的研究方向。

## 第二章 单载波系统中的无线分组调度

本章讨论单载波系统中的无线分组调度技术。在以 cdma2000 1x EV-DO 为代表的单载波系统中, 分组调度技术已经发展得非常成熟。研究这些调度算法不仅对分组调度的理论、方法有较为完整的认识, 更为多载波系统中的分组调度技术研究打下良好的基础。

本章首先介绍了 cdma2000 1x EV-DO 系统, 然后介绍了 cdma2000 1x EV-DO 中针对 BE 业务的调度算法。从系统方面考虑, 调度算法需要让系统吞吐量尽可能大, 而从用户角度考虑, 调度算法又必须保证用户之间的公平性。只从系统角度出发的最大化吞吐量调度算法是比较简单和直观的, 但从用户角度出发, 用户之间公平性却有不同衡量标准: 从资源公平方面考虑, 有资源公平的轮询调度; 从最终用户速率相同方面考虑, 有速率公平的最大-最小公平调度。而比例公平是综合考虑系统和用户这两方面的因素、在吞吐量和公平性之间做折中的策略。本章将对这几个概念做详细的介绍, 包括概念的定义以及推导等, 并介绍效益函数的概念, 由此将不同的调度算法统一到最优化不同形式的效益函数这个框架中, 为后续章节的推导打下了基础。

针对有速率限制要求的 NRT 业务, 调度算法在使各个用户满足速率限制的前提下, 要考虑系统吞吐量和用户公平性之间的折中。早期的算法研究提出修改比例公平调度的最优化目标函数, 以期在满足最小速率限制要求的同时, 在系统吞吐量和用户公平性之间做出折中, 这方面的经典算法是 Penalty 算法。近期的研究提出了不修改比例公平调度的最优化目标, 而另加上速率限制条件的方案, 较新的研究成果就是 PFMR 算法, 该算法利用令牌计数器来使得各个用户的速率限制要求得到满足。

本章的最后提出了针对 BE 业务的新算法 ATMA 算法, 和针对 NRT 业务的新算法 S-UO 算法, 并分别给出了仿真结果。ATMA 算法采用一种新的、以线性函数为渐近线的函数作为最优化目标函数, 以期在满足 3GPP2 公平性准则的同时提高系统吞吐量。S-UO 算法仍然采用修改比例公平调度最优化目标函数的方法, 提出利用 S 函数来求出最优化目标函数, 克服了 Penalty 算法中最优化目标函数不符合实际情况的缺陷。

### 2.1 cdma2000 1x EV-DO 系统简介

cdma2000 1x EV-DO<sup>[10][11]</sup>是随着无线网络的飞速发展和业务类型多样化, 为

了专门提供高速的下行数据服务而提出,并在近年来得到很大发展的。cdma2000 1x EV-DO 下行链路采用了时分为主、码分为辅的多址方式,因为结构典型,技术发展成熟,成为研究无线分组调度技术的经典研究对象。

在下文中,下行链路(downlink, down-stream)、前向链路(forward link)均指从基站到移动台的链路,统一称为下行链路;上行链路(uplink, up-stream)、反向链路(reverse link)均指从移动台到基站的链路,统一称为上行链路。

### 2.1.1 技术特点

提出 cdma2000 1x EV-DO 主要基于以下几点考虑:第一,数据业务和语音业务有着截然不同的 QoS 要求,数据业务对时延和时延抖动有一定容忍程度,但对差错率比较敏感,对速率要求比较严格,QoS 等级多,突发性强,不必长时间占用信道,而语音业务对差错率有一定的容忍程度,但对时延和时延抖动非常敏感,QoS 等级单一,需要长时间占用信道;第二,上行和下行数据业务呈现高度不对称的特点,如 Internet 服务,下行链路的流量要远高于上行链路,而语音业务的上行链路和下行链路流量基本对称。

cdma2000 1x EV-DO 系统有三个最主要的技术特点:

(1) cdma2000 1x EV-DO 系统的提出是为了提供高速高质量的数据业务服务,如果将语音和数据业务分别放在不同的载波上,对两者采取不同的传输和控制方法,则可以大大简化系统设备的结构,使得两种业务分别得到好的服务质量。所以,cdma2000 1x EV-DO 系统是在 cdma2000 1x 的 1.25MHz 带宽的旁边,另外部署的一个 1.25MHz 带宽的系统。

(2) 在多址方式上,cdma2000 1x EV-DO 系统针对数据业务流量不对称的特点,在下行链路采用了时分复用的方式,也就是说,cdma2000 1x EV-DO 系统中,一个时刻只有一个用户在接受服务,该用户占用全部的功率和码资源,如图 2.1 所示。但是,cdma2000 1x EV-DO 系统的调制解调和扩频方式仍然采用 CDMA 技术,这就保留了原来 CDMA 技术抗多用户干扰的特性。

(3) 由于每个时隙中只有一个用户在接受服务,基站使用其全部发射功率给这个用户发送数据,所以,cdma2000 1x EV-DO 系统中不再需要功率控制,取而代之的是速率控制,就是根据不同用户的不同信道条件调整数据发送速率,保证用户能以信道条件允许的最大速率正确接受信息。

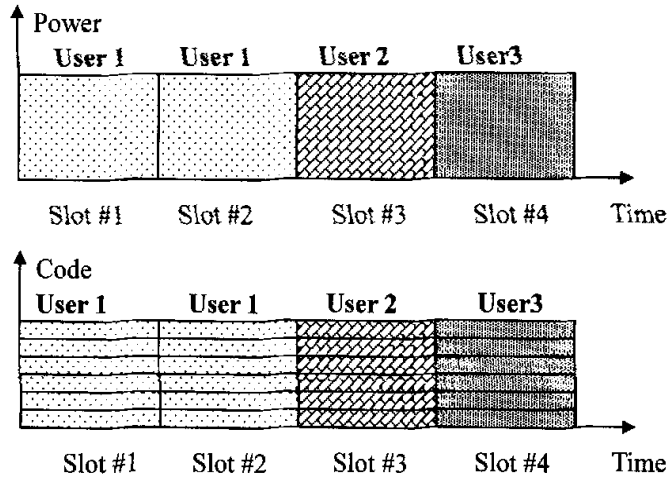


图 2.1 cdma2000 1x EV-DO 下行链路功率及码资源分配

cdma2000 1x EV-DO 系统在下行链路采用时分为主、码分为辅，各用户时分复用，基站一直以满功率发送数据的多址接入方式。其理论依据是信息论角度的一个证明<sup>[13]</sup>：无论是上行链路还是下行链路，采用简单的 TDMA 方式，同时利用灌水法则分配功率将会使得系统和容量最大。cdma2000 1x EV-DO 系统针对数据业务上行链路和下行链路流量高度不对称的特点，在上行链路中仍然采用码分为主、时分为辅的多址接入方式，而在下行链路中采用时分为主、码分为辅的多址接入方式，以期在下行链路实现高吞吐量。

### 2.1.2 系统模型

在 cdma2000 1x EV-DO 系统的下行链路中，时间被分成固定长度的时隙，各用户时分复用，发射机一直以满功率向用户发送数据。假设总共有  $I$  个用户，每个用户都在检测下行链路质量，由此预测在下一个时隙内的最大可达速率，并由上行链路将这些速率值报告给调度器。在每一时隙的开始时刻，调度器综合考虑各个用户报告的最大可达速率和各个用户的已达速率，决定将这一时隙分配给哪一个用户传输数据。假设一旦某一用户被允许在这一时隙传输，他的实际传输速率就是报告的预测最大可达速率。下行链路发送流程如图 2.2 所示，DRC 为 cdma2000 1x EV-DO 系统上行链路中报告下行最大可达速率的逻辑信道名。分组调度器挑选下一时隙发送数据的用户，其数据经过自适应调制编码后，串并转换为 16 路数据流，再通过 Walsh 码扩频，最后送到 PN 码调制部分和上变频部分。

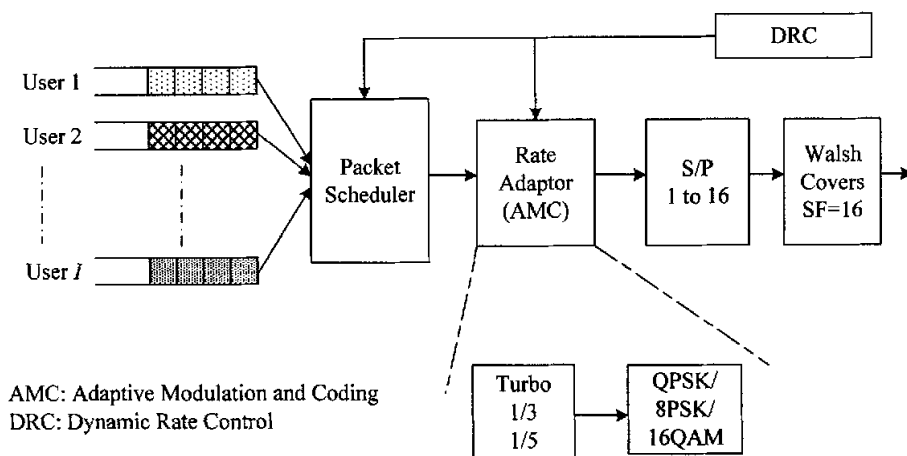


图 2.2 cdma2000 1x EV-DO 下行链路示意图<sup>[14]</sup>

用  $R_i(t)$  (单位为 bps) 表示用户  $i$  在时隙  $t$  结束时刻之前长度为多个时隙的窗口内的平均速率,  $i \in \{1, 2, \dots, I\}$ , 用  $C_i(t)$  (单位 bps) 表示用户  $i$  向调度器报告的最大可达速率,  $\mu_i(t)$  (单位 bps) 表示用户  $i$  在时隙  $t$  中实际达到的速率。如果调度器将时隙  $t$  分配给用户  $i$  发送数据, 则  $\mu_i(t)$  等于  $C_i(t)$ , 否则  $\mu_i(t)$  等于零。记滑动平均窗口长度为  $T_{win}$ , 单位为时隙个数。在时隙  $t$  的结束时刻, 各用户将其平均速率  $R_i(t)$  由上一个时间窗内的值更新为当前时间窗内的值<sup>[15]</sup>:

$$\begin{aligned}
 R_i(t) &= \left(1 - \frac{1}{T_{win}}\right) R_i(t-1) + \frac{1}{T_{win}} \mu_i(t) \\
 &= (1 - \beta) R_i(t-1) + \beta \mu_i(t)
 \end{aligned}
 \tag{2.1}$$

式中常数  $\beta = 1/T_{win}$ 。

## 2.2 cdma2000 1x EV-DO 中的无线分组调度

本节总结应用在 cdma2000 1x EV-DO 下行链路的已有的经典分组调度算法。2.2.1 节总结针对 BE 业务的调度算法, 包括最大 C/I(Max-C/I)算法, 轮询算法(Round Robin, RR)算法, 最大-最小公平(Max-Min Fairness, Max-Min)算法和比例公平(Proportional Fairness, PF)算法。随着无线网络的发展, 网络中的数据业务不再仅限于 BE 业务, 而是出现了有多种 QoS 要求的业务, 如有速率限制要求的 NRT 业务。2.2.2 节总结了针对 NRT 业务的调度算法, 包括只有最小速率限制时的惩罚(Penalty)算法, 有最小/最大速率限制时的比例公平(Proportional Fairness with Minimum/Maximum Rate constraints, PFMR)算法, 和有最小/最大速率限制时的最大吞吐量(Maximum Throughput with Minimum/Maximum Rate constraints, MTMR)算法。

## 2.2.1 针对 BE 业务的调度算法

BE 业务没有时延、速率要求，所以吞吐量和公平性成为针对 BE 业务的分组调度算法需要考虑的两个最重要方面。从系统角度考虑，调度算法必须充分利用信道的时变特性，将时隙分配给信道条件最好的用户；而从用户角度考虑，信道条件最好的用户必然是比较靠近基站的用户，如果只将时隙分配给信道条件最好的用户，势必造成离基站较远的信道条件较差的用户一直得不到传输机会，调度算法必须考虑用户之间的公平性，将一部分传输机会分配给信道条件较差的用户，这又导致了整体系统吞吐量的下降。如何在系统吞吐量和用户公平性之间做出折中，就成为调度算法必须考虑的问题。下面介绍几种常见的调度算法。

### 2.2.1.1 最大化吞吐量的最大 C/I 调度

系统吞吐量可以用各个用户的速率之和  $\sum_i R_i(t)$  表示。最大化吞吐量调度就是要让  $\sum_i R_i(t)$  最大，由式(2.1)推导如下。

若时隙  $t$  分配给用户  $i^*$  可以使得在时隙  $t$  的结束时刻系统吞吐量达到最大值，则根据(2.1)式，在时隙  $t$  的结束时刻  $\mu_i(t)$  和  $R_i(t)$  做如下更新：

$$\begin{aligned} \mu_i(t) &= \begin{cases} C_i(t), & \text{if } i = i^* \\ 0, & \text{if } i \neq i^* \end{cases} \\ R_i(t) &= \begin{cases} (1-\beta)R_i(t-1) + \beta C_i(t), & \text{if } i = i^* \\ (1-\beta)R_i(t-1), & \text{if } i \neq i^* \end{cases} \end{aligned} \quad (2.2)$$

则最大化吞吐量的目标等价于：

$$\begin{aligned} & \max \sum_i R_i(t) \\ & \Leftrightarrow \max \left\{ \left[ \sum_{i \neq i^*} (1-\beta)R_i(t-1) \right] + \left[ (1-\beta)R_{i^*}(t-1) + \beta C_{i^*}(t) \right] \right\} \\ & \Leftrightarrow \max \left\{ \left[ \sum_i (1-\beta)R_i(t-1) \right] + (1-\beta)R_{i^*}(t-1) + \beta C_{i^*}(t) - (1-\beta)R_{i^*}(t-1) \right\} \quad (2.3) \\ & \Leftrightarrow \max \left\{ \left[ \sum_i (1-\beta)R_i(t-1) \right] + \beta C_{i^*}(t) \right\} \end{aligned}$$

因为第一项  $\sum_i (1-\beta)R_i(t-1)$  是常数，可以省略， $\beta$  为大于零的常数，不影响函数的增减性质和极值点，所以最大化吞吐量调度的调度策略为：将时隙  $t$  分配给用户  $i^*$ ，

$$i^* = \arg \max_i C_i(t) \quad (2.4)$$

式中,  $C_i(t)$  为用户  $i$  向调度器报告的最大可达速率, 是用户  $i$  通过检测到的载波干扰比(Carrier to Interference Ratio, C/I)算出的相应的值, C/I 值越大,  $C_i(t)$  值越大, 所以最大化吞吐量的调度方法被称为最大 C/I(Max-C/I)调度<sup>[12][14]</sup>。Max-C/I 调度算法的本质就是在选择传输用户时, 只选择 C/I 最大的用户, 即让信道条件好的用户一直在传, 等其信道变差时, 再让其他信道变好的用户传输数据, 算法简单直观, 充分利用了多用户分集效果。

Max-C/I 算法得到的吞吐量是吞吐量的极限值, 无论采用哪种其他调度算法, 得到的吞吐量都不可能超过它。但 Max-C/I 算法对用户来说是最不公平的一种方案, 它将导致某些在小区边缘的用户因为信道条件一直较差而“饿死”。

### 2.2.1.2 资源公平的轮询调度

在考虑公平性时, 最直接的方案就是采用轮询(Round Robin)的方法, 即不考虑各个用户的信道状况, 依次将时隙分配给各个用户传输数据, 也就是说所有用户以相同的概率(概率为  $1/I$ )占用时隙和功率, 实现用户间的资源公平(Resource Fairness)。

这种方案在有线网络针对 BE 业务的调度中得到了非常广泛的应用。有线网络中共享信道的信道容量固定, 各个用户的信道容量相同, 实现了资源公平也就实现了各个用户之间的吞吐量公平(Throughput Fairness, 也称为速率公平, Rate Fairness)。这在无线网络中是不成立的, 轮询方案实现了资源公平, 但并不能保证各个用户之间的速率公平。靠近基站的用户信道条件好, 可以以较高的速率传输数据, 远离基站的用户信道条件差, 只能以较低的速率传输数据, 他们在得到相同传输机会和资源时, 达到的速率肯定是不相同的。

### 2.2.1.3 速率公平的最大-最小公平调度

文献[16]提出了最大-最小公平(Max-Min Fairness)的概念, 这个概念起初是在针对 BE 业务的拥塞控制(Congestion Control)中的。假设  $R_i$  (单位为 bps) 是用户  $i$  在很长时间内的平均速率,  $i \in \{1, 2, \dots, I\}$ , 调度器的目标为最大化所有用户速率的最小值(Maximize the Minimum  $R_i$ ), 所以这种公平准则被称为最大-最小公平准则。

最大-最小公平调度器的严格的定义为: 假设在调度器  $M$  下, 各个用户速率为  $R_i^M$ , 在调度器  $S$  下, 各个用户速率为  $R_i^S$ , 调度器  $M$  被称为最大-最小公平调

度器的充要条件是：如果调度器  $S$  使用户  $m$  得到了比调度器  $M$  下更大的速率，则必然导致另一个调度器  $M$  下，比用户  $m$  速率小的用户  $n$  的速率减少，即  $R_n^M < R_m^M$ ， $R_m^S > R_m^M$ ， $R_n^S < R_n^M$ 。换句话说，任意一个用户速率的增加，必然是以原本比他速率小的用户的速率减少为代价的，即导致了所有用户速率的最小值的减小。

很容易可以看到，最大-最小公平调度策略为：将时隙  $t$  分配给用户  $i^*$ ，

$$i^* = \arg \max_i \frac{1}{R_i(t-1)} \quad (2.5)$$

即将时隙  $t$  分配给已达速率最小的用户。这种方法最终渐近实现了用户间的速率公平，它与 Max-C/I 相反，信道条件较差的用户反而会得到较多的传输机会，而与轮询调度类似，都是完全从用户角度出发考虑而没有考虑系统吞吐量要求的。

#### 2.2.1.4 比例公平调度

上面三种调度方法，都是单方面从系统吞吐量或用户公平性角度出发得出的调度策略。在文献[15]中，QualComm 公司提出了一种称为比例公平的调度算法，因为它能够在系统总吞吐量和用户公平性之间做出较好的折中，近年来研究较多，现简述如下。

##### 比例公平的定义

在博弈论中，比例公平调度定义为<sup>[17]</sup>：假设在调度器  $P$  下，各个用户平均速率为  $R_i^P$ ，在调度器  $S$  下，各个用户平均速率为  $R_i^S$ ，当且仅当条件(2.6)满足时，调度器  $P$  被称为比例公平调度器，

$$\sum_i \frac{R_i^S - R_i^P}{R_i^P} \leq 0 \quad (2.6)$$

也就是说，任何一个用户速率上的变化，必然导致最终变化比例的总和为负值或零，即如果调度器  $S$  下其中一个用户  $k$  的速率为  $R_k^S$ ，与比例公平调度器  $P$  下的结果  $R_k^P$  相比，增大了  $x\%$ ，必然导致其他用户的  $R_i^S$  与  $R_i^P$  ( $i \neq k$ ) 相比，变化比例的总和并不小于  $x\%$ 。

我们可以用一个极端的例子来定性说明比例公平的原则。假设有两个用户时分复用共享一个信道，时隙长度非常小，且这两个用户在这个信道上的信道容量为各不相同的固定值，分别记为  $C_1$  (单位为 bps)和  $C_2$  (单位为 bps)。不失一般性，我们假设  $C_1 > C_2$ 。考察在很长一段时间后用户的速率  $R_1$  与  $R_2$ ，如果最终目标是使得吞吐量最大，则应该将所有传输机会都分配给用户 1，此时  $R_1 = C_1$ ， $R_2 = 0$ ，



很明显的看到,这对用户 2 来说是很不公平的:他一直得不到传输机会;进一步可以证明,存在一个用户速率为零的方案肯定不是比例公平的(式(2.6)不等号左边的项将变成无穷大)。如果最终目标是用户速率间完全公平,即  $R_1 = R_2$ , 则信道容量较大的用户的传输机会应该较少,最后得到  $R_1 = R_2 = C_1 C_2 / (C_1 + C_2)$ 。从附录 A 的证明中可知,在这个例子中,式(2.6)的充要条件是  $R_1 : R_2 = C_1 : C_2$ , 用户最终吞吐量与信道条件好坏成比例,并且,此时系统吞吐量为  $(C_1 + C_2) / 2$ 。比较在三个不同目标下最终导致的总吞吐量的大小,最大吞吐量调度使得系统吞吐量为  $C_1$ , 用户速率间完全公平导致系统吞吐量为  $2C_1 C_2 / (C_1 + C_2)$ , 而比例公平导致系统吞吐量为  $(C_1 + C_2) / 2$ , 介于  $2C_1 C_2 / (C_1 + C_2)$  与  $C_1$  之间。

由这个例子可以看出,比例公平的原则导致系统吞吐量和用户公平性之间的折中。一方面,它导致最终吞吐量比系统可达的最大吞吐量小,另一方面,它保证了无论用户 2 的信道条件有多差(即无论  $C_2$  有多小),用户 2 都能得到一定的传输机会,保证了一定的公平性。需要说明的是,在这个例子中,比例公平与用户速率满足  $R_1 : R_2 = C_1 : C_2$  是等价的,但在其他情况下,这两者不一定等价。

### 比例公平的等价条件

比例公平的定义式(2.6)并不能直接应用在实践中。当信道是时变信道时,如果根据定义来求解比例公平的方案的话,必须已经知道很长一段时间内的各用户的信道状况,然后在所有可能的分配方案中,通过全比较的方法来确定在比例公平意义下的最佳分配方案。这在实践中是不可能实现的,第一,在很长时间内信道状况很难预测准确,第二,随着时间的增长,可能的分配方案的总数呈指数增长,全比较是不现实的。第三,比例公平定义式(2.6)有解的前提条件是资源可以无限细分,也即用户所有可能的可达速率组成的可达速率空间  $R'$  是连通的,在这个前提条件下,可以证明比例公平定义式(2.6)必然有解且解唯一(见附录 B)。而在实际中,资源不是无限可分的(例如时间往往被分成等长的时隙),可达速率空间是由一个个离散的点组成的,在这些离散的可达速率中,可能并不存在某一点能够满足式(2.6),也即,实际情况下完全符合式(2.6)的解可能不存在。

因此,必须推导出实践中能够应用的比例公平等价条件。在这方面已有一些研究,文献[16]证明了比例公平调度器的目标是使用户速率的对数和最大,即比例公平调度器的目标可以写成:

$$\max \sum_i \log R_i \quad (2.7)$$

附录 B 中给出了式(2.6)与式(2.7)等价的证明。

式(2.7)可以从另一方面理解。当用户速率为  $R_i$  时,用效益函数来表示用户对  $R_i$  的满意程度<sup>[18][19]</sup>,调度器的任务就是让系统效益(即所有用户效益之和)最大。

对于 BE 类型的传统数据业务，效益函数通常写成  $\log R_i$  的形式，为递增、凸函数，表示效益递增、边际效益递减。

实践中常用用户  $i$  在时隙  $t$  结束时刻前固定长度窗口内的平均速率的  $R_i(t)$  代替  $R_i$ ，这样，式(2.7)就可以写成：

$$\max \sum_i \log R_i(t) \quad (2.8)$$

也就是说，调度器的目标可以表述成：在时隙  $t$  的开始时刻，调度器将时隙  $t$  分配给用户传输数据，使得时隙  $t$  结束时刻的各个用户平均速率  $R_i(t)$  的对数和达到最大值。由于未来很长时间内的信道状况是很难准确预测的，但在紧接的一个时隙内的信道状况还是能够预测得比较准确的，式(2.8)就为比例公平调度提供了一个可实现的目标。

### cdma2000 1x EV-DO 中的比例公平调度

下面我们推导 cdma2000 1x EV-DO 系统中的比例公平调度算法。与最大化吞吐量调度的推导过程类似，若时隙  $t$  分配给用户  $i^*$  可以使得在时隙  $t$  的结束时刻所有用户速率的对数和达到最大值，则在时隙  $t$  的结束时刻  $R_i(t)$  的更新仍如式(2.1)所示，将式(2.1)代入式(2.8)，得到，

$$\begin{aligned} & \max \sum_i \log R_i(t) \\ \Leftrightarrow & \max \left\{ \left[ \sum_{i \neq i^*} \log(1-\beta)R_i(t-1) \right] + \log \left[ (1-\beta)R_{i^*}(t-1) + \beta C_{i^*}(t) \right] \right\} \\ \Leftrightarrow & \max \left\{ \left[ \sum_i \log(1-\beta)R_i(t-1) \right] + \log \left[ (1-\beta)R_{i^*}(t-1) + \beta C_{i^*}(t) \right] \right. \\ & \quad \left. - \log(1-\beta)R_{i^*}(t-1) \right\} \end{aligned} \quad (2.9)$$

因为第一项  $\sum_i \log(1-\beta)R_i(t-1)$  为常数，可以省略，得到，

$$\begin{aligned} & \Leftrightarrow \max \left\{ \log \frac{(1-\beta)R_{i^*}(t-1) + \beta C_{i^*}(t)}{(1-\beta)R_{i^*}(t-1)} \right\} \\ & \Leftrightarrow \max \left\{ \log \left[ 1 + \frac{\beta}{1-\beta} \cdot \frac{C_{i^*}(t)}{R_{i^*}(t-1)} \right] \right\} \end{aligned} \quad (2.10)$$

因为  $\log$  函数是单调增函数， $\beta/(1-\beta)$  为常数，所以 cdma2000 1x EV-DO 系统中针对 BE 业务的比例公平调度策略为，将时隙  $t$  分配给用户  $i^*$ ：

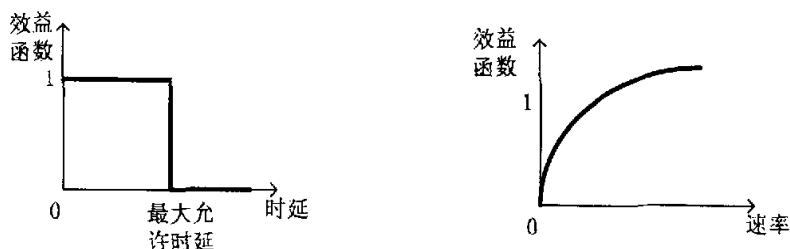
$$i^* = \arg \max_i \frac{C_i(t)}{R_i(t-1)} \quad (2.11)$$

式(2.11)的意义是明显的, 当用户信道条件较好, 会因为  $C_i(t)$  较大而得到较高优先权, 同时如果有用户长时间内得不到传输机会, 也会因为  $R_i(t-1)$  较小而得到较高的优先权。通过这种方式, 比例公平调度实现了系统吞吐量和用户公平性之间的折中。文献[20][21]详细分析了比例公平调度算法的数学性质, 文献[22-24]提出了几种改进的比例公平调度算法。

### 2.2.1.5 效益最优化调度

在介绍比例公平算法时, 提到了效益函数(Utility Function)的概念, 并且指出当效益函数写成对数函数的形式时, 比例公平调度的目的就是让系统效益(各用户效益函数之和)最大。由于下文的讨论将经常涉及到效益函数这个概念, 现在先做一个简单的介绍。

通常用效益函数  $U(\cdot)$  来表示用户对服务(包括达到的速率, 时延等)的满意程度<sup>[18][19]</sup>。实时业务因为其时延敏感性, 效益函数通常有如图 2.3(a)所示的特性, 而 BE 业务和 NRT 业务的效益函数通常有对数函数的特性, 如图 2.3(b)所示。对数函数是递增、凸函数, 表示随着速率的增大, 用户的满意程度是递增的, 但是增长的越来越慢(即边际效益递减)。



(a) 实时业务的效益函数

(b) BE 业务与 NRT 业务的效益函数

图 2.3 效益函数示意图

以上只给出了比较典型的两种效益函数示意图。随着 QoS 的等级、参数的不同, 效益函数也呈现出不同的特性。

引入效益函数的概念后, 就可以将不同调度算法的目标统一到最大化不同形式的系统效益(即各用户效益函数之和)的框架下来。BE 业务和 NRT 业务最常用的效益函数就是对数函数, 即  $U_i(R_i(t)) = \log R_i(t)$ , 比例公平调度的目标就是最大化对数函数形式的系统效益; 而最大 C/I 调度可以理解为效益函数写成  $U_i(R_i(t)) = R_i(t)$  时, 最大化系统效益的情况。

下面讨论一般形式的效益函数的最大化问题, 问题可以表示为:

$$\max \sum_i U_i(R_i(t)) \quad (2.12)$$

如果  $U_i(R_i(t))$  是凸函数, 则根据凸函数在凸集上极值唯一, 并且局部极值就是全局极值的性质可知, 式(2.12)有唯一解, 并且可以通过求导的方法求得局部极值点, 从而得到全局极值点<sup>[25]</sup>。利用式(2.1)简单推导如下:

$$\begin{aligned} & \max \sum_i U_i(R_i(t)) \\ \Leftrightarrow & \max \left\{ \left[ \sum_{i \neq i^*} U_i((1-\beta)R_i(t-1)) \right] + U_{i^*}((1-\beta)R_{i^*}(t-1) + \beta C_{i^*}(t)) \right\} \\ \Leftrightarrow & \max \left\{ \left[ \sum_i U_i((1-\beta)R_i(t-1)) \right] + U_{i^*}((1-\beta)R_{i^*}(t-1) + \beta C_{i^*}(t)) \right. \\ & \quad \left. - U_{i^*}((1-\beta)R_{i^*}(t-1)) \right\} \end{aligned} \quad (2.13)$$

因为第一项  $\sum_i U_i((1-\beta)R_i(t-1))$  是常数项, 可以省略, 得到,

$$\begin{aligned} & \Leftrightarrow \max \left\{ U_{i^*}((1-\beta)R_{i^*}(t-1) + \beta C_{i^*}(t)) - U_{i^*}((1-\beta)R_{i^*}(t-1)) \right\} \\ & \Leftrightarrow \max \left\{ \frac{U_{i^*}((1-\beta)R_{i^*}(t-1) + \beta C_{i^*}(t)) - U_{i^*}((1-\beta)R_{i^*}(t-1))}{\beta C_{i^*}(t)} \cdot \beta C_{i^*}(t) \right\} \end{aligned} \quad (2.14)$$

$\beta \rightarrow 0$  时, 式中乘号左边的项就是导数的定义式, 由此, 以式(2.12)为目标的调度算法的调度策略为: 将时隙  $t$  分配给用户  $i^*$

$$i^* = \arg \max_i U_i'(R_i(t-1)) \cdot C_i(t) \quad (2.15)$$

式中  $U_i'(\cdot)$  表示求导运算。式(2.15)更加严格的数学证明参见文献[26][27]。需要注意的是, 如果  $U_i(R_i(t))$  是凸函数, 则用求导的方法可以得到唯一的最优的调度策略, 但是如果  $U_i(R_i(t))$  是凹函数, 由于凹函数在凸集上的极值并不唯一, 用求导的方法求得的调度策略只是最优调度策略的必要而非充分条件。

## 2.2.2 针对 NRT 业务的调度算法

以上主要讨论了针对 BE 业务的比例公平调度。实际还有一些业务类型是有最小/最大速率限制等 QoS 要求的, 这些业务也称为非实时(Non-Real-Time, NRT)业务, 典型应用如 FTP 等。最初的 cdma2000 1x EV-DO 系统只针对 BE 业务, 现在已经有一些研究希望在 cdma2000 1x EV-DO 系统中加入有 QoS 要求的其他业务类型, 研究最多的就是有速率限制的 NRT 业务。

设用户  $i$  的速率限制为:

$$R_i^{\min} \leq R_i(t) \leq R_i^{\max} \quad (2.16)$$

其中  $R_i^{\min}$  (单位 bps) 和  $R_i^{\max}$  (单位 bps) 分别表示用户  $i$  的最小/最大速率限制。本节主要讨论当存在最小/最大速率限制时, 采用不同准则(如最大化吞吐量, 用户间比例公平等)的一些现有调度算法。

### 2.2.2.1 只有最小速率限制时的 Penalty 调度算法

只有最小速率限制时, 应用最广泛的算法是 Penalty 算法<sup>[28]</sup>。Penalty 算法只能处理只有最小速率限制的情况, 目标是满足各用户最小速率限制要求的同时, 在系统吞吐量和用户公平性之间做出良好的折中。它的调度策略可以写成: 在时隙  $t$  的开始时刻, 调度器将时隙  $t$  分配给用户  $i^*$  传输数据,

$$i^* = \arg \max_i \frac{C_i(t)}{R_i(t-1) - R_i^{\min}} \quad (2.17)$$

调度策略式(2.17)采用倒数函数, 当用户速率接近  $R_i^{\min}$  值时, 该用户将得到很高的优先权, 同时从系统的角度出发, 将用户信道情况考虑进来, 当用户信道条件很好,  $C_i(t)$  很大时, 该用户也将得到较高的优先权。

深入分析 Penalty 算法发现, 它是用修改比例公平调度的最优化目标效益函数的方式, 来实现保证最小速率的同时在系统吞吐量和用户公平性之间做折中的, Penalty 算法的目标可以写成:

$$\max \sum_i \log(R_i(t) - R_i^{\min}) \quad (2.18)$$

因为  $\log(R_i(t) - R_i^{\min})$  是递增凸函数, 所以可以用求导的方法来求得最优调度策略, 得到的结果与式(2.17)是一致的。

### 2.2.2.2 有最小/最大速率限制时的比例公平调度

与针对 BE 业务的调度算法相同, 针对 NRT 业务的调度算法也需要考虑系统吞吐量和用户公平性之间的折中。可以不修改比例公平调度的最优化目标函数, 而是另加上速率限制条件, 则比例公平调度的目标变为在满足速率限制的同时, 使得用户速率对数和达到最大, 即:

$$\begin{aligned} \max \sum_i \log R_i(t) \\ \text{Subject to: } R_i^{\min} \leq R_i(t) \leq R_i^{\max} \end{aligned} \quad (2.19)$$

文献[29]提出了 PFMR 算法(Proportional Fairness with Minimum/Maximum Rate constraints)并且证明了此算法可以实现目标(2.19)。PFMR 算法利用“令牌计

数器”使得速率限制得到满足。假设每个用户都有一个令牌计数器， $T_i(t)$ 表示时隙 $t$ 结束时用户 $i$ 的令牌数，量纲为bps。 $T_i(t-1)$ 值大于零表示时隙 $t$ 内需要增大用户 $i$ 的传输机会以满足其最小速率限制要求， $T_i(t-1)$ 值小于零表示时隙 $t$ 内需要减小用户 $i$ 的传输机会以满足其最大速率限制要求。用 $\exp\{a_i T_i(t-1)\}$ 项来反映速率限制的要求，用 $C_i(t)/R_i(t-1)$ 项反映比例公平的要求，PFMR算法可以表述为：在时隙 $t$ 的开始时刻，调度器将此时隙分配给用户 $i^*$ 传输数据，

$$i^* = \arg \max_i \exp\{a_i T_i(t-1)\} \frac{C_i(t)}{R_i(t-1)} \quad (2.20)$$

其中 $a_i$ 是量纲为1/bps的控制参数，在实践中设定， $a_i$ 的值越大，速率限制作用越明显；在时隙 $t$ 的结束时刻， $R_i(t)$ 仍按照式(2.1)做更新，而 $T_i(t)$ 做如下更新：

$$T_i(t) = 0, \text{ if } R_i^{\min} = 0 \text{ and } R_i^{\max} = \infty \quad (2.21)$$

$$T_i(t) = \max\{0, T_i(t-1) + R_i^{\min} - \mu_i(t)\}, \text{ if } R_i^{\min} > 0 \text{ and } R_i^{\max} = \infty \quad (2.22)$$

$$T_i(t) = \min\{0, T_i(t-1) + R_i^{\max} - \mu_i(t)\}, \text{ if } R_i^{\min} = 0 \text{ and } R_i^{\max} < \infty \quad (2.23)$$

$$T_i(t) = \begin{cases} T_i(t-1) + R_i^{\min} - \mu_i(t) & \text{if } T_i(t-1) \geq 0 \\ T_i(t-1) + R_i^{\max} - \mu_i(t) & \text{if } T_i(t-1) < 0 \end{cases}, \text{ if } 0 < R_i^{\min} < R_i^{\max} < \infty \quad (2.24)$$

PFMR算法中最为关键的技术在于令牌的更新规则。令牌的更新必须依循以下准则进行：

(i)当用户 $i$ 没有速率限制时，根据式(2.21)，用户 $i$ 的令牌个数始终为零，式(2.20)退化成式(2.11)，正是没有速率限制时的比例公平调度策略。

(ii)根据式(2.24)，可以看到令牌的增加量为：

$$\Delta T_i(t) = T_i(t) - T_i(t-1) = \begin{cases} R_i^{\min} - \mu_i(t), & \text{if } T_i(t-1) \geq 0 \\ R_i^{\max} - \mu_i(t), & \text{if } T_i(t-1) < 0 \end{cases} \quad (2.25)$$

可见，若 $\mu_i(t)$ 小于 $R_i^{\min}$ 则令牌数增加，若 $\mu_i(t)$ 大于 $R_i^{\max}$ 则令牌数减少，若 $\mu_i(t)$ 介于 $R_i^{\min}$ 与 $R_i^{\max}$ 之间，则 $T_i(t-1)$ 为正时，令牌减少， $T_i(t-1)$ 为负时，令牌增加，以使 $T_i(t)$ 保持在零值附近。

文献[29]中的仿真结果表明，PFMR算法的性能非常好，并能够和没有速率限制时的比例公平算法良好兼容。同时，文献[29]中也说明了PFMR算法并不是问题(2.19)的唯一解。用式(2.20)得到的各用户“长时间内”的平均速率能够达到式(2.19)的目标，而实际上，式(2.20)中的 $\exp\{a_i T_i(t-1)\}$ 项可以用另外的 $f(a_i T_i(t-1))$ 函数来代替，只要函数 $f(\cdot)$ 满足以下三个条件：

$$(1) f(0) = 1$$

$$(2) x \rightarrow +\infty \text{ 时, } f(x) \rightarrow +\infty$$

$$(3) x \rightarrow -\infty \text{ 时, } f(x) \rightarrow 0$$

则在式(2.20)中用  $f(a_i T_i(t-1))$  代替  $\exp\{a_i T_i(t-1)\}$  项, 最后得到各用户“长时间内”的平均速率也能够达到式(2.19)给出的目标<sup>[30]</sup>。

### 2.2.2.3 有最小/最大速率限制时的最大化吞吐量调度

最大化吞吐量调度的目标是要让系统吞吐量  $\sum_i R_i(t)$  最大。当有最小/最大速率限制时, 调度器目标为:

$$\begin{aligned} \max \sum_i R_i(t) \\ \text{Subject to: } R_i^{\min} \leq R_i(t) \leq R_i^{\max} \end{aligned} \quad (2.26)$$

上式与式(2.19)非常类似, 文献[29]也对其进行了研究, 提出了有最小/最大速率限制时的最大化吞吐量调度(Maximum Throughput with Minimum/Maximum Rate constraints, MTMR)算法。与 PFMR 算法类似, MTMR 算法利用令牌计数器使得用户的速率限制要求得到满足, 具体做法是用下式代替式(2.20):

$$i^* = \arg \max_i \exp\{a_i T_i(t-1)\} C_i(t) \quad (2.27)$$

其令牌的更新规则与 PFMR 算法相同。

### 2.2.2.4 有最小/最大速率限制时的效益最优化调度

我们用列表的方法来观察上述几个算法的关系。见表 2.1。

表 2.1 单载波中几种调度算法的总结

业务类型	QoS 要求 (限制条件)	调度算法	
		最大化吞吐量 目标: $\max \sum_i R_i(t)$	比例公平 目标: $\max \sum_i \log R_i(t)$
BE	无	$i^* = \arg \max_i C_i(t)$	$i^* = \arg \max_i \frac{C_i(t)}{R_i(t-1)}$
NRT	$R_i^{\min} \leq R_i(t) \leq R_i^{\max}$	$i^* = \arg \max_i \exp\{a_i T_i(t-1)\} C_i(t)$	$i^* = \arg \max_i \exp\{a_i T_i(t-1)\} \frac{C_i(t)}{R_i(t-1)}$

针对有最小/最大速率限制的 NRT 业务，面对多种多样的效益函数，梯度算法<sup>[29][30]</sup>(Gradient with Minimum/Maximum Rate constraints, GMR)给出了最一般的调度策略。当用户的效益函数可以表示为  $U_i(R_i(t))$ ，且为递增、凸函数，而最优化目标可以表示为  $U(\vec{R}(t)) = \sum_i U_i(R_i(t))$  时，有速率限制时的调度器的目标可以写成：

$$\begin{aligned} \max \sum_i U_i(R_i(t)) \\ \text{Subject to: } R_i^{\min} \leq R_i(t) \leq R_i^{\max} \end{aligned} \quad (2.28)$$

GMR 算法给每个用户一个令牌计数器，并在时隙  $t$  的开始时刻，将此时隙分配给用户  $i^*$  传输数据，

$$i^* = \arg \max_i \exp\{a_i T_i(t-1)\} U_i'(R_i(t-1)) C_i(t) \quad (2.29)$$

在时隙  $t$  的结束时刻，仍然分别按照式(2.1)和式(2.21)-式(2.24)更新所有用户的速率和令牌数。因为式(2.29)中有一个求导的运算，所以这个算法被称为 Gradient 算法。

可以看到，PFMR 算法和 MTMR 算法都是 GMR 算法的特例。当效益函数写成  $U_i(R_i(t)) = \log R_i(t)$  的形式时，GMR 算法就成为 PFMR 算法；当效益函数写成  $U_i(R_i(t)) = R_i(t)$  的形式时，GMR 算法就成为 MTMR 算法。

### 2.3 针对 BE 业务的渐近吞吐量最大化算法(ATMA)

2.2.1 节中介绍了针对 BE 业务的 Max-C/I 调度和比例公平调度。比例公平调度的一个缺陷就是为了用户间公平性牺牲了太多了吞吐量，实验表明比例公平调度下的吞吐量仅为 Max-C/I 调度下的 75%，这实际上与比例公平调度采用的效益函数有关，最大化吞吐量的 Max-C/I 算法的目标为最大化各用户速率之和，可以理解成效益函数被写成了线性函数的形式。线性函数给出了吞吐量的上界，效益



函数越靠近线性函数，吞吐量就越大。而比例公平的目标为最大化各用户速率对数和，效益函数具有对数函数的形式，自变量越大，与线性函数的差距就越大，由此导致吞吐量牺牲得太多。

本节提出一种新的、针对 BE 业务的调度算法，目标是满足 3GPP2 公平性准则的同时提高吞吐量。算法采用以线性函数为渐近线的函数作为效益函数，以实现吞吐量的提高。

### 2.3.1 算法描述

我们采用一种新的效益函数，因为它以线性函数为渐近线，称为渐近线性 (Asymptotic Linear) 效益函数，记为  $U_{i,AL}(R_i(t))$ ，如下所示：

$$U_{i,AL}(R_i(t)) = R_i(t) - \exp\{-R_i(t)\} \quad (2.30)$$

其第一项就是 Max-C/I 调度的目标效益函数；第二项对第一项做了非线性的校正， $R_i(t)$  趋于无穷的时，第二项为零，即不做校正。这是与对数效益函数很不相同的地方，因为后者在  $R_i(t)$  趋于无穷的时候也不会趋向于  $U_i(R_i(t)) = R_i(t)$  线。采用  $U_{i,AL}(R_i(t))$  函数作为效益函数的好处在于：(1) 它与最大化吞吐量调度的线性效益函数很靠近，所以系统吞吐量会高于以对数效益函数之和为最优化目标的算法；(2)  $U_{i,AL}(R_i(t))$  函数永远也不会与线性函数相交，这意味着以各用户  $U_{i,AL}(R_i(t))$  之和为最优化目标，不等同于最大化吞吐量调度，从而保证了各用户间满足 3GPP2 规定的公平性准则。

因为以上提到的性质，可以将以各用户  $U_{i,AL}(R_i(t))$  函数之和为最优化目标的调度算法称为渐近吞吐量最大化算法 (Asymptotic Throughput Maximization Algorithm, ATMA)，算法优化目标可以表示为：

$$\max \sum_i [R_i(t) - \exp\{-R_i(t)\}] \quad (2.31)$$

渐近效益函数是递增的凸函数，所以可以用求导的方法来求得最优化调度策略，即在时隙  $t$  的开始时刻，将时隙  $t$  分配给  $U'_{i,AL}(R_i(t-1)) \cdot C_i(t)$  最大的用户  $i^*$ ，

$$i^* = \arg \max_i (1 + \exp\{-R_i(t-1)\}) \cdot C_i(t) \quad (2.32)$$

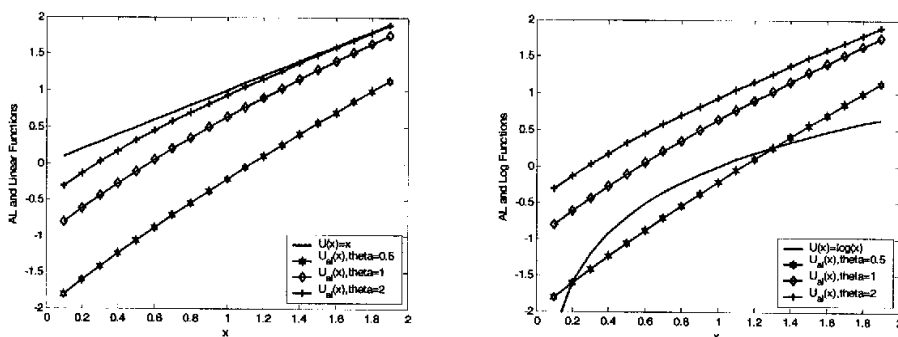
在时隙  $t$  的结束时刻，仍然按照式(2.1)更新各用户的已达速率。

式(2.32)的意义是明显的：它对 Max-C/I 的调度策略作出校正，当用户  $i$  的已达速率较大时，表征信道状况的  $C_i(t)$  决定了该用户的传输机会，类似于 Max-C/I 调度，而当用户  $i$  的已达速率较小时，校正项  $\exp\{-R_i(t-1)\}$  将会增大该用户的传输机会，ATMA 算法正是通过这种方式来保证用户间公平性。

为了对吞吐量的提高程度提供更加灵活的控制,渐近线性函数的表达式(2.30)还可以写成更普适的形式,如下所示,

$$U_{i,AL(\theta)}(R_i(t)) = R_i(t) - \frac{1}{\theta} \exp\{-\theta \cdot R_i(t)\} \quad (2.33)$$

参数 $\theta$ 决定了 $U_{i,AL(\theta)}(R_i(t))$ 函数在纵轴上的截距以及函数曲线与渐近线的靠近程度。取不同的 $\theta$ 值的 $U_{i,AL(\theta)}(x)$ 函数曲线与线性函数 $U(x) = x$ 的关系如图 2.4(a)所示,可见, $U_{i,AL(\theta)}(x)$ 函数曲线位于 $U(x) = x$ 曲线的下方, $\theta$ 值越大则曲线越靠近线性函数曲线。不同 $\theta$ 值下渐近线性函数 $U_{i,AL(\theta)}(x)$ 与对数函数 $U(x) = \log(x)$ 的关系如图 2.4(b)所示,可以看到,取不同的 $\theta$ 值时, $U_{i,AL(\theta)}(x)$ 函数曲线并不一定一直位于 $U(x) = \log(x)$ 的上方(例如,当 $\theta = 0.5$ 时, $U_{i,AL(0.5)}(x)$ 曲线与 $U(x) = \log(x)$ 曲线有两个交点),但是随着 $x$ 的增大,渐近线性函数的值肯定会超过对数函数。



(a)渐近线性函数与线性函数曲线

(b)渐近线性函数与对数函数曲线

图 2.4 渐近线性函数性质

当渐近线性函数写成式(2.33)的形式时,最优化调度策略为:在时隙 $t$ 的开始时刻,将时隙 $t$ 分配给 $U'_{i,AL(\theta)}(R_i(t-1)) \cdot C_i(t)$ 最大的用户 $i^*$ ,

$$i^* = \arg \max_i (1 + \exp\{-\theta \cdot R_i(t-1)\}) \cdot C_i(t) \quad (2.34)$$

当 $\theta \rightarrow \infty$ 时,式(2.34)所表示的调度策略就趋向于 Max-C/I 调度策略,这是与 $U_{i,AL(\theta)}(x)$ 函数在 $\theta$ 越大时越靠近线性函数的性质相符合的。

### 2.3.2 仿真与分析

本节通过仿真来分析 ATMA 算法的性能。仿真工具为 OPNET 和 Matlab。

针对 BE 业务的调度算法性能主要从以下几个方面考察：

(1)  $\sum_i R_i$ ，即系统吞吐量；

(2) 3GPP2 给出的公平性准则<sup>[33]</sup>。针对 BE 业务，3GPP2 在鼓励提出各种新的调度算法的同时，为了衡量各算法公平性，确定了一个定性的准则。该准则用各用户归一化速率的概率分布函数(Cumulative Distribution Function, CDF)来表示，各用户归一化速率按照下式计算，

$$\tilde{R}_i = \frac{R_i}{(\sum_i R_i)/I} \quad (2.35)$$

所有符合公平性准则的调度算法的归一化速率 CDF 曲线必须位于表 2.3 中三个点连线的右侧，否则就是违反了公平性准则。该准则实际上是限制了速率较低的用户占总用户数的比例，比如低于平均速率 0.1 倍的用户数不能超过用户总数的 10%。

表 2.3 CDF 公平性准则

归一化速率	0.1	0.2	0.5
CDF	0.1	0.2	0.5

### 2.3.2.1 仿真条件

总共仿真 90000 个时隙，每个时隙的持续时间为 1ms，每隔 10000 个时隙记录一次各个用户的平均速率。仿真环境如下：

#### (1) 系统参数

单小区单载波系统，基站以满功率向各个用户发送数据，各用户时分复用下行链路，16 个用户在小区内均匀分布，用户队列假设为 Infinitely Backlogged Queues，即假设所有用户一直有数据要传送，不出现空队列的情况。

#### (2) 信道条件

包括小尺度衰落(Rayleigh 衰落)和大尺度衰落，其中 Rayleigh 衰落用 Jakes 模型<sup>[32]</sup>产生。信道仿真主要参数如表 2.2 所示，图 2.5 给出由表 2.2 参数产生的多径 Rayleigh 衰落信道的时域冲激响应，可以看到各径衰落过程互相独立。

表 2.2 信道仿真参数

大尺度衰落 参数	小区半径(km)	1
	衰落系数	4
多径 Rayleigh 衰落参数	Jakes 模型振荡器个数	32
	最大多普勒频移(Hz)	10
	路径数	5
	各径相对功率(dB)	0, -3, -6, -9, -12
其他参数	信噪比(dB)	10
	带宽(kHz)	128

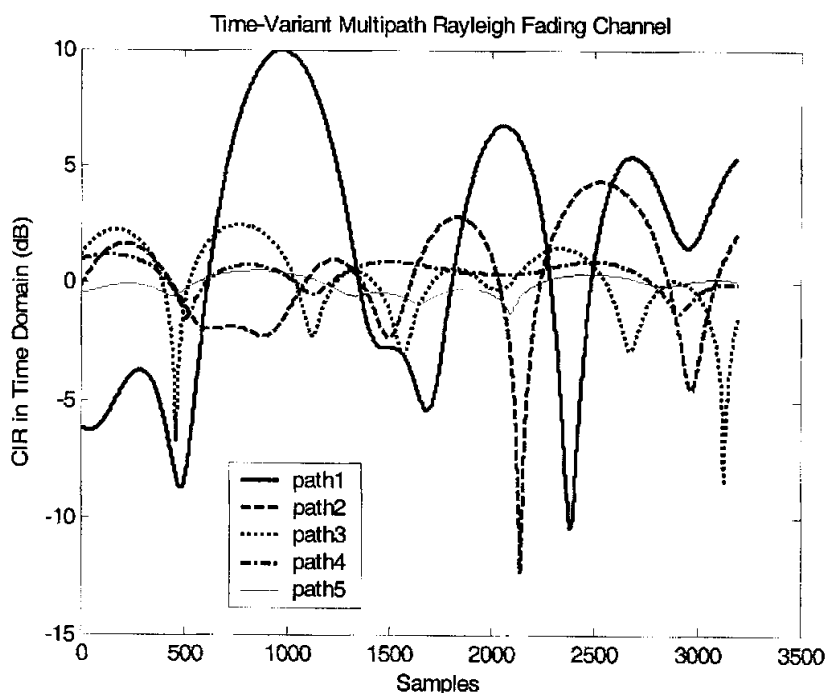


图 2.5 Jakes 模型产生的多径 Rayleigh 衰落信道冲激响应(CIR)

用户通过上行链路向基站报告下一个时隙内的最大可达速率，若第  $t$  时隙用户  $i$  的信道增益为  $h_i(t)$ ，那么用户  $i$  在时隙  $t$  内最大可达速率  $C_i(t)$  为：

$$C_i(t) = \frac{B}{2} \log_2 \left( 1 + \frac{|h_i(t)|^2 P}{N_i \cdot B} \right) \quad (2.36)$$

式中  $B$  是信道带宽， $P$  和  $N_i$  分别为基站发送功率和用户  $i$  的噪声功率谱密度，假

设每个用户噪声功率相同。每时隙采样 32 点，用信道响应幅度最小的采样值 (worst case) 作为  $h_i(t)$  代入式(2.36)计算。

### 2.3.2.2 仿真结果与分析

首先固定  $\theta=1$ ，比较 ATMA 算法与 PF 算法；然后比较  $\theta$  值不同时 ATMA 算法的性能。

$\theta=1$  时<sup>1</sup>，ATMA 算法和 PF 算法下各用户速率的 CDF 曲线如图 2.6 所示。与 PF 算法相比，ATMA 算法下用户速率的范围明显要大。进一步统计可得，此时 ATMA 算法下的吞吐量比 PF 算法下的吞吐量提高了 17%，为 Max-C/I 算法下吞吐量的 88%，吞吐量有较大的提高。

再考察用户间的公平性。根据式(2.35)计算各算法下用户归一化速率，其 CDF 曲线如图 2.7 所示，公平性准则的要求由图中直线表示，可见这两种算法下用户归一化速率的 CDF 曲线都位于公平性准则直线的右侧，也就是说， $\theta=1$  时的 ATMA 算法是满足公平性准则的。

从 2.3.1 节的分析中可知，随着  $\theta$  值的增大，ATMA 趋近于 Max-C/I 算法，但是由于 ATMA 算法最优化目标函数的渐近性质，即使  $\theta$  值很大，它也不会与 Max-C/I 算法等同，保证了用户间的公平性。下面我们用仿真考察  $\theta$  值很大时 ATMA 算法的性能。

取  $\theta$  值分别为 1、2、10、1000。实验结果表明使用 Max-C/I 算法时有些用户因为信道条件一直很差，最终的速率为零，而取这四个  $\theta$  值的 ATMA 算法下均不会有用户速率为零的情况出现。ATMA 算法得到的系统吞吐量，以及 Max-C/I 算法得到的系统吞吐量值在表 2.4 中给出，可见，ATMA 算法下系统吞吐量随着  $\theta$  值的增大而增大，但  $\theta$  值很大时，仍然与 Max-C/I 算法下的吞吐量值有差距，保证了用户间公平性。需要注意的是， $\theta=2$  时的 ATMA 算法仍是满足公平性准则的，但  $\theta=10$  和  $\theta=1000$  时的 ATMA 算法是不满足公平性准则的，应用时不应选用太大的  $\theta$  值。

由此得到结论：ATMA 算法克服了 PF 算法吞吐量牺牲太大的缺陷，合适  $\theta$  值的 ATMA 算法满足 3GPP2 公平性准则，保证了用户间的公平性。

<sup>1</sup> 注意到用户速率  $R_i(t)$  以 bps 为单位时是个很大的数值，直接代入式(2.32)将会导致整个算法几乎等同于 Max-C/I 调度，所以在仿真时， $R_i(t)$  和  $C_i(t)$  分别用一个时隙内传输的平均比特数和最大可达比特数代替。

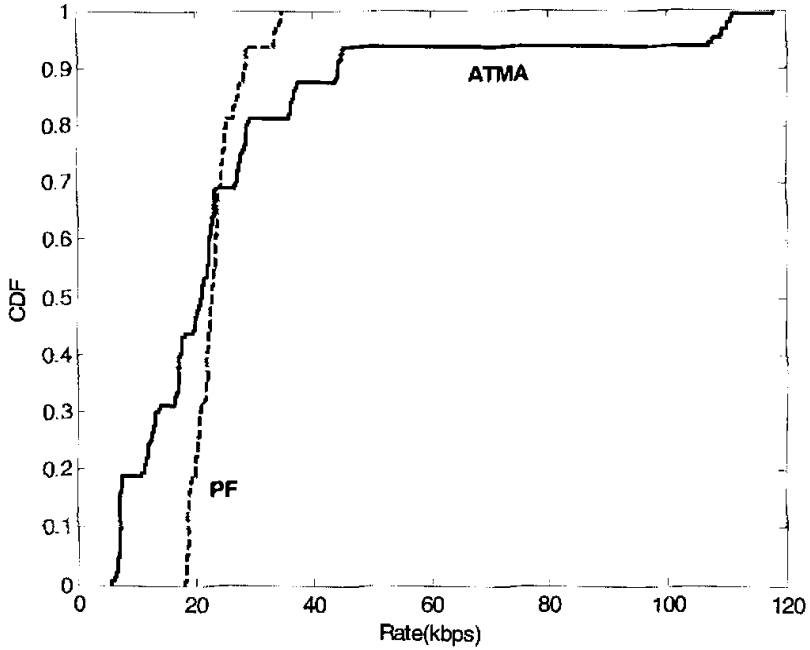
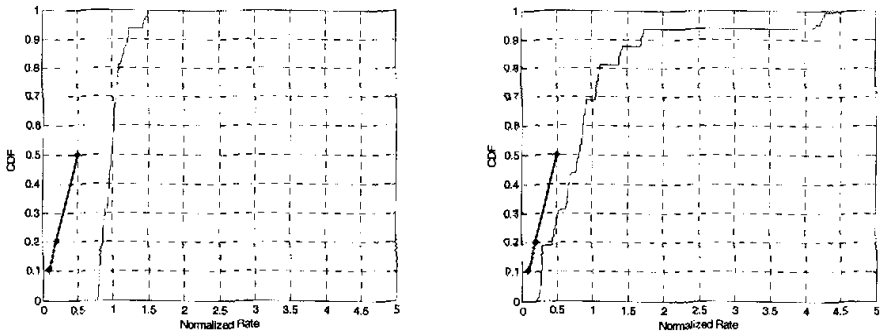


图 2.6 ATMA( $\theta=1$ )算法与 PF 算法下用户速率 CDF 曲线



(a)PF 算法下归一化速率 CDF 曲线 (b)ATMA 算法下归一化速率 CDF 曲线

图 2.7 PF 算法和 ATMA( $\theta=1$ )算法下用户归一化速率 CDF 曲线

表 2.4  $\theta$  值不同的 ATMA 算法与 Max-C/I 算法下的吞吐量

算法	$\sum_i R_i$
ATMA( $\theta=1$ )	3835.7
ATMA( $\theta=2$ )	4087.8
ATMA( $\theta=10$ )	4284.8
ATMA( $\theta=1000$ )	4364.9
Max-C/I	4381.2

## 2.4 针对 NRT 业务的 S-效益最优化算法(S-UO)

在只有最小速率限制时, 经典算法 Penalty 算法利用倒数函数来得到调度策略。本节将分析 Penalty 算法的不足之处, 并提出一种新的算法克服 Penalty 算法的缺陷。

当有最小速率限制时, Penalty 算法的思路是直接将最小速率限制要求写进效益函数, 不像 GMR 算法那样需要令牌计数器, 节省了开销。算法中的效益函数为  $U_i(R_i(t)) = \log(R_i(t) - R_i^{\min})$ , 最优化目标为  $\sum_i \log(R_i(t-1) - R_i^{\min})$ , 调度策略是将时隙  $t$  分配给  $C_i(t) / [R_i(t-1) - R_i^{\min}]$  值最大的用户。

Penalty 算法的缺陷是明显的: 第一, 它的效益函数是以  $R = R_i^{\min}$  为渐近线的, 当用户速率从右侧趋近于  $R_i^{\min}$  时, 效益函数趋于负无穷, 这不符合实际情况, 因为当用户速率正好等于  $R_i^{\min}$  时, 仍然算是满足了速率限制要求的, 而且 Penalty 算法调度策略采用倒数函数, 这就给速率稍稍大于  $R_i^{\min}$  的用户非常高的优先权, 表征信道条件好坏的  $C_i(t)$  在决定时隙的分配时几乎不起作用。第二, Penalty 算法的效益函数在  $R_i^{\min}$  左边没有定义, 算法实现时必须分情况讨论, 优先将时隙分配给速率小于  $R_i^{\min}$  的用户。

为了克服第一个缺陷, 一种方案是利用阶跃函数  $Sgn(R_i(t) - R_i^{\min})$ , 将效益函数写成阶跃函数与  $\log R_i(t)$  函数的乘积<sup>[18]</sup>, 称此方案为阶跃方案。Penalty 算法和阶跃方案的效益函数如图 2.8 所示。

阶跃方案虽然解决了效益函数在  $R_i^{\min}$  处的定义与实际意义问题, 但是阶跃方案效益函数在  $R_i^{\min}$  的右边就是普通的  $\log R_i(t)$  函数, 也就是说, 它没有给速率靠近  $R_i^{\min}$  的用户更多的优先权。

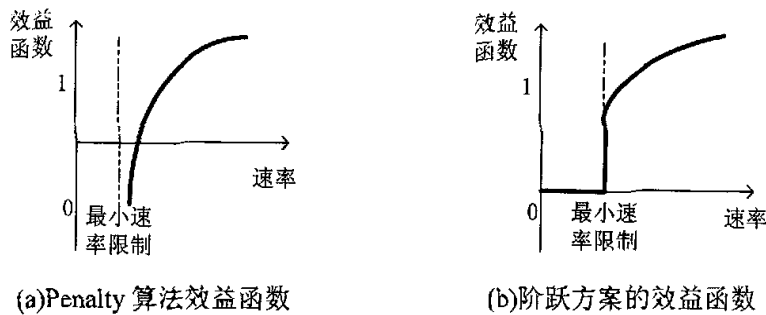


图 2.8 Penalty 算法和阶跃方案的效益函数

## 2.4.1 算法描述

从上面的分析可知, Penalty 算法效益函数在  $R_i^{\min}$  处无定义, 且给速率稍大于  $R_i^{\min}$  的用户的优先权过多, 而阶跃方案虽然解决了效益函数在  $R_i^{\min}$  处的定义问题, 但给速率稍大于  $R_i^{\min}$  的用户的优先权过少, 无法有效保证最小速率限制。

我们利用 S 函数<sup>[34]</sup>, 提出 S-效益最优化(Sigmoid-Utility Optimization, S-UO) 算法。S 函数在各个科学领域都得到了广范的应用, 它有如下的表达式:

$$S(x) = \frac{1}{1 + \exp\{-\theta(x - x_0)\}} \quad (2.37)$$

其中参数  $\theta$  决定了 S 数曲线的斜率,  $\theta$  越大, 曲线越陡峭; 参数  $x_0$  决定了 S 曲线的拐点, 函数在  $x_0$  处的值固定为 0.5, 在  $x_0$  右边为凸函数, 在  $x_0$  的左边为凹函数。下图表示了  $x_0 = 5$ ,  $\theta$  值不同时 S 函数曲线, 可以看到  $\theta$  值为 10 时, S 函数已经很接近阶跃函数了。

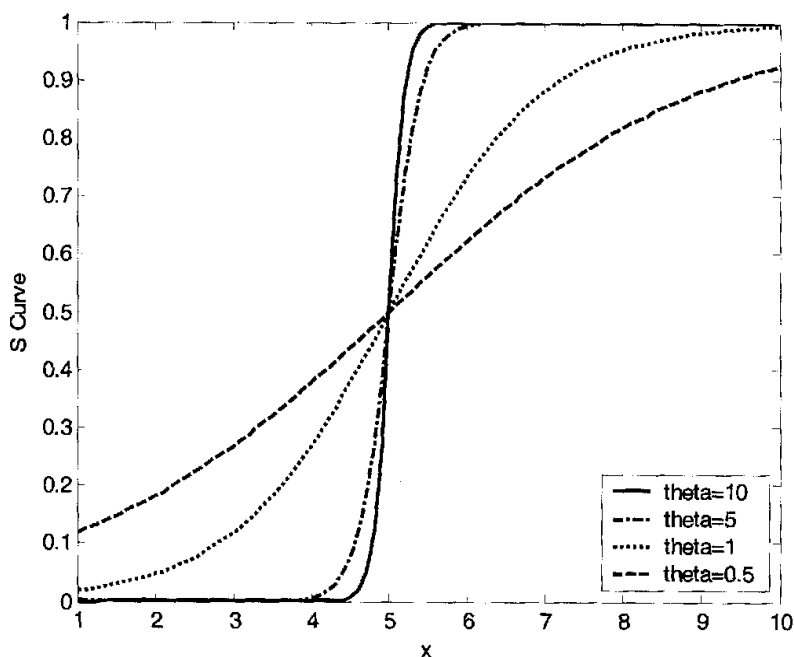


图 2.9  $x_0 = 5$ ,  $\theta$  值不同时 S 函数曲线

利用 S 函数, 将效益函数写成 S 函数与  $\log R_i(t)$  函数相乘的形式, 记为  $U_{i,S}(R_i(t))$ , 其中 S 函数的拐点横坐标设为最小速率限制值, 得到:

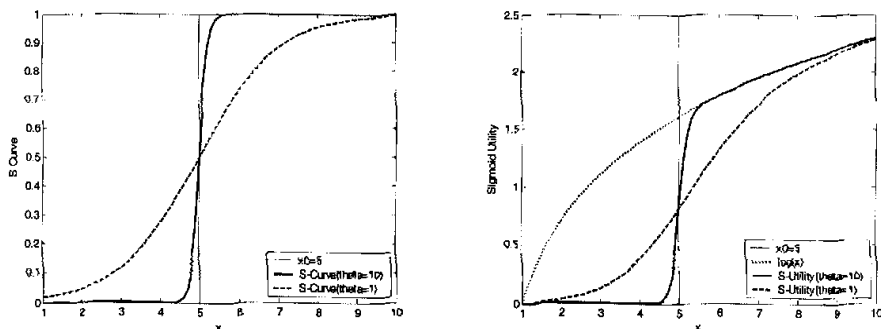


$$U_{i,S}(R_i(t)) = S(R_i(t)) \cdot \log R_i(t) = \frac{1}{1 + \exp\{-\theta(R_i(t) - R_i^{\min})\}} \cdot \log R_i(t) \quad (2.38)$$

用这种方法得到的效益函数  $U_{i,S}(R_i(t))$  实际上就是经过 S 函数调制的对数函数, 我们把它称为 S-效益(Sigmoid-Utility)函数。图 2.10 给出了  $\theta=1$  和  $\theta=10$  时, S 函数曲线与由此得到的 S-效益函数曲线,  $x_0$  值用竖线标出。

首先观察  $\theta=10$  时的 S-效益函数曲线。它在  $R_i^{\min}$  点有定义, 而且它在  $R_i^{\min}$  点右边的斜率介于  $\log(R_i(t) - R_i^{\min})$  函数和  $\log R_i(t)$  函数在此点的斜率之间, 解决了 Penalty 算法效益函数在  $R_i^{\min}$  点处的定义和斜率大小问题。

其次比较  $\theta=1$  和  $\theta=10$  时的 S-效益函数。 $\theta=10$  时的 S-效益函数在  $R_i^{\min}$  点右边斜率较大, 这表示对速率限制能力较强; 而  $\theta=1$  时 S-效益函数变化得很平缓,  $R_i^{\min}$  点右边的斜率比  $\theta=10$  时的曲线此处的斜率小很多, 这将导致 S-UO 算法对速率的限制能力下降, 是使用者所不希望的。应用时  $\theta$  的取值应当使 S-效益函数曲线形状接近阶跃方案效益函数曲线形状。



(a)  $\theta=1$  和  $\theta=10$  时的 S 函数曲线      (b)  $\theta=1$  和  $\theta=10$  时的 S-效益函数曲线

图 2.10  $\theta=1$  和  $\theta=10$  时的 S 函数曲线和 S-效益函数曲线

$U_{i,S}(R_i(t))$  函数是两个连续可导函数的乘积, 所以它也是连续可导的, 但是由于  $U_{i,S}(R_i(t))$  并不是凸函数(它在  $x_0$  右边为凸函数, 在  $x_0$  的左边为凹函数), 从附录 B 的证明过程可以知道, 凸函数在凸集上的有唯一的极值, 局部极值就是全局极值, 从而可以用求导的方法求得最佳调度策略。而定义在凸集上的凹函数很可能有多个局部极值, 用求导方法求得的局部极值仅仅是全局极值的必要而非充分条件。由此, 通过求导的方法求得的调度策略很可能只是一个次优的策略。有很多文献对凹函数在凸集上的极值问题作出研究<sup>[35-37]</sup>, 指出这是个 NP 难(NP-hard)问题。

我们沿用 Penalty 算法实现时使用的分情况讨论的方法, 即在于时隙  $t$  的开始时刻, 优先将时隙分配给速率小于  $R_i^{\min}$  的用户, 而当所有用户的速率都大于  $R_i^{\min}$  时, 将时隙  $t$  分配给  $U_{i,S}(R_i(t-1)) \cdot C_i(t)$  最大的用户  $i^*$ ,

$$i^* = \arg \max_i C_i(t) \cdot \left[ \frac{1}{1+\eta} \cdot \frac{1}{R_i(t-1)} + \theta \log R_i(t-1) \frac{\eta}{(1+\eta)^2} \right] \quad (2.39)$$

式中  $\eta = \exp\{-\theta(R_i(t-1) - R_i^{\min})\}$ 。在时隙  $t$  的结束时刻，仍然按照式(2.1)更新各用户的已达速率。

S-UO 算法与 Penalty 算法相比，解决了效益函数在  $R_i^{\min}$  点处的定义和斜率大小问题。S-UO 算法也存在缺陷，式(2.39)的计算较为复杂，而且与 Penalty 算法一样，实现时需要分情况处理。

## 2.4.2 仿真与分析

本节通过仿真来分析 S-UO 算法的性能。仿真工具为 OPNET 和 Matlab。

针对 NRT 业务的分组调度算法，首先应该考虑怎样满足速率限制的要求，还应当考虑系统吞吐量和用户公平性。本节将从以下几个方面考察算法的性能：

- (1) 最小速率限制要求是否满足，这是衡量算法性能的最重要指标；
- (2) S-UO 和 Penalty 算法均采用了对数函数形式的效益函数作为最优化目标函数，所以，用户速率对数和  $\sum_i \log R_i$  成为衡量算法性能的重要指标；
- (3) 系统吞吐量  $\sum_i R_i$  将作为一个次要的指标。

总共仿真 90000 个时隙，每隔 10000 个时隙记录一次各个用户的平均速率，仿真条件均与 2.3.2 节中的仿真条件相同。

首先比较 S-UO 算法与传统的无速率限制的比例公平调度(PF)算法，验证 S-UO 算法的有效性；然后比较 S-UO 算法和 Penalty 算法；最后研究参数  $\theta$  对算法性能的影响。

PF 算法下各用户速率大概分布在 18kbps 到 36kbps 之间。取  $R_i^{\min} = 20\text{kbps}$ ，并设 S-UO 算法中的参数  $\theta = 10$ 。PF 算法和 S-UO 算法下各用户速率 CDF 曲线如图 2.11 所示。最小速率限制用竖线画出，落在竖线右边表示用户速率满足了最小速率限制要求。可以看到，在 S-UO 算法下，各用户的速率都满足了最小速率限制要求，从而验证了 S-UO 算法的有效性。

下面比较 S-UO 算法与 Penalty 算法。取  $R_i^{\min} = 20\text{kbps}$ ，S-UO 算法和 Penalty 算法下用户速率 CDF 曲线如图 2.12 所示。用竖线标明最小速率限制，可见 S-UO 算法和 Penalty 算法均能使用户速率满足最小速率限制要求。此时这两种算法下的  $\sum_i \log R_i$  值与  $\sum_i R_i$  值如表 2.5 所示，可见，使用 S-UO 算法将得到比 Penalty 算法大的系统效益与系统吞吐量。

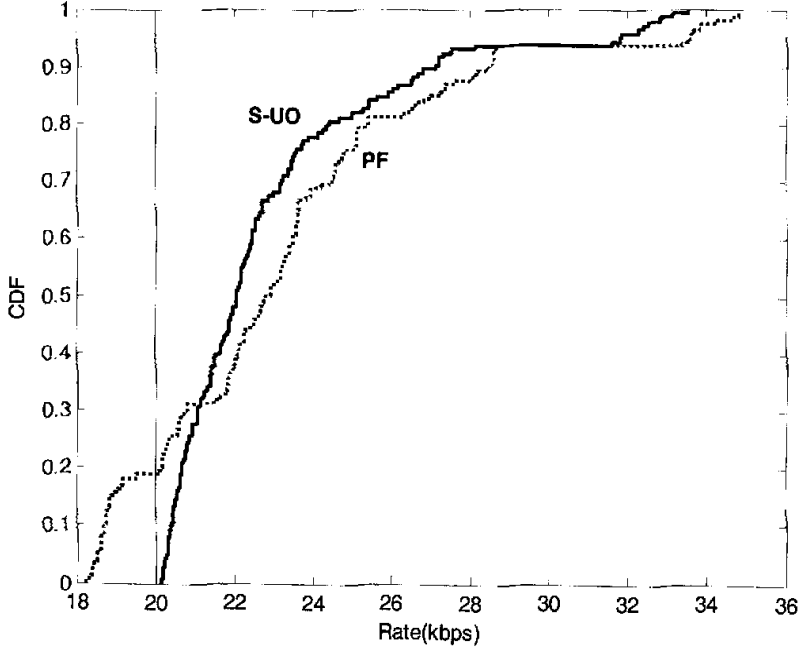


图 2.11 PF 与 S-UO( $R_i^{\min} = 20\text{kbps}$ ,  $\theta = 10$ )算法下用户速率 CDF 曲线

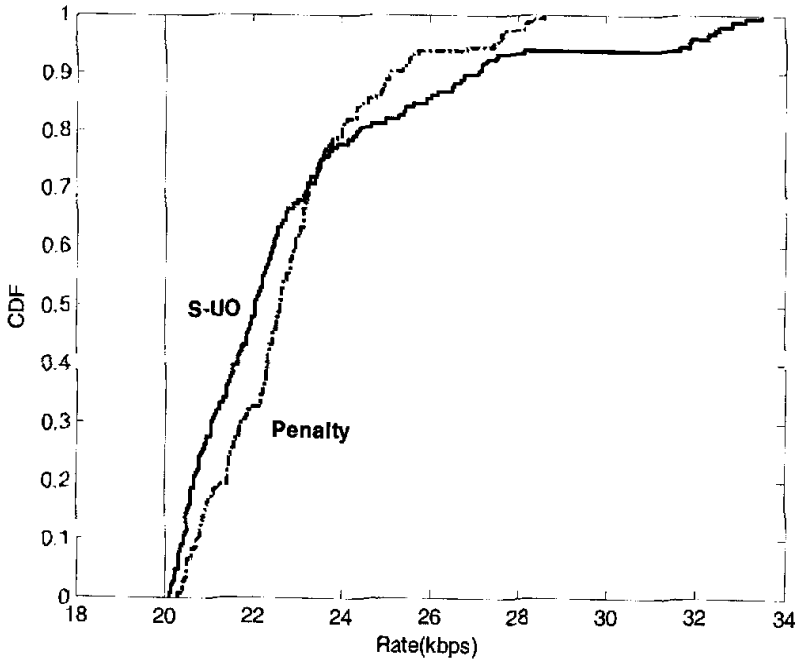


图 2.12  $R_i^{\min} = 20\text{kbps}$  时 Penalty 与 S-UO( $\theta = 10$ )算法下用户速率 CDF 曲线

表 2.5 PF 与 S-UO 算法下系统效益与系统吞吐量

算法	$\sum_i \log R_i$	$\sum_i R_i$	参数设置
S-UO	450.33	3310.9	$R_i^{\min} = 20\text{kbps}$
Penalty	449.14	3256.6	

在 2.4.1 节我们已经分析过 S 函数曲线的特性，S 曲线由两个参数决定，一是参数  $x_0$ ，它决定了 S 曲线的拐点，二是参数  $\theta$ ，它决定了 S 函数曲线的斜率， $\theta$  越大，曲线越陡峭，越接近阶跃函数。从理论分析可知， $\theta$  值越小时，对速率的限制能力就越低，应用时  $\theta$  的取值应当使 S-效益函数曲线形状接近阶跃方案效益函数曲线形状。下面通过仿真来验证这个特性。

取  $R_i^{\min} = 20\text{kbps}$ ， $\theta = 1$  和  $\theta = 10$  时 S-UO 算法下各用户的速率 CDF 曲线如图 2.13 所示。很明显， $\theta = 1$  时的 S-UO 算法不能使所有用户速率都满足最小速率要求，这正是由  $\theta = 1$  时 S-效益函数在拐点的右侧斜率较小，没有给速率仅稍大于  $R_i^{\min}$  的用户更多的优先权所致。

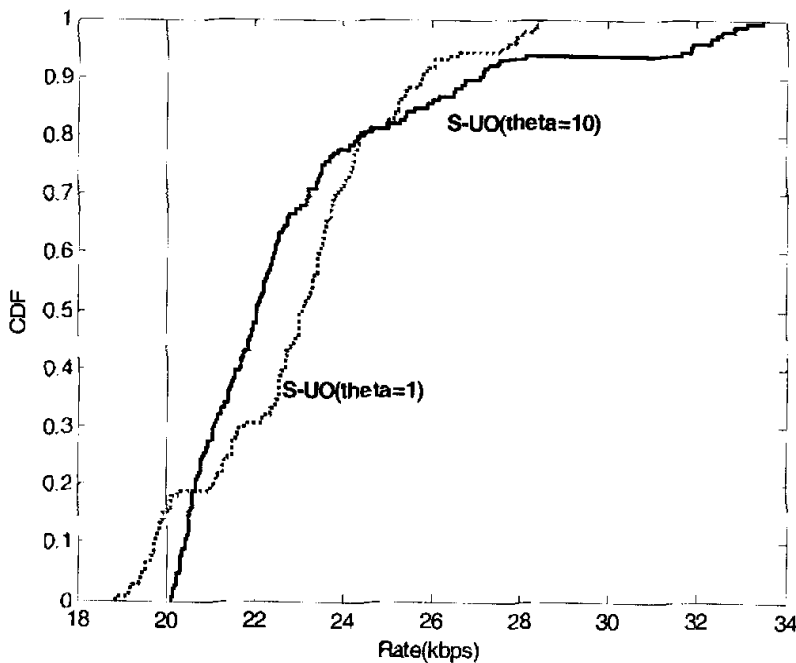


图 2.13  $R_i^{\min} = 20\text{kbps}$  时，不同  $\theta$  值的 S-UO 算法下用户速率 CDF 曲线

## 本章小结

本章主要讨论单载波系统中的分组调度技术。因为 cdma2000 1x EV-DO 系统的结构典型，技术发展成熟，所以本章以 cdma2000 1x EV-DO 为例，详细讨论了单载波系统中的经典调度算法。

针对 BE 业务，经典算法包括最大 C/I 算法、轮询算法、最大-最小算法和比例公平算法。本章重点介绍了比例公平调度算法，并由此引出了效益函数的概念，将这些调度算法都统一到最大化不同形式的系统效益(即各用户的效益函数之和)的框架中。理解效益函数对调度算法是很有益的，因为它给各种调度算法提供了可以量化的目标。

针对 NRT 业务，经典算法 Penalty 算法只能处理只有最小速率限制时的情况，深入研究发现，Penalty 算法是采用将最小速率限制写进最优化目标函数的方法来实现保证最小速率的。较新的 GMR 算法(包括 PFMR, MTMR)则采用了给最优化目标函数加上限制条件的方法，用令牌计数器的方法保证最小/最大速率限制。

本章的后半部分提出了两种新的算法，一种是针对 BE 业务的 ATMA 算法，针对 PF 算法造成较大吞吐量损失的缺陷，该算法采用一种新的、以线性函数为渐近线的效益函数，使得 ATMA 算法下的吞吐量方面要明显高于传统的比例公平算法，并且由于算法的优化目标函数以线性函数为渐近线，所以不会等同于最大 C/I 调度，保证了用户间的公平性。另一种是针对 NRT 业务的 S-UO 算法，为了克服 Penalty 算法效益函数在最小速率限制处无定义，且给速率稍大于最小速率限制的用户过多的优先权，而阶跃方案则给此类用户过少的优先权的缺陷，该算法用 S 函数调制过的对数函数作为效益函数，解决了效益函数在最小速率限制处的定义和斜率大小问题。

## 第三章 多载波系统中的无线分组调度

本章讨论多载波系统中的无线分组调度技术。因为可以有效对抗频率选择性衰落,多载波传输技术近年来得到了广泛的关注与研究,最具代表性的技术就是采用频谱相互重叠的正交子载波的正交频分复用(Orthogonal Frequency Division Multiplexing, OFDM)技术。本章以基于 OFDM 技术的正交频分多址(Orthogonal Frequency Division Multiple Access, OFDMA)系统为研究对象,讨论适用于 OFDMA 系统的无线分组调度技术。

首先简单介绍 OFDM 技术和 OFDMA 系统,讨论 OFDMA 系统中的功率分配问题,指出可以让功率在各个子载波上均分,分组调度只需要考虑时隙内的子载波分配。在此基础上,讨论现有的针对 BE 业务的最大吞吐量调度算法和三种比例公平调度方案。而用户间资源公平和速率公平的调度算法类似于单载波系统中的情况,在此不再给出。

本章的后半部分提出了针对 BE 业务的多载波比例公平调度改进算法,和针对 NRT 业务的有最小/最大速率限制时的多载波比例公平和最大化吞吐量调度算法。

### 3.1 OFDMA 系统简介

本节首先介绍 OFDM 技术和 OFDMA 系统。因为在 OFDMA 系统中,比特、功率、子载波都是可以自适应分配的,资源调度的灵活度要远远大于单载波系统,所以本节对 OFDMA 系统中的功率分配问题进行了简单的讨论,并在最后给出在各个子载波上固定功率,只对比特和子载波进行自适应分配这种方案的理论依据。

#### 3.1.1 OFDM 技术与 OFDMA 系统

##### OFDM 技术

OFDM 技术是多载波传输技术的一种,它的主要思想是用多个频谱相互重叠的正交子载波并行传输数据<sup>[3][38]</sup>。当 OFDM 信号在频率选择性衰落信道中传输时,每个子载波上的数据都只经历了平坦衰落。OFDM 技术具有天生的抗多径衰落的能力;与 CDMA 等系统相比,OFDM 技术的频谱利用率高,大大节省

了频谱资源。这两个优点使得 OFDM 非常适于多径信道下的高数据率通信。

OFDM 的原理虽然很早就被提出,但是由于它对子载波之间的正交性要求非常高,在数字信号处理还不是很发达的年代,用模拟振荡器提供这么多频谱重叠又完全正交的载波频率是很困难的。直到 70 年代 DSP 技术发展起来,文献[38]提出利用 IFFT 和 FFT 分别实现 OFDM 信号的调制与解调,才使 OFDM 技术从理论走向实用。

为了消除多径造成的符号间干扰(Inter-Symbol-Interference, ISI),必须在 OFDM 的符号中插入循环前缀(Cyclic Prefix, CP, 也称保护间隔, Guard Interval, GI),即把原来符号末端一段时间内的波形复制到符号的前面,构成新的符号。插入循环前缀后新符号的波形仍然是连续的,当存在多径时,只要多径的最大时延不超过循环前缀的长度,接收机在 FFT 窗口内“看到”的就是属于同一符号的多径分量而没有引入其它符号的信息,因此不存在 ISI。

### 基于 OFDM 技术的多址方式

在无线通信系统中,无论上行链路还是下行链路都需要使用一种技术来区分不同的用户,即多址技术。多址技术允许多个移动用户共享有限的资源。频分多址(FDMA),时分多址(TDMA),和码分多址(CDMA)是无线通信系统中共享资源的三种主要接入技术。

将 OFDM 技术与传统的多址接入方案如 FDMA 和 TDMA 相结合可以实现多用户 OFDM 系统。OFDM 本身也可既作为调制方案又作为多址接入方案,即 OFDMA 系统。这三种多址接入方案具有不同的原理和特点<sup>[39]</sup>。

OFDM-FDMA 多址接入方案将传输带宽划分成正交的子载波集,将不同的子载波集分配给不同的用户,从而避免了不同用户间的多址干扰(Multiple Access Interference, MAI)。这是一种以频率来区分用户的多址接入方式。

OFDM-TDMA 中,在一段时间内将全部带宽资源(所有子载波)分配给一个特定的用户,所有子载波在一个或几个时隙内为该用户独占,这是一种以时间区分用户的多址接入方式,可看作将全部资源在时间轴上进行分割。

OFDM 是一个宽带系统。对每个用户而言,信号在整个系统带宽上都要经历频率选择性衰落,不同的子载波上的衰落都不相同,对一个用户来说信道响应很差的子载波在另一个用户看来却可能很好。在下行链路中,根据各用户 QoS 的要求,可以采用一种自适应的资源分配策略,用户在不同时间内可以占用不同的子载波,构成 OFDMA 系统。

图 3.1 表示了这几种不同的多址方式。相对于其它几种多址接入方式,OFDMA 方式的优点在于更小的资源分配粒度,从而利用不同用户在不同子载波

上信道响应各不相同的特点实现多用户分集(Multi-user Diversity), 不仅能够对抗时间选择性衰落, 还能对抗频率选择性衰落, 最终达到较高的信道利用率。

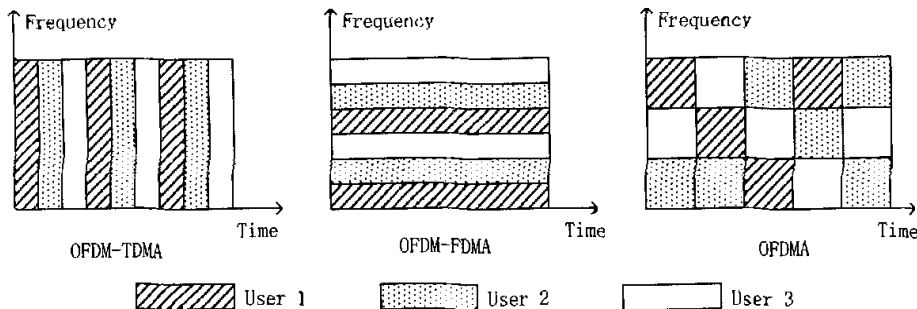


图 3.1 基于 OFDM 的多址接入技术示意图

### 3.1.2 OFDMA 系统中的功率分配问题

与 cdma2000 1x EV-DO 系统相比, OFDMA 系统中的资源分配问题要复杂得多。cdma2000 1x EV-DO 下行链路是时分复用的, 调度器只需要考虑时隙的分配, 不需要考虑功率的分配。而在 OFDMA 系统下行链路中, 调度器不仅需要考虑时隙的分配, 还需要考虑子载波的分配, 以及各个子载波上的功率分配<sup>[40-42]</sup>。

功率是无线通信系统中至关重要的资源之一, 系统的容量直接受限于系统的发射功率。功率分配的问题一般认为属于无线资源管理而不是无线分组调度的范畴<sup>[43]</sup>, 它的理论依据是信息论中经典的灌水法则<sup>[13][44]</sup>, 其基本思想是给优质信道多分配功率给较差信道少分配功率, 给劣质信道不分配功率。本小节从信息论的角度说明, 将总功率在各个子载波上均匀分配, 可以在性能没有很大降低的情况下, 使得整个资源分配问题大为简化。

设有  $J$  个独立并行高斯信道, 噪声功率分别为  $N_1, N_2, \dots, N_J$ 。根据香农信道容量公式,  $J$  个信道的总容量为<sup>[13]</sup>

$$C = \max_p \frac{1}{2} \sum_{j=1}^J \log_2 \left( 1 + \frac{p_j}{N_j} \right) \quad (3.1)$$

其中  $p_j$  是信道  $j$  的信号功率, 满足功率限制  $\sum_{j=1}^J p_j = P$ 。利用拉格朗日乘数法求解, 得到使总容量最大的功率分配为  $p_j = (\gamma - N_j)^+$ , 其中,  $(x)^+$  为正函数,  $\gamma$  为待定常数, 利用  $\sum_{j=1}^J (\gamma - N_j)^+ = P$  确定。因此并行高斯信道的总容量为

$$C = \frac{1}{2} \sum_{j=1}^J \log_2 \left( 1 + \frac{(\gamma - N_j)^+}{N_j} \right) \quad (3.2)$$



从  $p_j = (\gamma - N_j)^+$  可以看出信道的噪声功率越小, 分配的功率越多, 噪声功率越大则分配的功率越少。  $\gamma$  和  $p_j$  可以由灌水(Water Filling)法则求出。将总功率分配给各个子信道时, 首先分配给噪声最小的信道, 然后分配给噪声第二小的信道, 就像灌水一样进行, 直到水涨到一定水平  $\gamma$ , 使总的灌水量等于总功率  $P$ , 噪声功率超过一定值( $\gamma$ )的信道干脆不使用。灌水原理的详细说明参见附录 C。

现在我们讨论功率足够大的情况。此时,  $\sum_{j=1}^J (\gamma - N_j)^+ = P$  退化为

$$\sum_{j=1}^J (\gamma - N_j) = P \quad (3.3)$$

由式(3.3)解得  $\gamma = \frac{1}{J}(P + \sum_{j=1}^J N_j) \approx \frac{P}{J}$ , 则信道容量表达式简化为:

$$C \approx \frac{1}{2} \sum_{j=1}^J \log_2 \left( \frac{P}{JN_j} \right) \quad (3.4)$$

由此可见在高信噪比下, 平均功率分配方法所取得的信道容量与最优的灌水法则取得的信道容量基本相同。因此为了简化资源优化分配算法, 在高信噪比下可以使用简单的平均功率分配策略, 所取得的信道容量接近于最优分配方法。

在 OFDM 系统中, 当 OFDM 的子载波近似相互独立时, 可以用灌水法则将总功率优化分配到各个子载波, 取得最大的信道容量, 推导参见附录 C。附录 C 证明了讨论 BE 业务和 NRT 业务时可以将功率在各个子载波上平均分配, 然后集中考虑子载波分配, 这将大大降低复杂度<sup>[45]</sup>。在本章下面的讨论中, 均假设发射功率在各个子载波上均匀分布。

### 3.1.3 系统模型

系统框图如图 3.2 所示<sup>[39][40]</sup>。总共有  $I$  个用户和  $J$  个子载波,  $i \in \{1, 2, \dots, I\}$ ,  $j \in \{1, 2, \dots, J\}$ , 基站的总发射功率平均分配给各个子载波。用户检测下行链路质量, 由此预测在各个子载波上的最大可达速率, 时隙  $t$  内用户  $i$  在子载波  $j$  上的最大可达速率记为  $C_{i,j}(t)$  (单位为 bps)。调度器通过上行链路知道  $I \times J$  的最大可达速率矩阵。在每个时隙的开始时刻, 调度器根据这个矩阵以及各用户的 QoS 要求等, 决定将每个子载波各分配给哪一个用户。一个子载波只能分配给一个用户, 但一个用户可以占用多个子载波。当所有子载波都分配完毕时, 被分配到子载波的用户就将数据送到物理层, 调制到一个 OFDM 符号中同时发送。

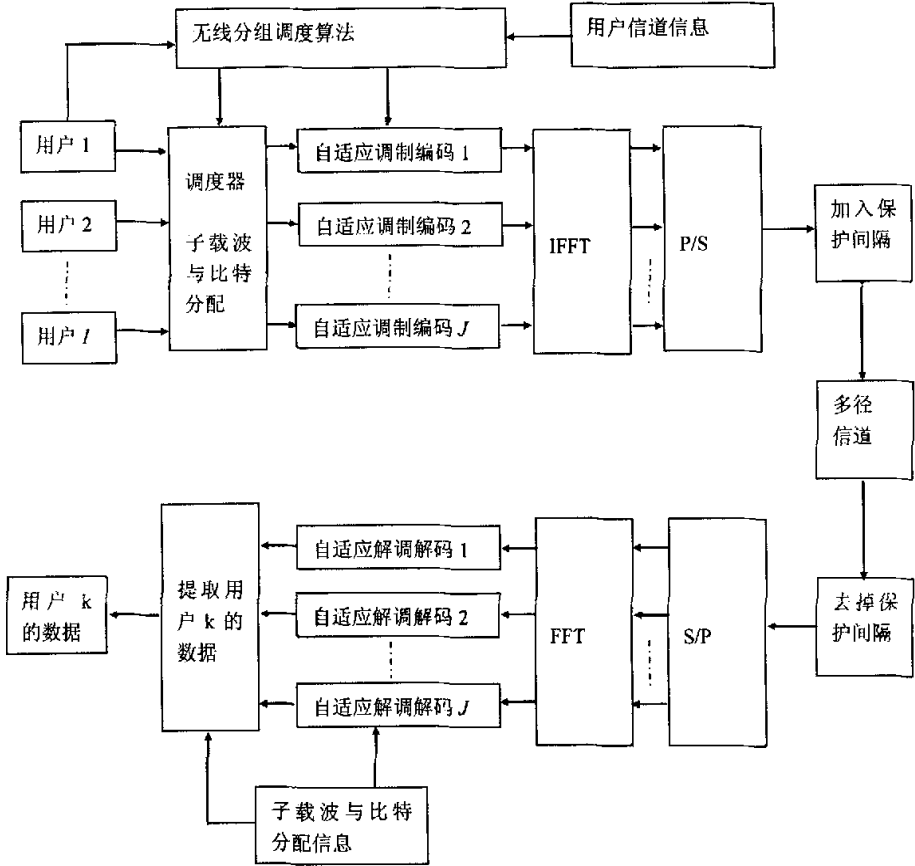


图 3.2 OFDMA 系统框图

假设若子载波  $j$  被分配给了用户  $i$ ，则用户  $i$  在子载波  $j$  上的实际传输速率就是报告的预测最大可达速率  $C_{i,j}(t)$ 。时隙  $t$  内用户  $i$  在子载波  $j$  上实际达到的速率记为  $\mu_{i,j}(t)$  (单位为 bps)，如果子载波  $j$  被分配给了用户  $i$ ，则  $\mu_{i,j}(t)$  等于  $C_{i,j}(t)$ ，否则  $\mu_{i,j}(t)$  为零。仍用  $T_{win}$  表示滑动平均窗口长度，在时隙  $t$  的结束时刻，各用户将其平均速率  $R_i(t)$  由上一个时间窗内的值更新为当前时间窗内的值：

$$\begin{aligned}
 R_i(t) &= \left(1 - \frac{1}{T_{win}}\right)R_i(t-1) + \frac{1}{T_{win}} \sum_j \mu_{i,j}(t) \\
 &= (1 - \beta)R_i(t-1) + \beta \sum_j \mu_{i,j}(t)
 \end{aligned}
 \tag{3.5}$$

式中常数  $\beta = 1/T_{win}$ 。

## 3.2 OFDMA 系统中的无线分组调度

针对 OFDMA 系统 BE 业务的分组调度算法已经有一些研究,算法的目标或者是最大化系统吞吐量,或者是用户间公平,或者是在系统吞吐量和用户公平性之间折中(比例公平)。文献[46]利用单载波系统中的比例公平算法的成果,提出了适用于 OFDMA 系统的三种比例公平调度方案。针对 NRT 业务的最大化吞吐量调度和比例公平调度现在均在研究当中。

### 3.2.1 针对 BE 业务的最大化吞吐量调度算法

当调度的目标为最大化系统吞吐量时,针对 BE 业务的调度算法因为不需要考虑时延或速率要求而变得很简单。文献[47]给出吞吐量最大化算法 TMA(Throughput Maximization Allocation): 调度器在时隙 $t$ 的开始时刻对 $I \times J$ 的最大可达速率矩阵进行全搜索,找到最大的 $C_{i,j}(t)$ 的值,将该子载波分配给相应的用户,再将这个子载波从可分配的子载波集合中删除;对剩下的 $I \times (J-1)$ 的最大可达速率矩阵进行全搜索,反复进行下去直到所有子载波都被分配完毕。这种方法可以达到吞吐量的上界。

### 3.2.2 针对 BE 业务的三种比例公平调度算法

比例公平调度因为在系统吞吐量和用户公平性之间做出较好的折中,在 OFDMA 系统中也得到了重视。尽管固定功率的假设使得 OFDMA 系统的资源分配得到了简化,比例公平仍然是个较难解的问题。比例公平的目标是使得用户速率的对数和最大,在单载波 cdma2000 1x EV-DO 中,一个时隙只能分配给一个用户,所以总共只有 $I$ 种可能的分配方案,只需要比较 $I$ 次即可以得到最优解;而在 OFDMA 系统中,子载波分配方案共有 $I^J$ 种,需要比较的次数非常大。文献[48]用全搜索的方法进行了 OFDMA 系统中比例公平调度算法的初步研究,仿真时采用的载波数目比较少。怎样简化 OFDMA 系统中的比例公平调度算法成为研究热点之一。

研究人员利用单载波 cdma2000 1x EV-DO 系统中比例公平调度的研究成果,针对没有 QoS 要求的 BE 业务,提出了三种适用于 OFDMA 系统的比例公平调度方案<sup>[46][49]</sup>。相对于全搜索的方案,这三种方案都是次优解,但是它们在计算复杂度上要远远低于全搜索方案,现简述如下。

方案一：

每个子载波都有一个单独的调度器，并且各个子载波的调度器是互相独立的。用  $R_{i,j}(t)$  (单位为 bps) 表示时隙  $t$  结束时刻前  $T_{win}$  窗口内，用户  $i$  在子载波  $j$  上达到的平均速率。在时隙  $t$  的开始时刻，依次将子载波  $j$  分配给用户  $i_j^*$ ，

$$i_j^* = \arg \max_i \frac{C_{i,j}(t)}{R_{i,j}(t-1)} \quad (3.6)$$

在时隙  $t$  的结束时刻，按照式(3.7)更新  $R_{i,j}(t)$ ，

$$R_{i,j}(t) = (1 - \beta)R_{i,j}(t-1) + \beta\mu_{i,j}(t) \quad (3.7)$$

算法流程如图 3.3 所示。这种方案实质上将 OFDMA 系统看作是互相独立的各个子载波的集合，在各子载波上独立地做单载波系统的比例公平调度，可以将其称为 PF-per-SC(Proportional Fairness per Sub-Carrier)方案。这种方案只需要搜索  $I \times J$  次，搜索次数远低于全搜索方案。

实际上，OFDMA 系统各个子载波之间并不是完全独立的。用在 OFDMA 系统中，这种方案的性能有待进一步研究。

方案二：

整个系统只有一个中央调度器。在时隙  $t$  的开始时刻，初始化  $\mu_i(t)$ ，

$$\mu_i(t) = 0, \quad i = 1, 2, \dots, I \quad (3.8)$$

依次将子载波  $j$  分配给用户  $i_j^*$ ，

$$i_j^* = \arg \max_i \frac{C_{i,j}(t)}{R_i(t-1)} \quad (3.9)$$

在子载波  $j$  被分配后，立即更新  $\mu_i(t)$ ，

$$\mu_i(t) = \begin{cases} \mu_i(t) + C_{i,j}(t) & \text{if } i = i_j^* \\ \mu_i(t) & \text{if } i \neq i_j^* \end{cases} \quad (3.10)$$

并且在所有子载波都被分配后，按照式(3.5)更新  $R_i(t)$ 。算法流程如图 3.4 所示。

这种方案相当于依次在各个子载波上做比例公平调度，可以将其称为 PF-S(Proportional Fairness-Straightforward)方案。

PF-S 方案在所有子载波都被分配后才做各个用户速率的更新，可能导致调度器“反应过度”，即某用户因为  $R_i(t-1)$  特别小而在时隙  $t$  内持续不断地占用过多的子载波。但是 PF-S 方案一个时隙内只更新一次用户速率，计算复杂度低。

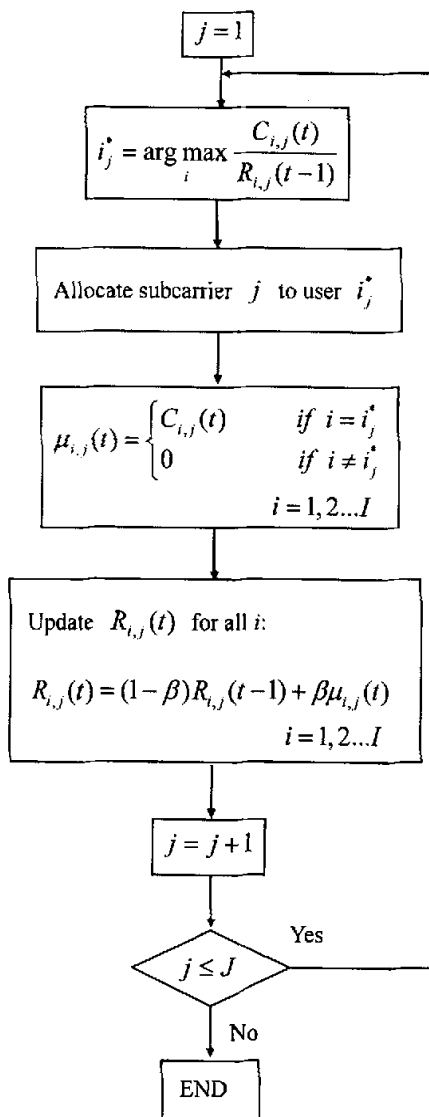


图 3.3 PF-per-SC 方案算法流程图

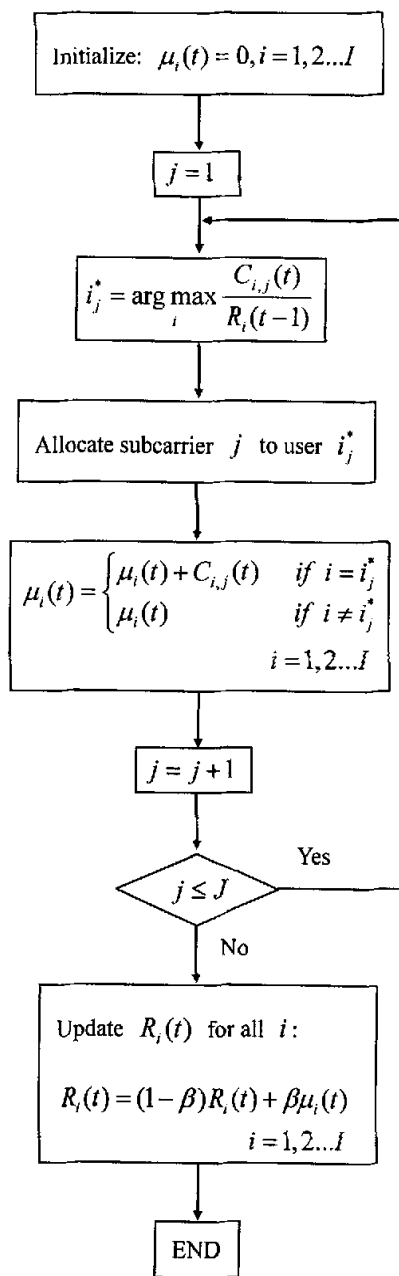


图 3.4 PF-S 方案算法流程图

方案三:

方案三仍然用一个中央调度器。在时隙 $t$ 的开始时刻,依次将子载波 $j$ 分配给用户最大可达速率与用户已达速率比值最大的用户,每分配一个子载波都更新用户已达速率。

用 $R_i^j$ (单位为 bps)表示子载波 $j$ 被分配后用户 $i$ 的已达速率。在时隙 $t$ 的开始时刻,设已达速率初始值

$$R_i^0 = (1 - \beta)R_i(t-1) \quad i = 1, 2, \dots, I \quad (3.11)$$

依次将子载波 $j$ 分配给用户 $i_j^*$ ,

$$i_j^* = \arg \max_i \frac{C_{i,j}(t)}{R_i^{j-1}} \quad (3.12)$$

每个子载波被分配后,立即更新 $R_i^j$ ,

$$\mu_{i,j}(t) = \begin{cases} C_{i,j}(t), & \text{if } i = i_j^* \\ 0, & \text{if } i \neq i_j^* \end{cases} \quad (3.13)$$

$$R_i^j = R_i^{j-1} + \beta \mu_{i,j}(t), \quad i = 1, 2, \dots, I$$

最后令 $R_i(t) = R_i^J$ ,这样就保证了最后的 $R_i(t)$ 与式(3.5)结果一致。整个算法流程如图 3.5 所示。

这种方案考虑到了前面的子载波分配对后面子载波分配的影响,可以将其称为多载波比例公平(Multi-Carrier Proportional Fairness, MC-PF)方案。

### 3.3 针对 BE 业务的比例公平新算法(Modified-MC-PF)

3.2.2 节提到的比例公平调度的三种方案都是依次分配子载波的,搜索次数为 $I \times J$ 次,而由 3.2.1 节中最大化系统吞吐量的 TMA 算法可知,对整个 $I \times J$ 最大可达速率矩阵进行全搜索找到最大值,然后把最大值相对应的子载波分配给对应的用户,能够使系统吞吐量达到上限。

为了提高比例公平调度的系统效益与系统吞吐量,借鉴 TMA 算法的思路,我们提出改进多载波比例公平(Modified Multi-Carrier Proportional Fairness, Modified-MC-PF)算法。

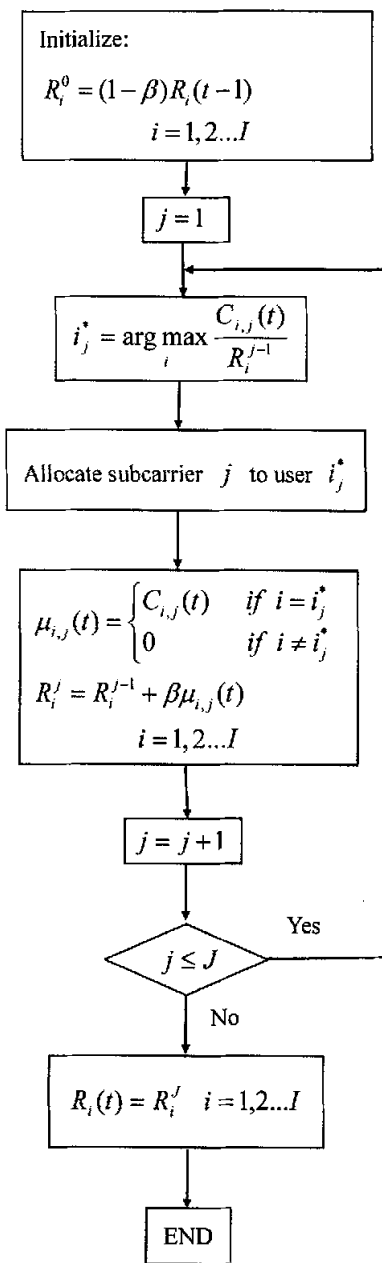


图 3.5 MC-PF 方案算法流程图

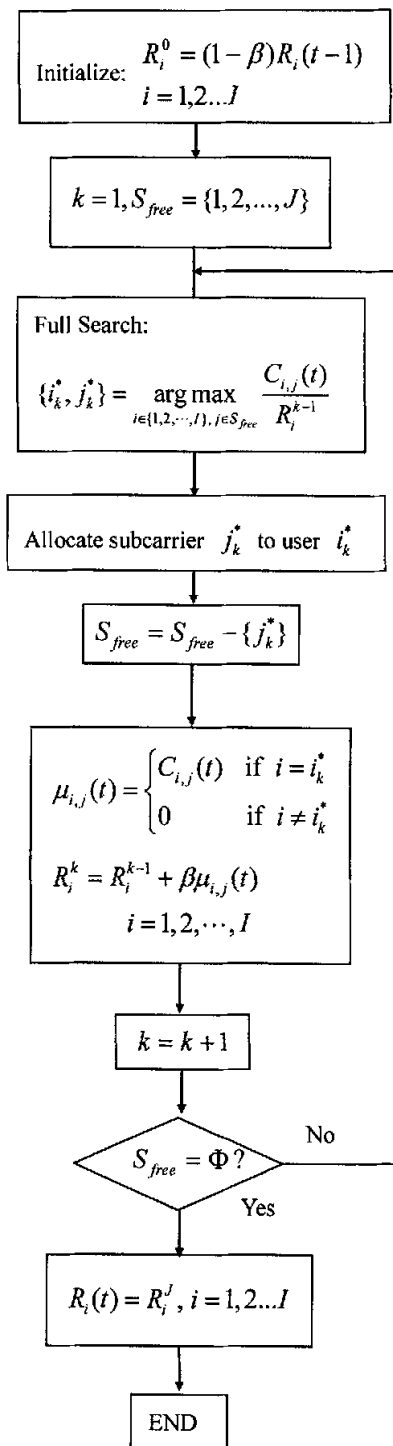


图 3.6 Modified-MC-PF 方案算法流程图

### 3.3.1 算法描述

用  $k$  来记录目前进行的是第几次分配,  $1 \leq k \leq J$ ,  $S_{free}$  为尚未被分配的子载波序号的集合,  $S_{free} \subseteq \{1, 2, \dots, J\}$ ,  $R_i^k$  (单位 bps) 表示第  $k$  次分配后用户  $i$  的已达速率。在时隙  $t$  的开始时刻, 初始化  $R_i^0$  与  $k$ 、 $S_{free}$ ,

$$R_i^0 = (1 - \beta)R_i(t-1) \quad i = 1, 2, \dots, I \quad (3.14)$$

$$k = 1 \quad (3.15)$$

$$S_{free} = \{1, 2, \dots, J\} \quad (3.16)$$

计算用户最大可达速率与已达速率之比的矩阵, 并搜索这个矩阵, 找出最大值,

$$\{i_k^*, j_k^*\} = \arg \max_{i \in \{1, 2, \dots, I\}, j \in S_{free}} \frac{C_{i,j}(t)}{R_i^{k-1}} \quad (3.17)$$

将子载波  $j_k^*$  分配给用户  $i_k^*$ , 并把子载波  $j_k^*$  从  $S_{free}$  集合中移除。  $k+1$ , 按照下式计算  $\mu_{i,j}(t)$  和更新  $R_i^k$ ,

$$\mu_{i,j}(t) = \begin{cases} C_{i,j}(t) & \text{if } i = i_k^* \\ 0 & \text{if } i \neq i_k^* \end{cases} \quad (3.18)$$

$$R_i^j = R_i^{j-1} + \beta \mu_{i,j}(t)$$

重复这个搜索过程, 直到  $S_{free}$  为空集, 最后令  $R_i(t) = R_i^j$ 。整个算法流程如图 3.6 所示。

### 3.3.2 仿真与分析

本节对提出的改进多载波比例公平(Modified-MC-PF)算法和已有的三种算法(PF-per-SC, PF-S, MC-PF)进行仿真和性能比较。仿真工具为 OPNET 和 Matlab。

各个算法性能将由下面几个方面来考察:

- (1) 系统效益  $\sum_i \log R_i$ ;
- (2) 系统吞吐量  $\sum_i R_i$ , 次要的指标;
- (3) 3GPP2 定义的公平性准则, 具体定义参见 2.3.2 节。

总共仿真 90000 个时隙, 每个时隙的持续时间为 1ms, 每隔 10000 个时隙记录一次各个用户的平均速率。依次仿真用户数为 8、16、32 和 64 的情况。仿真环境如下:

- (1) 系统参数: 单小区的 OFDMA 系统, 共有 32 个子载波, 用户在小区内均匀分



布，用户队列假设为 Infinitely Backlogged Queues。

(2)信道条件：包括小尺度衰落(Rayleigh 衰落)和大尺度衰落，其中 Rayleigh 衰落用 Jakes 模型<sup>[32]</sup>产生，信道参数如表 2.2 中所示。

各个用户的时域信道响应产生后，利用 FFT 得到该用户在各个子载波上的频域信道响应，FFT 点数即为子载波的个数。

图 3.7 给出用户数不同时，Modified-MC-PF 算法下用户速率 CDF 曲线。可见，子载波数固定、用户数增加时，各个用户达到的平均速率明显减少，这是符合实际的。在其他三个算法下，也有类似的规律。

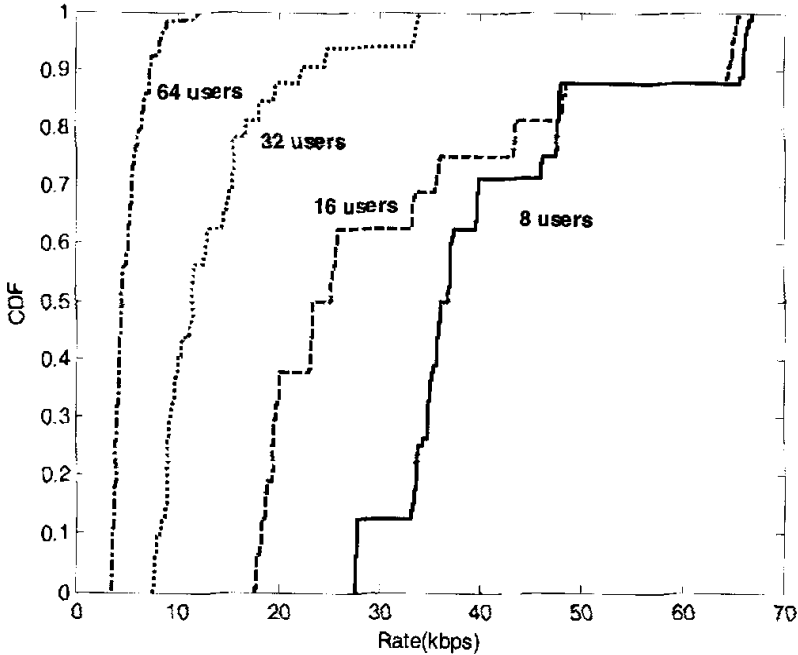


图 3.7 Modified-MC-PF 算法下用户数目不同时用户速率 CDF 曲线

表 3.1 给出这四种算法在不同用户数目下  $\sum_i \log R_i$  与  $\sum_i R_i$  的值，其中用户数目与  $\sum_i \log R_i$  的关系在图 3.8 中画出。从图和表中可见：

第一，随着用户数的增加， $\sum_i \log R_i$  是单调增加的。这表明各个算法都在最大化用户速率对数和的意义上实现了多用户分集。

第二，系统吞吐量  $\sum_i R_i$  并没有随着用户数的增加而单调增加，原因是用户数目比较多时，可能有好几个用户处于小区边缘，调度器要保证这些用户有一定的传输速率，为了公平性牺牲了吞吐量，导致整个系统吞吐量下降。这从另一方面也说明了比例公平算法的目标是最大化用户速率对数和，而不是最大化系统吞吐量。

第三，从表中数据可以看到，各个算法性能之间的差别并不是很大。性能从好到差的算法顺序为 Modified-MC-PF、MC-PF、PF-S、PF-per-SC。用户数目增加时，各算法之间的差别越来越大。

表 3.1 用户数不同时，不同算法下的系统效益与系统吞吐量

用户数	算法	$\sum_i \log R_i$	$\sum_i R_i$
8	PF-per-SC	274.35	3359.56
	PF-S	275.24	3396.26
	MC-PF	275.31	3399.67
	Modified-MC-PF	275.33	3404.46
16	PF-per-SC	480.15	4482.79
	PF-S	481.18	4492.74
	MC-PF	481.55	4504.36
	Modified-MC-PF	482.59	4521.48
32	PF-per-SC	722.71	3859.07
	PF-S	730.28	3956.12
	MC-PF	731.73	3976.14
	Modified-MC-PF	732.86	3992.01
64	PF-per-SC	895.24	2839.79
	PF-S	924.23	2975.68
	MC-PF	927.94	2998.24
	Modified-MC-PF	929.11	3006.19

图 3.9 给出用户数为 16 时，各算法下用户归一化速率的 CDF 曲线，公平性准则的要求由图中直线标出。可以看到这四种算法下用户归一化速率的 CDF 曲线都位于公平性准则直线的右侧，也就是说，这四种算法都满足了公平性准则。

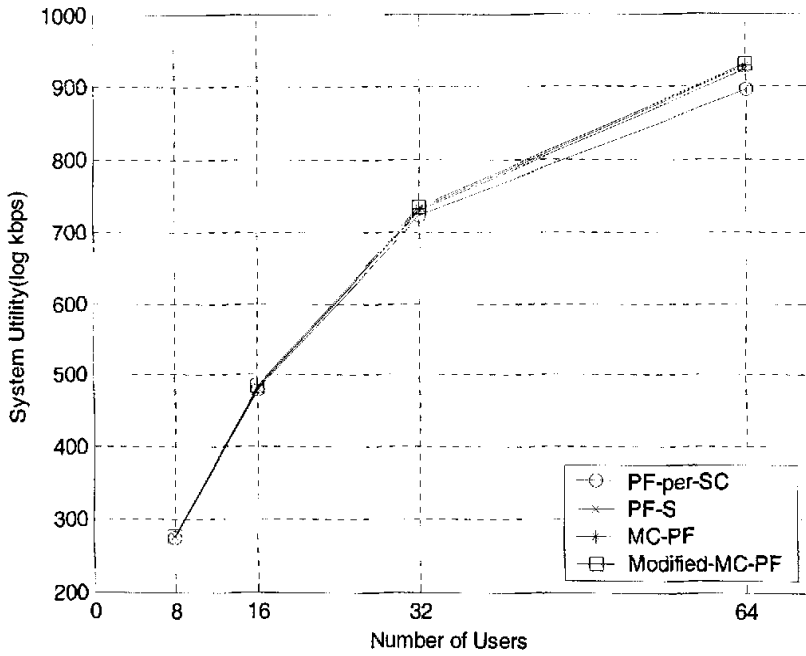


图 3.8 不同算法下系统效益与用户数目关系曲线

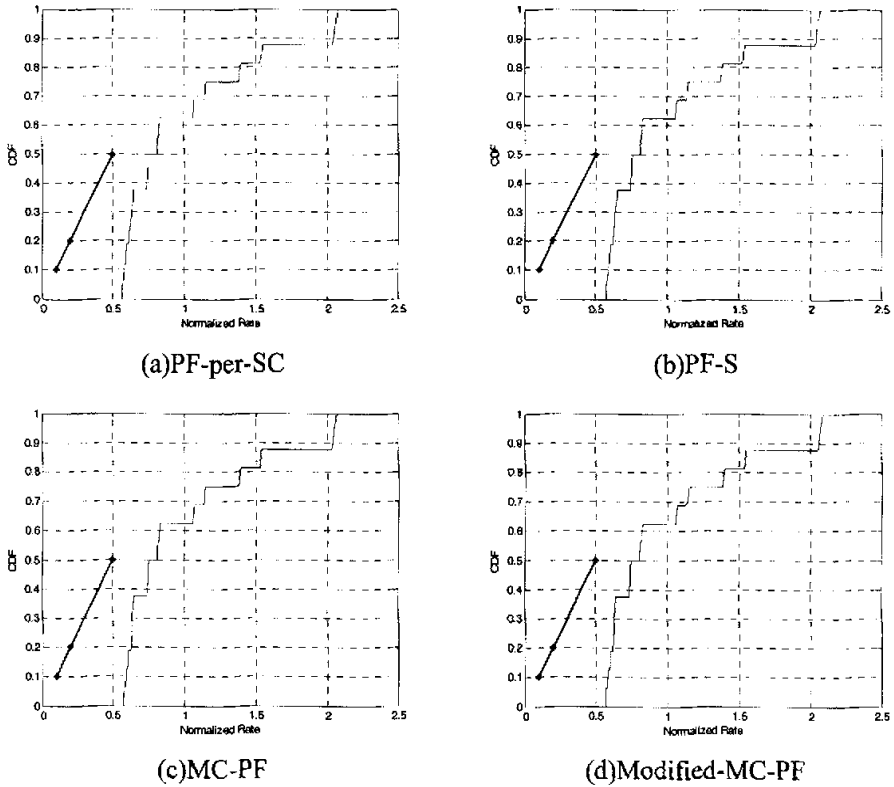


图 3.9 不同算法下用户归一化速率 CDF 曲线与公平性准则的比较(用户数为 16)

### 3.4 针对 NRT 业务的比例公平新算法(MC-PFMR)

OFDMA 系统中保证最小/最大速率限制的比例公平调度算法迄今尚未见公开报道。在 OFDMA 系统中，针对 NRT 业务比例公平调度的目标仍然是在有速率限制的条件下使得用户速率的对数和最大，重写如下：

$$\begin{aligned} \max \sum_i \log R_i(t) \\ \text{Subject to: } R_i^{\min} \leq R_i(t) \leq R_i^{\max} \end{aligned} \quad (3.19)$$

其中  $R_i^{\min}$  和  $R_i^{\max}$  分别表示用户  $i$  的最小、最大速率限制。

与针对 BE 业务的比例公平调度算法类似，这个问题的求解也可以通过全搜索与全比较来实现，即先列举全部  $I'$  种分配方案，去掉其中不能让用户速率满足最小/最大速率限制的方案，再在剩下的方案中做全比较。显然，这种方法的实现复杂度非常高。由于单载波系统中的 PFMR 算法已经从理论上被证明可以达到目标(3.19)，并且仿真结果也证明了此算法的优异性能<sup>[29]</sup>，我们利用 PFMR 算法的思路与方法，提出一种适用于 OFDMA 系统的有最小/最大速率限制时的比例公平调度(Multi-Carrier Proportional Fairness with Minimum/Maximum Rate constraints)算法。

最直接的方法是给每个用户一个令牌计数器，然后依次将子载波  $j$  分配给  $\exp\{\alpha T_i(t-1)\} \cdot [C_{i,j}(t) / R_i(t-1)]$  值最大的用户，在所有子载波被分配完毕后，更新所有用户的令牌数与已达速率。这种算法与针对 BE 业务的 PF-S 算法类似，是在所有子载波被分配后才更新用户的已达速率和令牌数，我们将这种算法称为 PFMR-S(PFMR-Straightforward)算法。3.2.3 节已经指出 PF-S 算法可能导致调度器“反应过度”。这个问题在有速率限制的时候表现得更为明显，即可能某一用户前面的几个时隙内都没有得到传输机会，以致于在时隙  $t$  的开始时刻，他的优先级很高，前面几个子载波都分配给了该用户，该用户已经能够在时隙  $t$  的结束时刻达到最小速率限制要求，然而由于用户已达速率和令牌数都没有更新，后面的子载波很可能还是接连不断的被他占用，不仅其他用户失去传输机会，甚至可能导致该用户在时隙  $t$  结束时刻的速率超过最大速率限制。

为了克服上述缺点，我们引入时隙  $t$  内用户  $i$  在子载波  $j$  上的最小/最大速率限制，分别记为  $R_{i,j}^{\min}(t)$  (单位 bps) 和  $R_{i,j}^{\max}(t)$  (单位 bps)，且满足：

$$\begin{aligned} \sum_j R_{i,j}^{\min}(t) &= R_i^{\min} \\ \sum_j R_{i,j}^{\max}(t) &= R_i^{\max} \end{aligned} \quad (3.20)$$

这样，当用户  $i$  在时隙  $t$  内在子载波  $j$  上满足了  $R_{i,j}^{\min}(t)$  和  $R_{i,j}^{\max}(t)$  的速率限制时，滑动窗口内的长时间平均速率值  $R_i(t)$  就满足了  $R_i^{\min}$  和  $R_i^{\max}$  的速率限制。为了避

免调度器“反应过度”，我们又借鉴 MC-PF 算法的思路，在每个子载波被分配后都更新所有用户的已达速率和令牌数。

由要求的  $R_i^{\min}$  和  $R_i^{\max}$  分别确定  $R_{i,j}^{\min}(t)$  和  $R_{i,j}^{\max}(t)$  时，一种简单的方法是把  $R_i^{\min}$  和  $R_i^{\max}$  在  $J$  个子载波上进行均分，即令  $R_{i,j}^{\min}(t) = R_i^{\min} / J$ ， $R_{i,j}^{\max}(t) = R_i^{\max} / J$ 。这种简单均分的方法并没有充分利用 OFDMA 系统不同用户在不同子载波上信道响应各不相同的特点。对于用户  $i$  而言，如果信道条件相对较好的子载波上  $R_{i,j}^{\min}(t)$  和  $R_{i,j}^{\max}(t)$  值较大，信道条件相对较差的子载波上  $R_{i,j}^{\min}(t)$  和  $R_{i,j}^{\max}(t)$  值较小，将会使子载波上的速率限制要求更容易满足，从而在不改变整体速率限制  $R_i^{\min}$  和  $R_i^{\max}$  的同时，提高系统效益和系统吞吐量。

为了计算合适的  $R_{i,j}^{\min}(t)$  和  $R_{i,j}^{\max}(t)$  值，引进反映用户  $i$  在子载波  $j$  上的相对信道状况的加权系数  $\lambda_{i,j}(t)$ ，

$$\lambda_{i,j}(t) = \frac{C_{i,j}(t)}{\sum_j C_{i,j}(t)} \quad (3.21)$$

$\lambda_{i,j}(t)$  值越大表示相对信道状况越好。由  $\lambda_{i,j}(t)$  计算  $R_{i,j}^{\min}(t)$  和  $R_{i,j}^{\max}(t)$  如下，

$$\begin{aligned} R_{i,j}^{\min}(t) &= \lambda_{i,j}(t) R_i^{\min} \\ R_{i,j}^{\max}(t) &= \lambda_{i,j}(t) R_i^{\max} \end{aligned} \quad (3.22)$$

调度器在时隙  $t$  的开始时刻依次为各子载波分配用户，并在每个子载波被分配后都更新用户已达速率及令牌数。

### 3.4.1 算法描述

设  $R_i^j$  和  $T_i^j$  分别表示子载波  $j$  被分配后用户  $i$  的已达速率和令牌数。用  $\exp\{\alpha_i T_i^{j-1}\}$  项反映子载波上最小/最大速率限制要求，用  $C_{i,j}(t) / R_i^{j-1}$  项反映用户间比例公平的要求，MC-PFMR 算法流程如图 3.10 所示。在时隙  $t$  的开始时刻，用户  $i$  的已达速率和令牌数初始值设为

$$R_i^0 = (1 - \beta) R_i(t-1) \quad i = 1, 2, \dots, I \quad (3.23)$$

$$T_i^0 = T_i(t-1) \quad i = 1, 2, \dots, I \quad (3.24)$$

再依次将子载波  $j$  分配给用户  $i_j^*$ ，

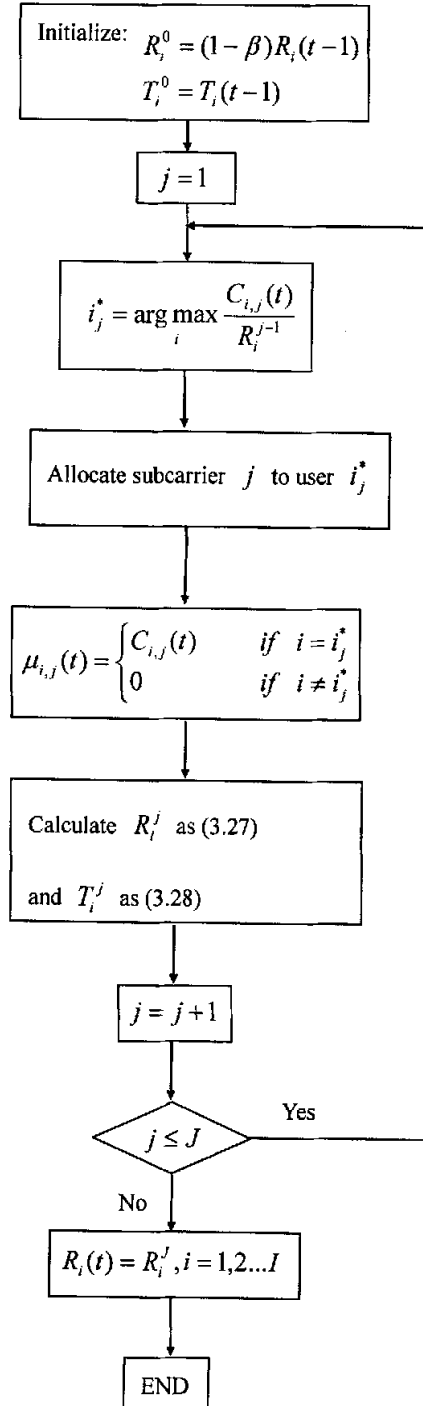


图 3.10 MC-PFMR 算法流程图

$$i_j^* = \arg \max_i \exp\{a_i T_i^{j-1}\} \frac{C_{i,j}(t)}{R_i^{j-1}} \quad j=1,2,\dots,J \quad (3.25)$$

在每个子载波被分配后，都更新所有用户的速率及令牌数。第  $j$  个子载波被分配后，用户  $i$  的已达速率  $R_i^j$ ，令牌数  $T_i^j$  以及在子载波  $j$  上实际达到的速率  $\mu_{i,j}(t)$  可分别计算如下，

$$\mu_{i,j}(t) = \begin{cases} C_{i,j}(t) & \text{if } i = i_j^* \\ 0 & \text{if } i \neq i_j^* \end{cases} \quad i=1,2,\dots,I \quad (3.26)$$

$$R_i^j = R_i^{j-1} + \beta \mu_{i,j}(t) \quad i=1,2,\dots,I \quad (3.27)$$

$$T_i^j = \begin{cases} 0, & \text{if } R_i^{\min} = 0 \text{ and } R_i^{\max} = \infty \\ \max\{0, T_i^{j-1} + R_{i,j}^{\min}(t) - \mu_{i,j}(t)\}, & \text{if } R_i^{\min} > 0 \text{ and } R_i^{\max} = \infty \\ \min\{0, T_i^{j-1} + R_{i,j}^{\max}(t) - \mu_{i,j}(t)\}, & \text{if } R_i^{\min} = 0 \text{ and } R_i^{\max} < \infty \\ T_i^{j-1} + R_{i,j}^{\min}(t) - \mu_{i,j}(t), & \text{if } 0 < R_i^{\min} < R_i^{\max} < \infty \text{ and } T_i^{j-1} \geq 0 \\ T_i^{j-1} + R_{i,j}^{\max}(t) - \mu_{i,j}(t), & \text{if } 0 < R_i^{\min} < R_i^{\max} < \infty \text{ and } T_i^{j-1} < 0 \end{cases} \quad (3.28)$$

时隙  $t$  结束时刻的  $R_i(t)$  和  $T_i(t)$  分别为  $R_i^j$  和  $T_i^j$ 。

MC-PFMR 算法中令牌按照式(3.28)更新，保留了 PFMR 算法的关键技术——令牌更新规则，保证了  $\mu_{i,j}(t)$  小于  $R_{i,j}^{\min}(t)$  时令牌数增加， $\mu_{i,j}(t)$  大于  $R_{i,j}^{\max}(t)$  时令牌数减少， $\mu_{i,j}(t)$  介于  $R_{i,j}^{\min}(t)$  与  $R_{i,j}^{\max}(t)$  之间时，若  $T_i^{j-1}$  为正，则令牌数减少，若  $T_i^{j-1}$  为负，则令牌数增加，从而使  $T_i^j$  保持在零值附近。当没有速率限制时，MC-PFMR 算法退化为 MC-PF 算法。

### 3.4.2 算法的几种变形

根据子载波分配顺序、 $\lambda_{i,j}(t)$  的计算方法不同，MC-PFMR 算法还有几种变形算法，简述如下。

#### Modified-MC-PFMR 算法

与 Modified-MC-PF 算法类似，使用全搜索的方法找到最大值，

$$\{i_k^*, j_k^*\} = \arg \max_{i \in \{1,2,\dots,I\}, j \in S_{free}} \exp\{a_i T_i^{k-1}\} \frac{C_{i,j}(t)}{R_i^{k-1}} \quad (3.29)$$

将最大值对应的子载波  $j^*$  分配给对应的用户  $i^*$ ，并在每次子载波分配后都更新用户已达速率和令牌数。

### Exponent-MC-PFMR 算法

在计算  $\lambda_{i,j}(t)$  时, 让  $R_{i,j}^{\min}(t)$  和  $R_{i,j}^{\max}(t)$  在信道条件较差的子载波上的值进一步减小, 以期提高系统效益与系统吞吐量。用指数律的方法计算  $\lambda_{i,j}(t)$  如下:

$$\lambda_{i,j}(t) = \frac{[C_{i,j}(t)]^n}{\sum_j [C_{i,j}(t)]^n} \quad (3.30)$$

其中  $n$  为无量纲的可调控制参数, 这样得到的  $R_{i,j}^{\min}(t)$  和  $R_{i,j}^{\max}(t)$  仍然满足所有子载波上的和分别为  $R_i^{\min}$  和  $R_i^{\max}$ 。算法在每次子载波分配后都更新用户已达速率和令牌数。

很容易看出, 当  $n=0$  时, 就是整体速率限制在各个子载波上均分的策略; 当  $n=1$  时, 就是 MC-PFMR 算法; 而当参数  $n$  变得较大 ( $n>1$ ) 时, 信道条件较差的子载波上  $R_{i,j}^{\min}(t)$  和  $R_{i,j}^{\max}(t)$  进一步减小。参数  $n$  还可以根据信道的状况进行自适应的调整。

### 3.4.3 仿真与分析

算法性能由以下几个方面来考察:

- (1) 各个用户的速率限制要求是否得到满足, 这是衡量算法性能的最重要指标;
- (2) 系统效益  $\sum_i \log R_i$ ;
- (3) 系统吞吐量  $\sum_i R_i$ , 作为一个次要的指标;

本节固定用户数目为 16, 其他仿真条件与 3.3.2 节的仿真条件类似, 单小区的 OFDMA 系统, 共有 32 个子载波, 16 个用户在小区内均匀分布; 信道包括 Rayleigh 衰落和大尺度衰落两种衰落, 参数由表 2.2 给出。

首先仿真只有最小速率限制时的情况, 这在实践中非常常见, 并且有很多已有算法可以进行比较。第一步先给出针对 BE 业务的 MC-PF 算法与只有最小速率限制时的 MC-PFMR 算法的比较, 讨论为满足速率限制付出的代价; 第二步给出 MC-PFMR 算法与已有算法的比较。

然后仿真最小/最大速率限制同时存在, 且各个用户速率限制值不相同时的情况, 分析 MC-PFMR 算法在这种情况下性能。

#### MC-PF 算法与 MC-PFMR 算法比较

3.3.2 节的仿真可见, 当用户数为 16 时, 用户速率大概分布在 17kbps 到 65kbps



之间。取最小速率限制  $R_i^{\min} = 22\text{kbps}$ ，MC-PFMR 算法和 MC-PF 算法下用户速率的 CDF 曲线如图 3.11 所示。竖线标出最小速率限制，落在竖线的右边表示速率大于最小速率限制。这两种算法下的系统效益和系统吞吐量在表 3.2 中给出。

从图中可以看到，MC-PFMR 算法几乎所有用户速率都能达到要求，证明了 MC-PFMR 算法的有效性。从表中可以看到，MC-PFMR 算法下的系统效益和系统吞吐量都比 MC-PF 算法得到的值小，这表明最小速率限制的满足是以牺牲系统效益和系统吞吐量为代价的。

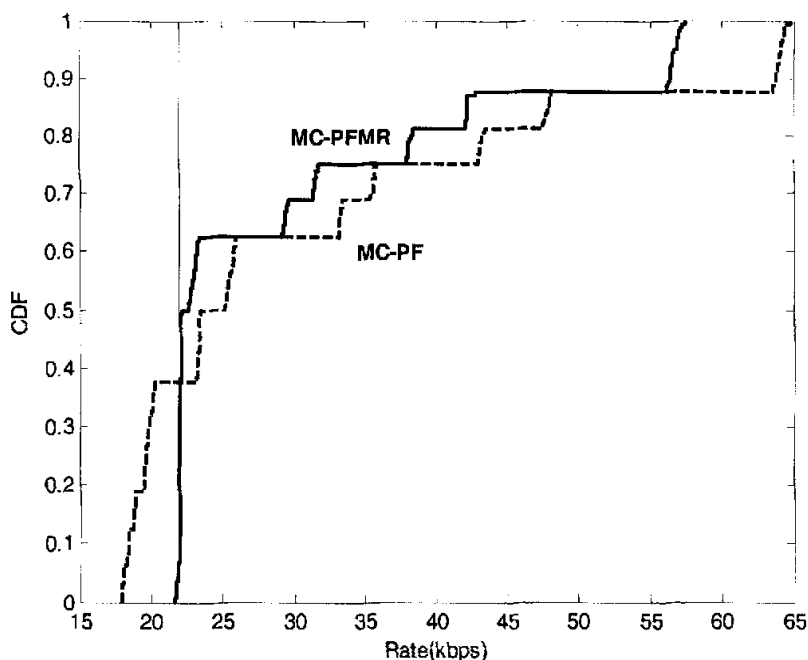


图 3.11 MC-PF 与 MC-PFMR( $R_i^{\min} = 22\text{kbps}$ )算法下用户速率 CDF 曲线

表 3.2 MC-PF 与 MC-PFMR( $R_i^{\min} = 22\text{kbps}$ )算法下系统效益与系统吞吐量

算法	$\sum_i \log R_i$	$\sum_i R_i$
MC-PF	481.55	4504.36
MC-PFMR	479.52	4288.10

### MC-PFMR 算法与 Penalty、Reservation 算法比较

比较 MC-PFMR 算法与已有的 Penalty 算法和 Reservation 算法<sup>[50]</sup>。Penalty 算法用倒数函数得到调度策略，依次将子载波  $j$  分配给  $C_{i,j}(t) / [R_i(t-1) - R_i^{\min}]$  值最

大的用户。Reservation 算法为各用户分别保留一定的子载波以满足最小速率限制要求，在剩下的子载波上做无速率限制的比例公平调度。

取最小速率限制为 25kbps，不同算法下用户速率 CDF 曲线如图 3.12 所示。最小速率限制用竖线画出，CDF 曲线与竖线交界点的纵坐标值表示未满足速率要求的用户数与用户总数的比值。可见，在 MC-PFMR 算法、Penalty 算法和 Reservation 算法下平均速率小于最小速率限制的用户分别占 12.5%、37.5%和 43.75%。这说明 MC-PFMR 算法比另两种算法能使更多用户达到最小速率要求。

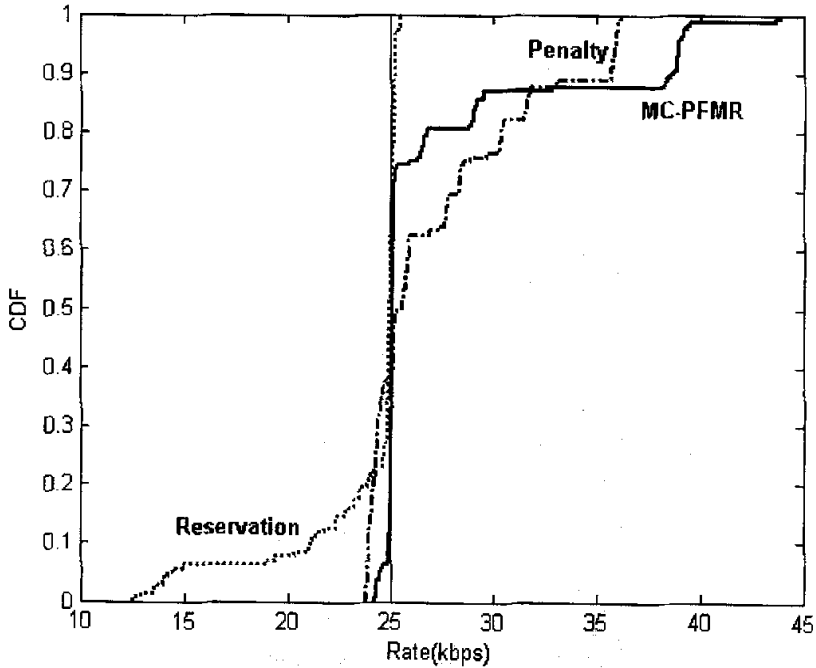


图 3.12  $R_i^{\min} = 25\text{kbps}$ ,  $R_i^{\max} = \infty$  时不同算法下用户速率 CDF 曲线

再取  $R_i^{\min} = 22\text{kbps}$ ，不同算法下的用户速率 CDF 曲线如图 3.13 所示。可见，当最小速率限制较小时，MC-PFMR 和 Penalty 算法下几乎所有用户速率都能达到要求，Reservation 算法下 87.5% 左右的用户速率也能达到要求。此时的系统效益和系统吞吐量在表 3.3 中给出。可以看到，MC-PFMR 算法在更好地满足了最小速率限制的同时，系统效益和系统吞吐量都比其他算法大。

表 3.3 不同算法下系统效益和系统吞吐量的值

算法	$\sum_i \log R_i$	$\sum_i R_i$	参数设置
MC-PFMR	479.52	4288.1	$R_i^{\min} = 22\text{kbps}$
Penalty	478.19	4098.3	
Reservation	448.56	3303.7	

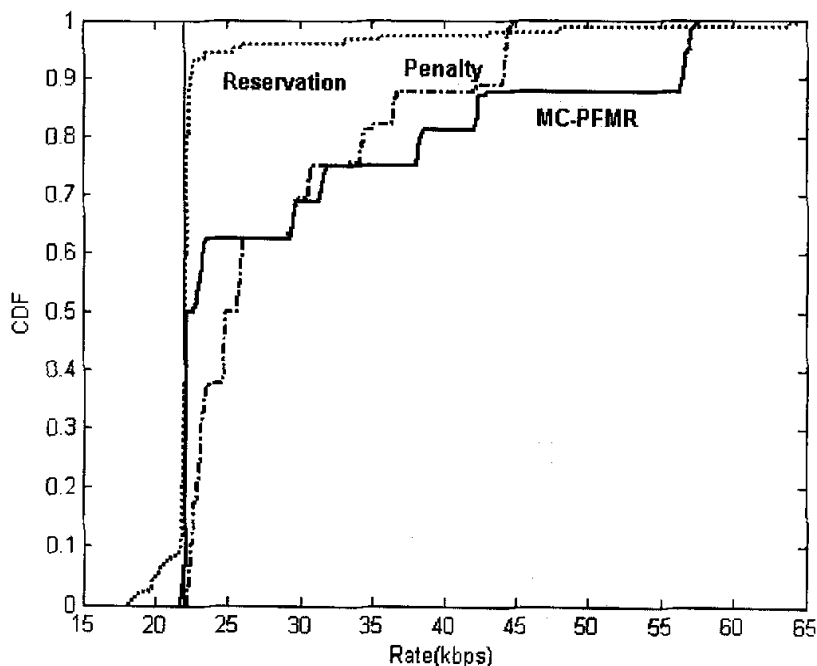


图 3.13  $R_i^{\min} = 22\text{kbps}$  ,  $R_i^{\max} = \infty$  时用户速率 CDF 曲线

### 存在不同用户类型时 MC-PFMR 算法的性能

实际应用中还有许多业务是同时有最小/最大速率限制要求的，而且很可能各个用户的速率限制值并不相同。设总共有三种类型的用户，用户数和最小/最大速率限制值如表 3.4 所示。

表 3.4 三种类型的用户参数设置

类型	$R_i^{\min}$	$R_i^{\max}$	用户数
I	32kbps	36kbps	2
II	22kbps	46kbps	8
II	0	$\infty$	6

MC-PFMR 算法下三种类型的用户速率 CDF 曲线见图 3.14。可以看到，各个类型的速率限制要求都得到了满足，说明 MC-PFMR 算法在用户速率限制不同时仍然表现很好。

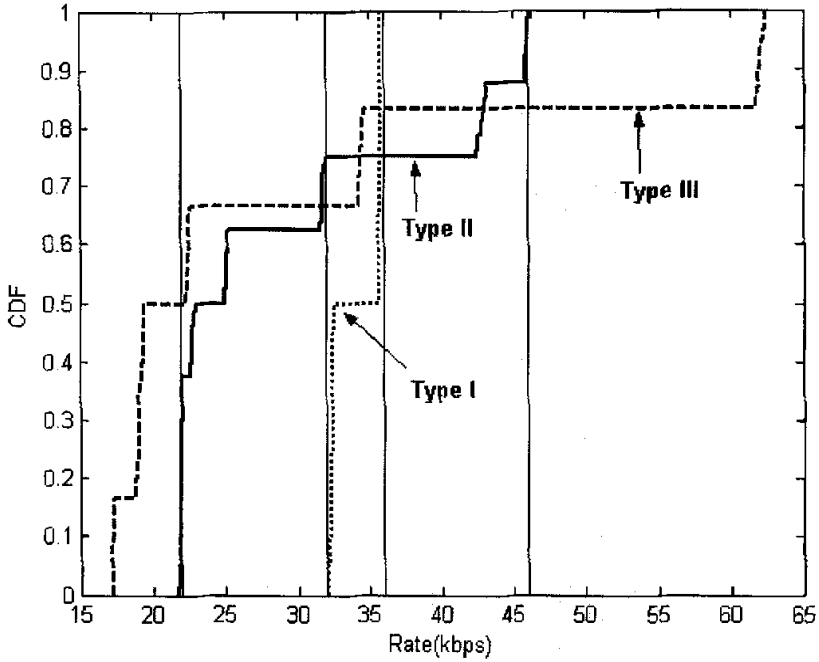


图 3.14 MC-PFMR 算法下速率限制不同的用户速率 CDF 曲线

### MC-PFMR 算法与 PFMR-S 算法比较

从 3.3.2 节的针对 BE 业务的比例公平调度算法的仿真结果可以看到，在每个子载波被分配后都更新所有用户已达速率的 MC-PF 算法与所有子载波都被分配后再更新用户已达速率的 PF-S 算法相比，性能十分接近。在有速率限制的时候，用户已达速率的更新方式对算法性能的影响是否存在相似的结论，我们比较 MC-PFMR 算法与 PFMR-S 算法。前者在每个子载波被分配后都更新所有用户的令牌数和已达速率，后者在所有子载波都被分配后更新所有用户的令牌数和已达速率。

取  $R_i^{\min} = 22\text{kbps}$ ， $R_i^{\max} = 46\text{kbps}$ ，MC-PFMR 算法和 PFMR-S 算法下用户速率的 CDF 曲线如图 3.15 所示，可以看到，这两种算法均能使几乎所有用户速率满足最小/最大速率要求，但是这两条曲线存在较大的差异。表 3.5 给出了这两种算法下系统效益和系统吞吐量的值。与 MC-PFMR 算法相比，PFMR-S 算法将会造成系统效益和系统吞吐量较小，这是由 PFMR-S 算法在所有子载波都被分配后才更新用户的令牌数和已达速率，导致调度器“反应过度”造成的，与 3.4 节开头的分析一致。

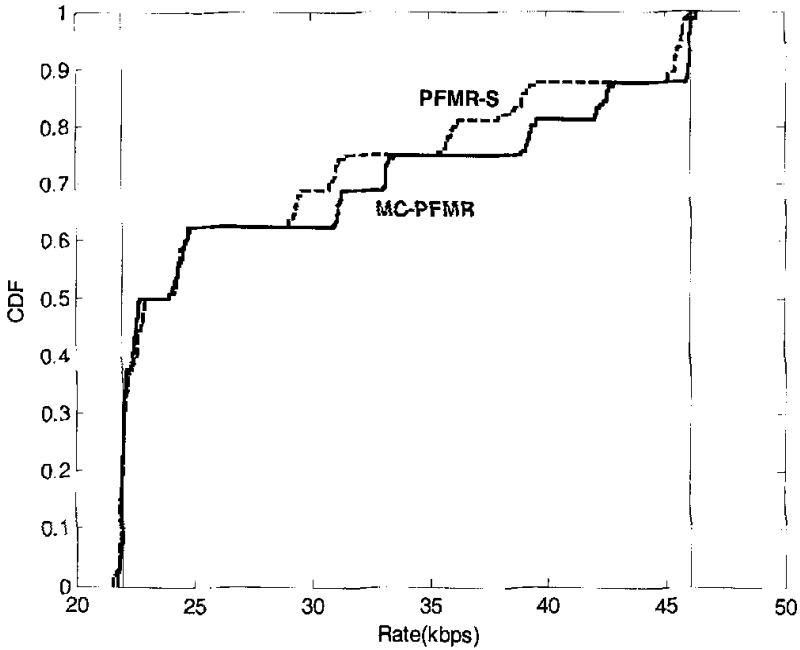


图 3.15  $R_i^{\min} = 22\text{kbps}$  ,  $R_i^{\max} = 46\text{kbps}$  时不同算法下用户速率 CDF 曲线

表 3.5 MC-PFMR 与 PFMR-S 算法下系统效益与系统吞吐量的值

算法	$\sum_i \log R_i$	$\sum_i R_i$	参数设置
MC-PFMR	478.59	4174.04	$R_i^{\min} = 22\text{kbps}$ $R_i^{\max} = 46\text{kbps}$
PFMR-S	475.74	4068.79	

### MC-PFMR 算法与几种变形算法

在 3.4.2 节中提到的几种变形算法，均采用了每个子载波被分配后都更新用户速率和令牌数的方式。仿真结果表明这几种算法得到的结果十分相近，用户速率 CDF 曲线很靠近，故不再给出，而各算法下的系统效益与系统吞吐量值在表 3.6 中给出。其中  $n=0$  时的 Exponent-MC-PFMR 算法就是将整个速率限制在各个子载波上做均分得到子载波上速率限制的方案。从表中可以看到，这几种算法的性能是很相近的。

表 3.6 MC-PFMR 与 PFMR-S 算法下系统效益与系统吞吐量的值

算法	$\sum_i \log R_i$	$\sum_i R_i$	参数设置
MC-PFMR	478.5904	4174.04	$R_i^{\min} = 22\text{kbps}$
Exponent-MC-PFMR(n=0)	478.5766	4173.28	
Exponent-MC-PFMR(n=5)	478.5927	4174.09	$R_i^{\max} = 46\text{kbps}$
Modified-MC-PFMR	478.5908	4174.07	

与上面的 PFMR-S 算法比较可以看到, 当有速率限制时, 每个子载波被分配后都作速率和令牌数的更新将较大幅度地提高系统效益与系统吞吐量, 这是与没有速率限制时的情况很不同的。所以, 针对 BE 业务, 可以选择计算复杂度低的 PF-S 算法来做比例公平调度, 而针对有速率限制的 NRT 业务, 应当选用 MC-PFMR 算法及其变形算法。

### 3.5 针对 NRT 业务的最大化吞吐量新算法(MC-MTMR)

本节讨论满足速率限制的同时最大化吞吐量的调度。

#### 3.5.1 算法描述

利用单载波最大化吞吐量 MTMR 算法(2.2.2.3 节)和多载波比例公平 MC-PFMR 算法(3.4.1 节), 我们提出适用于 OFDMA 系统的多载波有最小/最大速率限制时的最大化吞吐量调度 (Multi-Carrier Maximum Throughput with Minimum/Maximum Rate constraints) 算法。子载波上的最小/最大速率限制记为  $R_{i,j}^{\min}(t)$  和  $R_{i,j}^{\max}(t)$ , 按照式(3.22)计算。MC-MTMR 算法简单表述如下: 在时隙  $t$  的开始时刻, 按照式(3.23)和式(3.24)分别初始化  $R_i^0$  和  $T_i^0$ , 再依次将子载波  $j$  分配给用户  $i_j^*$ ,

$$i_j^* = \arg \max_i \exp\{a_i T_i^{j-1}\} C_{i,j}(t) \quad j=1,2,\dots,J \quad (3.31)$$

在每个子载波被分配后, 都更新所有用户的速率及令牌数。第  $j$  个子载波被分配后, 用户  $i$  的已达速率  $R_i^j$ , 令牌数  $T_i^j$  以及在子载波  $j$  上实际达到的速率  $\mu_{i,j}(t)$  按照式(3.26)-式(3.28)计算, 时隙  $t$  结束时刻的  $R_i(t)$  和  $T_i(t)$  分别为  $R_i^j$  和  $T_i^j$ 。

#### 3.5.2 算法的几种变形

MC-MTMR 算法也可以得到几种变形算法, 如 Modified-MC-MTMR 算法, MC-MTMR 算法也可以得到几种变形算法, 如 Modified-MC-MTMR 算法,

用全搜索全比较的方法分配子载波；还有 Exponent-MC-MTMR 算法等，在此不再赘述。

### 3.5.3 仿真与分析

仿真条件与 3.4.3 节相同。

取  $R_i^{\min} = 22\text{kbps}$ ， $R_i^{\max} = 46\text{kbps}$ ，以最大化吞吐量为目标的 MC-MTMR 算法与以比例公平为目标的 MC-PFMR 算法下用户速率 CDF 曲线如图 3.16 所示。可见 MC-MTMR 算法下几乎所有用户速率都满足了速率限制要求。表 3.7 给出了这两种算法下系统效益与系统吞吐量的值，MC-MTMR 算法的系统吞吐量比 MC-PFMR 算法大，但系统效益比 MC-PFMR 算法小，这种结果与两个算法各自的目标相符合。

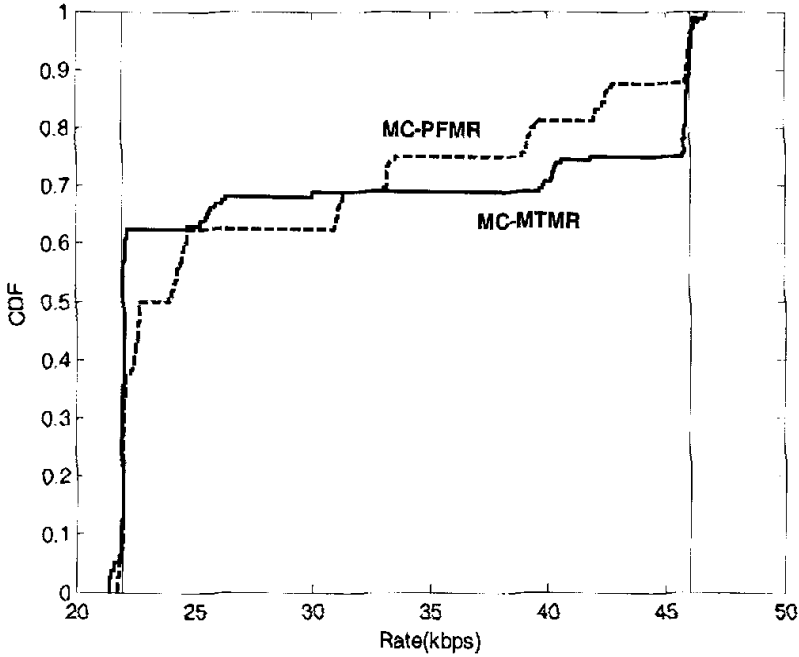


图 3.16  $R_i^{\min} = 22\text{kbps}$ ， $R_i^{\max} = 46\text{kbps}$  时不同算法下用户速率 CDF 曲线

表 3.7 MC-PFMR 与 PFMR-S 算法下系统效益与系统吞吐量的值

算法	$\sum_i \log R_i$	$\sum_i R_i$	参数设置
MC-MTMR	473.43	4289.18	$R_i^{\min} = 22\text{kbps}$ $R_i^{\max} = 46\text{kbps}$
MC-PFMR	478.59	4174.04	

## 本章小结

本章讨论多载波系统中的分组调度技术,借鉴第二章单载波系统中的经典算法,提出适用于多载波系统的调度算法。

本章讨论均以 OFDMA 系统为研究对象,这是因为 OFDMA 系统基于 OFDM 技术,具有频谱利用率高、能够对抗时间选择性衰落和频率选择性衰落的优点,近年来得到了广泛的关注与研究。在讨论适用于 OFDMA 系统的分组调度技术之前,首先讨论了 OFDMA 系统中的功率分配问题,指出在分组调度技术研究中常常采用总功率在各个子载波上均分这种方法的理论依据。

针对 BE 业务,调度算法考虑系统吞吐量和用户间的公平两方面的要求。现有的 TMA 算法实现了最大化系统吞吐量。将单载波系统中的比例公平调度策略移植到 OFDMA 系统中来,3.2.2 节讨论了已有的三种在 OFDMA 系统中实现比例公平的方案。

本章的后半部分提出了几种新的算法。一种是针对 BE 业务的改进多载波比例公平(Modified-MC-PF)调度算法,采用全搜索的方法寻找用户在各子载波上最大可达速率与该用户已达速率比值的最大值,将最大值对应的子载波分配给对应的用户,能够提高系统效益与系统吞吐量。一种是针对 NRT 业务的最小/最大速率限制时的比例公平(MC-PFMR)调度算法,借鉴单载波系统中 PFMR 算法的思路,并利用 OFDMA 系统的各用户在各子载波上信道响应各不相同的特点,引入子载波上最小/最大速率限制的概念,用令牌数反映用户速率限制的要求,用子信道状况和已达速率的比值反映比例公平的要求。最后一种是针对 NRT 业务的最小/最大速率限制时的最大化吞吐量(MC-MTMR)调度算法。



## 第四章 总结与展望

现代移动通信网络中不仅有传统的低速语音业务和尽力而为的数据业务，而且出现了多种不同 QoS 要求的业务类型，如视频、FTP 等等。通信底层的传输技术也得到了很大的发展，多载波系统尤其是具有优良性能的 OFDMA 系统得到了越来越多的关注。这对无线分组调度技术提出了新的挑战：一方面，必须尽力满足用户的 QoS 需求，另一方面，必须充分利用无线链路的特点，保证用户公平性的同时提高带宽利用率。

本文主要研究单载波和多载波系统中，针对尽力而为业务和有速率要求的非实时业务的无线分组调度技术。在对前人工作做了充分理解的基础上，提出了几种新算法或者改进算法，小结如下：

### ◆ 单载波系统中，针对 BE 业务的 ATMA 算法

算法目标： 保证用户间满足 3GPP2 规定的公平性的同时，提高系统吞吐量。克服比例公平调度造成较大的吞吐量损失的缺陷。

算法特点： 采用一种新的渐近线性效益函数，作为算法最优化目标。

算法性能： 与比例公平算法相比，ATMA 算法在满足了 3GPP2 公平性准则的同时提高了系统吞吐量。

### ◆ 单载波系统中，针对 NRT 业务的 S-UO 算法

算法目标： 提供最小速率保证的同时，提高系统效益与系统吞吐量。克服 Penalty 算法采用的效益函数在最小速率限制点处无定义，且给速率稍大于最小速率限制的用户过多的优先权的缺陷。

算法特点： 利用 S 函数曲线类似于阶跃函数曲线的特性，将效益函数写成 S 函数与对数函数相乘的形式。

算法性能： 与 Penalty 算法相比，能提供更高的系统效益与系统吞吐量。

### ◆ 多载波系统中，针对 BE 业务的 Modified-MC-PF 算法

算法目标： 提高系统效益与系统吞吐量。

算法特点： 采用全搜索的方法寻找用户在各子载波上最大可达速率与该用户已达速率比值的最大值，将最大值对应的子载波分配给对应的用户。

算法性能： 与已有的三种比例公平调度方案相比，能够提供更高的系统效益与系统吞吐量，并且这种提高随着用户数的增大而增大。但算法复杂度也有所增加。

◆ 多载波系统中, 针对 NRT 业务的 MC-PFMR 算法

算法目标: 满足最小/最大速率限制的条件下尽可能达到用户速率间的比例公平。

算法特点: 利用 OFDMA 系统的各用户在各子载波上信道响应各不相同的特点, 引入子载波上最小/最大速率限制的概念。用令牌数反映用户速率限制的要求, 用子载波信道状况和已达速率的比值反映比例公平的要求, 并在每个子载波被分配后都更新各用户已达速率和令牌数。

算法性能: 算法在容纳用户数、系统效益和系统吞吐量等方面性能优于现有其他常用算法, 在用户速率限制不同时仍然表现很好, 并且能与传统的无速率限制的比例公平调度完全兼容。

◆ 多载波系统中, 针对 NRT 业务的 MC-MTMR 算法

算法目标: 满足最小/最大速率限制的条件下最大化系统吞吐量。

算法特点: 用令牌反映速率限制的要求, 用子载波信道状况反映最大化吞吐量的要求。

算法性能: 有很强的速率限制能力, 其变形算法能与传统的无速率限制的 TMA 算法兼容。

当然, 本文的工作还存在很多不足之处, 比如在研究多载波系统中的分组调度技术时, 做了功率在各个子载波上平均分配的假设。在以往的研究当中, 功率分配问题往往被划入无线资源管理的范畴, 需要解决的是物理层资源优化分配的问题, 而分组调度技术面对的是上层, 需要解决的是上层各个用户的 QoS 保证的问题。这种分层分模块的思想在单载波传输, 而且网络中的业务类型单一、数据传输率较低的时代得到了广泛的应用。然而随着无线传输技术的发展, 能够提供的数据率越来越高, 同时用户也希望网络能够提供多种 QoS 等级的服务, 网络中的业务类型逐渐变得多种多样。此时, 这种将无线资源管理与无线分组调度分开的方法显出了一定局限性: 它很难提供整体上的最优化方案。面对这样的问题, 研究人员提出了跨层设计(Cross-Layer Design)的思想<sup>[51-53]</sup>, 目的是将无线资源管理与无线分组调度结合起来, 以达到整体最优化。由此, 如何将功率分配问题与分组调度技术结合起来就成为下一步的研究重点。

## 参考文献

- [1] J.Proakis, *Digital Communications*, New York, McGraw Hill, Inc., 1995.
- [2] J.Lee, L.Miller 等著, 徐希斌, 周世东等译, CDMA 系统工程手册, 北京, 人民邮电出版社, 2002
- [3] J.Chuang, N.Sollenberger, "Beyond 3G: Wideband Wireless Data Access Based on OFDM and Dynamic Packet Assignment", *IEEE Communications Magazine*, 2000, vol.38, issue 7, pp.78-87
- [4] Air Interface for Fixed Broadband Wireless Access Systems, IEEE STD.802.16, 2004
- [5] Jean Walrand, Pravin Varaiya 著, 张艳, 孙瑞志, 姜进磊等译, 高性能通信网络, 北京, 机械工业出版社, 2002
- [6] A.Raha, N. Malcolm, Wei Zhao. "Hard real-time communications with weighted round robin service in ATM local area networks", *IEEE International Conference on Engineering of Complex Computer Systems*, Nov 1995, vol.1, pp.96-103
- [7] A.K.Parekh, R.G. Gallager, "A Generalized Processor Sharing Approach to Flow Control in Integrated Services Networks: The Single-Node Case", *IEEE/ACM Transactions on Networking*, June 1993, vol.1, issue 3, pp.344-357
- [8] A.K.Parekh, R.G. Gallager, "A Generalized Processor Sharing Approach to Flow Control in Integrated Services Networks: The Multi-Node Case", *IEEE/ACM Transactions on Networking*, June 1993, vol.2, issue 2, pp.137-150
- [9] V.Bharghavan, Songwu Lu, T.Nandagopal, "Fair Queuing in Wireless Networks: Issues and Approaches", *IEEE Personal Communications*, Feb 1999, vol.6, issue 1, pp.44-53
- [10] 3GPP2, C.S0024, cdma2000 High Rate Packet Data Air Interface Specification, Version 4.0, October 25, 2002
- [11] P.Bender, et al, "CDMA/HDR: A Bandwidth-Efficient High-Speed Wireless Data Service for Normadic Users", *IEEE Communications Magazine*, 2000, vol.38, issue 7, pp.70-77
- [12] Harri Holma, Antti Toskala 著, 陈译强 周华 付景兴等译, WCDMA 技术与系统设计: 第三代移动通信系统的无线接入, 北京, 机械工业出版社, 2005

- [13] 仇佩亮, 信息论与编码, 北京, 高等教育出版社, 2003
- [14] 杨大成, cdma2000 1x 移动通信系统, 北京, 机械工业出版社, 2003
- [15] A.Jalali, R.Padovani, R.Pankaj, "Data Throughput of CDMA-HDR: A High Efficiency High Data Rate Personal Communication Wireless System", *IEEE Vehicular Technology Conference May 2000 (VTC2000-Spring)*, vol.3, pp.1854-1858
- [16] Frank Kelly, A. Maulloo, and D. Tan, "Rate control for communication networks: shadow prices, proportional fairness and stability", *Journal of the Operational Research Society*, Mar. 1998, vol.49, pp. 237-252
- [17] R.Mazumdar, L.G. Mason, C.Douligieris, "Fairness in Network Optimal Flow Control", *IEEE International Telecommunications Symposium*, Sep 1990, (ITS'90 Symposium Record), pp.590-596
- [18] S.Shenker, "Fundamental Design Issues for the Future Internet", *IEEE Journal on Selected Areas in Communications*, Sep 1995, vol.13, issue 7, pp.1176-1188
- [19] Frank Kelly, "Charging and Rate Control for Elastic Traffic", *European Transactions on Telecommunications*, 1997, vol. 8, pp. 33-37
- [20] Harold J. Kushner, Philip A. Whiting, "Asymptotic Properties of Proportional-Fair Sharing Algorithms", Division of Applied Mathematics, Brown University, Providence; Bell Labs, Lucent Technologies, Murray Hill, New Jersey; May 2002
- [21] Matthew Andrews. "Instability of the Proportional Fair Scheduling Algorithms for HDR" Bell Laboratories, Lucent Technologies, Murray Hill, New Jersey, August 28, 2002
- [22] Hoon Kim, Keunyoung Kim, Youngnam Han and Jiwoong Lee. "An Efficient Scheduling Algorithm for QoS in Wireless Packet Data Transmission" *IEEE Personal, Indoor, and Mobile Radio Communications*, Sep 2002 (PIMRC'02), vol.5, pp.2244-2248
- [23] Ghassane Aniba and Sonia Aissa, "Adaptive Proportional Fairness for Packet Scheduling in HSDPA" *IEEE Global Telecommunications Conference*, Nov 2004, (GLOBECOM'04), vol.6, pp.4033-4037
- [24] C. Wengerter, J. Ohlhorst, Golitschek Edler von Elbwart, "Fairness and Throughput Analysis for Generalized Proportional Fair Frequency Scheduling in OFDMA", *IEEE Vehicular Technology Conference*, June 2005, (VTC 2005-Spring), vol.3, pp.1903-1907
- [25] Guocong Song, Ye Li, "Adaptive Subcarrier and Power Allocation in OFDM based on

- Maximizing Utility”, *IEEE Vehicular Technology Conference*, 2003 (VTC 2003-Spring), vol.2, pp.905-909
- [26] Guocong Song, Ye Li, “Cross-Layer Optimization for OFDM Wireless Networks—part I: Theoretical Framework”, *IEEE Transactions on Wireless Communications*, 2005, vol.4, pp.614-624
- [27] Guocong Song, Ye Li, “Cross-Layer Optimization for OFDM Wireless Networks—part II: Algorithm Development”, *IEEE Transactions on Wireless Communications*, 2005, vol.4, pp.625-633
- [28] Jeroen Wigard, Nina A.H. Madsen, Pablo Ameigeiras Gutierrez, Isaias Lopez Sepulveda and Preben Mogensen, “Packet Scheduling with QoS Differentiation”, *ACM Wireless Personal Communications*, 2002, vol. 23, issue 1, pp. 147-160
- [29] Matthew Andrews, Lijun Qian, Alexander Stolyar, “Optimal Utility Based Multi-User Throughput Allocation subject to Throughput Constraints”, *Annual Joint Conference of the IEEE Computer and Communications Societies, Proceeding of IEEE INFOCOM*, March 2005, vol. 4, pp.2415 – 2424
- [30] Alexander L. Stolyar, “On the Asymptotic Optimality of the Gradient Scheduling Algorithm for Multiuser Throughput Allocation”, Bell Labs, Lucent Technologies, Murray Hill, New Jersey, Jan-Feb 2005, pp.12-25
- [31] Z.Jiang, Y.Ge, Y.Li, “Max-Utility Wireless Resource Management for Best-Effort Traffic”, *IEEE Transactions on Wireless Communications*, Jan 2005, vol.4, issue 1, pp. 100-111
- [32] 杨大成等, 移动传播环境: 理论基础, 分析方法和建模技术, 北京, 机械工业出版社, 2003
- [33] 3GPP2, C.R1002, cdma2000 Evaluation Methodology, Version 1.0, December 10, 2004
- [34] Seungwan Ryu, Byunghan Ryu, Hyunhwa Seo, and Mooyong Shin, “Urgency and Efficiency based Packet Scheduling Algorithm for OFDMA Wireless System”, *IEEE International Conference on Communications*, May 2005, (ICC’05), vol.4, pp.2779-2785
- [35] M. Fazel, Mung Chiang, “Network Utility Maximization With Nonconcave Utilities Using Sum-of-Squares Method”, *IEEE Conference on Decision and Control*, Dec 2005, pp.1867-1874
- [36] V.K.N. Lau, Yu-Kwong Kwok, “Performance Analysis of SIMO Space-Time Scheduling with

- Convex Utility Function: Zero-Forcing Linear Processing”, *IEEE Transactions on Vehicular Technology*, March 2004, vol.53, issue 2, pp.339-350
- [37] J.W. Lee, R.R. Mazumdar, N.B.Shroff, “Non-Convex Optimization and Rate Control for Multi-Class Services in the Internet”, *IEEE/ACM Transactions on Networking*, Aug 2005, vol.13, issue 4, pp.827-840
- [38] S. B. Weinstein and P. M. Ebert, “Data Transmission by Frequency Division Multiplexing Using the Discrete Fourier transform”, *IEEE Transactions on Communication Technology*, Oct. 1971, vol. COM-19, pp.628-634
- [39] 佟学俭, 罗涛, OFDM 移动通信技术原理与应用, 北京, 人民邮电出版社, 2003
- [40] C.Y. Wong, R.S. Cheng, et al, “Multiuser OFDM with Adaptive Subcarrier, Bit and Power Allocation”, *IEEE Journal on Selected Areas in Communications*, 1999, vol.17, pp.1747-1757,
- [41] Jiho Jang, Kwang Lee, “Transmit Power Adaptation for Multiuser OFDM Systems”, *IEEE Journal on Selected Areas in Communications*, 2003, pp.171-178
- [42] M.Ergen, S.Coleri, “QoS Aware Adaptive Resource Allocation Techniques for Fair Scheduling in OFDMA Based Broadband Wireless Access Systems”, *IEEE Transactions on Broadcasting*, 2003, vol.49, pp.362-370
- [43] J.Zander, S.L.Kim, *Radio Resource Management for Wireless Networks*, Artech House, USA, 2001
- [44] R.Knopp, P.A.Humbler, “Information Capacity and Power Control in Single-Cell Multiuser Communications”, *IEEE International Conference on Communications*, 1995, (ICC'95), vol.1, pp.331-335
- [45] V.Bharghavan, S.Lu, “A Sub-Optimal Joint Subcarrier and Power Allocation Algorithm for Multiuser OFDM”, *IEEE Communication Letters*, 1999, vol.9, issue 8, pp.44-53
- [46] Wang Anchun, Xiao Liang, Zhou Shidong, Xu Xibin, Yao Yan, “Dynamic Resource Management in the Fourth Generation Wireless Systems” *Proceeding of International Conference on Communication Technology*, April 2003, (ICCT'03) vol. 2, pp.1095 -1098
- [47] 阳洁, 基于 IEEE802.16 的 QoS 技术研究, 杭州, 浙江大学硕士论文, 2005.2
- [48] Hoon Kim, Keunyoung Kim Youngnan Han and Sangboh Yun, “A Proportional Fair Scheduling for Multicarrier Transmission Systems”, *IEEE Vehicular Technology Conference*,

Sep 2004, (VTC 2004-Fall), vol.1, pp.409-413

- [49] Wang Anchun, X.She, et al, "Asymptotic Analysis of Fair Scheduling in the OFDM Systems", *IEEE Personal, Indoor, and Mobile Radio Communications*, Sep 2003 (PIMRC'03), vol.2, pp.1186-1191
- [50] W. Wang, K. C. Hwang, K. B. Lee, and S. Bahk, "Resource Allocation for Heterogeneous Services in Multiuser OFDM Systems", *Proceeding of IEEE Global Telecommunications Conference*, Nov 2004, (GLOBECOM'04), vol. 6, pp. 3478-3481
- [51] S.Shakkottai, T.S.Rappaport, P.C.Karlsson, "Cross-layer Design for Wireless Networks", *IEEE Communications Magazine*, Oct 2003, vol.41, issue 10, pp.74-80
- [52] V.Srivastava, M.Motani, "Cross-Layer Design: A Survey and the Road Ahead", *IEEE Communications Magazine*, Dec 2005, vol.43, issue 12, pp.112-119
- [53] Taesoo Kwon, Howon Lee, Sik Choi, Juyeop Kim, Dong-Ho Cho, Sunghyun Cho, Sangboh Yun, Won-Hyoung Park, Kiho Kim, "Design and Implementation of a Simulator based on a Cross-Layer Protocol Between MAC and PHY Layers in a WiBro Compatible.IEEE 802.16e OFDMA System", *IEEE Communications Magazine*, Dec 2005, vol.43, issue 12, pp.136-146

## 附录 A 比例公平一个特例的证明

比例公平调度定义为：假设在调度器  $P$  下，各个用户速率为  $R_i^P$ ，在调度器  $S$  下，各个用户速率为  $R_i^S$ ，当且仅当条件(A.1)满足时，调度器  $P$  被称为比例公平调度器，

$$\sum_i \frac{R_i^S - R_i^P}{R_i^P} \leq 0 \quad (\text{A.1})$$

用一个极端的例子来定性说明比例公平的原则。假设有两个用户时分复用共享一个信道，时隙长度非常小，且这两个用户在这个信道上的信道容量为不相等的固定值  $C_1$  (单位为 bps) 和  $C_2$  (单位为 bps)。不失一般性，我们假设  $C_1 > C_2$ 。考察在很长一段时间后用户的速率  $R_1$  与  $R_2$ ，下面我们将证明，在这个例子中，式(A.1)的充要条件是  $R_1 : R_2 = C_1 : C_2$ 。

称最终使得用户速率满足  $R_1 : R_2 = C_1 : C_2$  的方案为方案一，最终使得用户速率满足  $R_1' : R_2' \neq C_1 : C_2$  的方案为方案二，证明如下。

充分性：即要证明式(A.1)必然导致最终的方案为方案一。用反证法，假设式(A.1)导致的方案是方案二。由  $R_1 : R_2 = C_1 : C_2$  可以推得，方案一中用户占用总传输时间的比例为  $k_1 = k_2 = 0.5$ ，设方案二中用户占用总传输时间的比例分别为  $k_1'$  和  $k_2'$ ，满足  $k_1' = 0.5 + \Delta k$ ， $k_2' = 0.5 - \Delta k$ ，其中  $-0.5 < \Delta k < 0.5$  且  $\Delta k \neq 0$ ，得到  $R_1' = k_1' C_1$ ， $R_2' = k_2' C_2$ 。方案二满足(A.1)式，即：

$$\frac{R_1 - R_1'}{R_1'} + \frac{R_2 - R_2'}{R_2'} \leq 0 \quad (\text{A.2})$$

将  $R_1$ ， $R_2$ ， $R_1'$ ， $R_2'$  的值代入上式，得到：

$$\begin{aligned} \frac{R_1 - R_1'}{R_1'} + \frac{R_2 - R_2'}{R_2'} &= \frac{k_1 C_1 - k_1' C_1}{k_1' C_1} + \frac{k_2 C_2 - k_2' C_2}{k_2' C_2} \\ &= \frac{k_1 - k_1'}{k_1'} + \frac{k_2 - k_2'}{k_2'} = \frac{-\Delta k}{0.5 + \Delta k} + \frac{\Delta k}{0.5 - \Delta k} \\ &= \frac{2(\Delta k)^2}{0.25 - (\Delta k)^2} \leq 0 \end{aligned} \quad (\text{A.3})$$

这与  $-0.5 < \Delta k < 0.5$  且  $\Delta k \neq 0$  的条件不符，所以假设不成立，由此证明了，式(A.1)必然导致最终的方案使得用户速率满足  $R_1 : R_2 = C_1 : C_2$ 。

必要性：即要证明最终用户速率满足  $R_1 : R_2 = C_1 : C_2$  的方案一必然是比例公平的。将  $R_1$ ， $R_2$ ， $R_1'$ ， $R_2'$  的值代入式(A.1)，得到：



$$\begin{aligned} \frac{R_1' - R_1}{R_1} + \frac{R_2' - R_2}{R_2} &= \frac{k_1' C_1 - k_1 C_1}{k_1 C_1} + \frac{k_2' C_2 - k_2 C_2}{k_2 C_2} \\ &= \frac{k_1' - k_1}{k_1} + \frac{k_2' - k_2}{k_2} = \frac{\Delta k}{0.5} + \frac{-\Delta k}{0.5} = 0 \end{aligned} \quad (\text{A.4})$$

即方案一满足式(A.1)，方案一是比例公平的方案。由此我们证明了，在这个例子中，比例公平的充要条件是最终用户速率满足  $R_1 : R_2 = C_1 : C_2$ 。此时，用户最终吞吐量与信道条件好坏成比例，并且系统吞吐量为  $(C_1 + C_2)/2$ 。

## 附录 B 比例公平等价条件的证明

比例公平调度定义为：假设在调度器  $P$  下，各个用户速率为  $R_i^P$ ，在调度器  $S$  下，各个用户速率为  $R_i^S$ ，当且仅当条件(B.1)满足时，调度器  $P$  被称为比例公平调度器，

$$\sum_i \frac{R_i^S - R_i^P}{R_i^P} \leq 0 \quad (\text{B.1})$$

文献[16]中证明了比例公平调度器的目标是使用户速率的对数和最大，即比例公平调度器的目标可以写成：

$$\max \sum_i \log R_i \quad (\text{B.2})$$

简单证明如下：设  $\bar{R} = (R_1, R_2, \dots, R_I)$  是可达速率空间  $R'$  中的点， $f(\bar{R})$  是定义在  $R'$  上的  $I$  元标量函数，定义如下，

$$f(\bar{R}) = f(R_1, R_2, \dots, R_I) = \sum_i \log R_i \quad (\text{B.3})$$

若  $R'$  是连通的区域，则  $f(\bar{R})$  在  $R'$  上连续且可微，且  $f(\bar{R})$  的梯度为：

$$\begin{aligned} \nabla f(\bar{R}) &= \left( \frac{\partial}{\partial R_1} f(\bar{R}), \frac{\partial}{\partial R_2} f(\bar{R}), \dots, \frac{\partial}{\partial R_I} f(\bar{R}) \right) \\ &= \left( \frac{\partial}{\partial R_1} \sum_i \log R_i, \frac{\partial}{\partial R_2} \sum_i \log R_i, \dots, \frac{\partial}{\partial R_I} \sum_i \log R_i \right) \\ &= \left( \frac{1}{R_1}, \frac{1}{R_2}, \dots, \frac{1}{R_I} \right) \end{aligned} \quad (\text{B.4})$$

若在比例公平调度器调度器  $P$  下，各个用户速率为  $R_i^P$ ，向量形式为  $\bar{R}^P = (R_1^P, R_2^P, \dots, R_I^P)$ ，而在调度器  $S$  下，各个用户速率为  $R_i^S$ ，向量形式为  $\bar{R}^S = (R_1^S, R_2^S, \dots, R_I^S)$ 。设  $\bar{R}^S$  点在  $\bar{R}^P$  点附近，速率的变化量  $\Delta \bar{R}$  可以表示为：

$$\Delta \bar{R} = \bar{R}^S - \bar{R}^P = (R_1^S - R_1^P, R_2^S - R_2^P, \dots, R_I^S - R_I^P) \quad (\text{B.5})$$

而函数  $f(\bar{R})$  的值由  $\bar{R}^P$  到  $\bar{R}^S$  点的变化量  $\Delta f(\bar{R}) \Big|_{\bar{R}^P \rightarrow \bar{R}^S}$  可以表示为：

$$\begin{aligned}
 \Delta f(\bar{R})\Big|_{\bar{R}^p \rightarrow \bar{R}^s} &= f(\bar{R}^s) - f(\bar{R}^p) = \left( \nabla f(\bar{R})\Big|_{\bar{R}^p} \right) \cdot (\Delta \bar{R}) \\
 &= \left( \frac{1}{R_1^p}, \frac{1}{R_2^p}, \dots, \frac{1}{R_I^p} \right) \cdot (R_1^s - R_1^p, R_2^s - R_2^p, \dots, R_I^s - R_I^p) \\
 &= \sum_i \frac{R_i^s - R_i^p}{R_i^p}
 \end{aligned} \tag{B.6}$$

因为调度器  $P$  是比例公平调度器，即  $\bar{R}^p$  与  $\bar{R}^s$  的关系满足(B.1)式，得到：

$$\sum_i \frac{R_i^s - R_i^p}{R_i^p} \leq 0 \Leftrightarrow \Delta f(\bar{R})\Big|_{\bar{R}^p \rightarrow \bar{R}^s} \leq 0 \tag{B.7}$$

这表示对于任意  $\bar{R}^s$ ，函数增量  $\Delta f(\bar{R})\Big|_{\bar{R}^p \rightarrow \bar{R}^s}$  不大于零，也就是说，函数  $f(\bar{R})$  在  $\bar{R}^p$  取局部最大值。

由于  $f(\bar{R})$  是严格凸函数之和，所以它也是严格凸函数；而根据多用户信息论，所有可达速率组成的集合  $R^I$  是凸集，而严格凸函数在凸集上最多只有一个最大值，并且局部最大值就是全局最大值。由此我们证明了比例公平调度器的目标就是(B.2)式，并且在  $R^I$  为连通区域时有唯一解。文献[16]给出了更为严格的证明。

## 附录 C OFDMA 系统功率分配问题

对于  $J$  个独立并行高斯信道，噪声功率分别为  $N_1, N_2, \dots, N_J$ ，根据香农信道容量公式， $J$  个信道的总容量为：

$$C = \max_{p_i} \frac{1}{2} \sum_{j=1}^J \log_2 \left( 1 + \frac{P_j}{N_j} \right) \quad (\text{C.1})$$

其中  $p_j$  是信道  $j$  的信号功率，满足功率限制  $\sum_{j=1}^J p_j = P$ 。

利用拉格朗日乘法，设目标函数为：

$$F(p_1, p_2, \dots, p_J) = \frac{1}{2} \sum_{j=1}^J \log_2 \left( 1 + \frac{P_j}{N_j} \right) - \lambda \sum_{j=1}^J p_j \quad (\text{C.2})$$

上式对  $p_j$  求导，且令导数等于零，得到，

$$\frac{1}{2} \frac{1}{p_j + N_j} - \lambda = 0 \quad (\text{C.3})$$

令  $\gamma = 1/2\lambda$ ， $\gamma$  为待定常数，由功率限制确定。则由式(C.3)可得  $p_j = \gamma - N_j$ ， $j=1, 2, \dots, J$ 。由于  $p_j$  必须大于或者等于零，所以该式不一定代表可实现解。下面对解作一修正，令

$$p_j = (\gamma - N_j)^+ \quad (\text{C.4})$$

其中，正函数  $(x)^+$  定义为：

$$(x)^+ = \begin{cases} x & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases} \quad (\text{C.5})$$

利用下式确定  $\gamma$ ，即可求出最优解  $p_j$ ，

$$\sum_{j=1}^J (\gamma - N_j)^+ = P \quad (\text{C.6})$$

利用对分搜索法等数值计算方法可以很容易的求得  $\gamma$ 。因此并行高斯信道的总容量为，

$$C = \frac{1}{2} \sum_{j=1}^J \log_2 \left( 1 + \frac{(\gamma - N_j)^+}{N_j} \right) \quad (\text{C.7})$$

从式(C.4)可以看出信道的噪声功率越小，分配的功率越多，噪声功率越大则分配

的功率越少。在总功率分配给各个子信道时, 首先把功率分配给噪声最小的信道, 然后分配给第二小噪声信道, 就像灌水一样进行, 直到水涨到一定水平  $\gamma$ , 使总的灌水量等于总功率  $P$ , 如图 C.1 所示。噪声功率超过一定值( $\gamma$ )的信道干脆不用, 这是灌水(Water Filling)法则的基本思想。

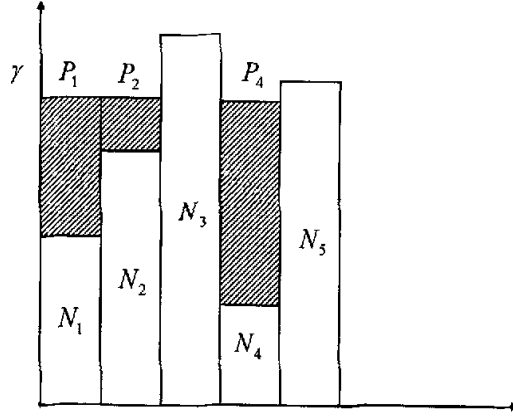


图 C.1 灌水原理示意图

现在我们讨论功率足够大时的情况。此时, 式(C.6)退化为  $\sum_{j=1}^J (\gamma - N_j) = P$ ,

则  $\gamma = \frac{1}{J} (P + \sum_{j=1}^J N_j) \approx \frac{P}{J}$ , 信道容量简化为:

$$C \approx \frac{1}{2} \sum_{j=1}^J \log_2 \left( \frac{P}{J N_j} \right) \quad (C.8)$$

由此可见在高信噪比下, 平均功率分配方法所取得的信道容量与最优的灌水法则取得的信道容量基本相同。为了简化资源优化分配算法, 高信噪比下可以使用简单的平均功率分配策略, 所取得的信道容量接近于最优分配方法。

OFDM 系统中可以得到类似的结论。当 OFDM 的子载波近似相互独立时, 可以用灌水法则将总功率优化分配到各个子载波, 取得最大的信道容量, 具体推导如下。

OFDM 系统一般具有较宽的频谱, 其信道是频率选择性的。某一时刻的信道冲激响应可以写成:

$$\mathbf{a} = [a_0, a_1 \dots a_{L-1}] \quad (C.9)$$

其中  $a_l$  是第  $l$  径上的信道衰落系数。对于 Rayleigh 衰落,  $a_l$  是一个复高斯随机变量。 $L$  是多径数目。假设子载波的带宽足够小, 子载波上信道可认为是平坦的。则利用傅立叶变换, 可以得到子载波上的频域信道响应,

$$H_k = \sum_{l=0}^{L-1} a_l e^{-j2\pi k l / K}, k = 1, 2, \dots, J \quad (C.10)$$

其中  $J$  为子载波的数目。假设各个子载波上的接收噪声功率相等，记为  $\sigma^2$ ，则子载波  $k$  上的等效噪声功率可以写为，

$$N_k = \frac{\sigma^2}{|H_k|^2} \quad (C.11)$$

利用式(C.4)的结论可以得到 OFDM 系统中的灌水功率分配结果，

$$p_k = \left( \gamma - \frac{\sigma^2}{|H(k)|^2} \right)^+ \quad (C.12)$$

根据香农信道容量公式，子载波  $k$  上可传输的最大比特数为

$$b_k = \lceil \log_2(\gamma |H(k)|^2) \rceil^+ \quad (C.13)$$

OFDM 系统中的功率分配问题可以分为两类，一类是总功率约束下的最大化传输速率问题，称为速率优化问题(Rate Adaptive, RA)，数学上可以表示成：

$$\begin{aligned} \max R_T &= \sum_{k=1}^J b_k \\ \text{Subject to: } &\sum_{k=1}^J p_k = P_T \end{aligned} \quad (C.14)$$

其中  $P_T$  为系统总功率约束。另一类是传输速率约束下的最小化总功率问题(Margin Adaptive, MA)，称之为边际优化问题，数学上可以表示为：

$$\begin{aligned} \min P_T &= \sum_{k=1}^J p_k \\ \text{Subject to: } &\sum_{k=1}^J b_k = R_T \end{aligned} \quad (C.15)$$

其中  $R_T$  为系统传输速率约束。第一类问题一般应用在针对 BE 业务和 NRT 业务的功率分配或者分组调度中。第二类问题一般应用在针对 RT 业务的功率分配或者分组调度中。

对于这两类问题，最优的功率分配结果都由公式(C.12)给出，所不同的是灌水线  $\gamma$  的选择。速率优化问题中的  $\gamma$  满足：

$$\sum_{k=1}^J \left( \gamma - \frac{\sigma^2}{|H(k)|^2} \right)^+ = P_T \quad (C.16)$$

边际优化问题中的  $\gamma$  满足

$$\sum_{k=1}^J \left[ \log_2(\gamma |H(k)|^2) \right]^+ = R_T \quad (\text{C.17})$$

由于本文主要针对 BE 业务和 NRT 业务进行研究，所以这里主要讨论速率最优化问题。同样，当功率足够大时，由式(C.16)可以得到：

$$\gamma = \frac{1}{J} \left( P + \sum_{k=1}^J \frac{\sigma^2}{|H_k|^2} \right) \approx \frac{P}{J} \quad (\text{C.18})$$

也就是说，在 OFDMA 系统中，可以将功率在各个子载波上平均分配，然后只考虑子载波分配和比特分配，大大降低了复杂度。

## 在读期间参与的科研项目和完成的论文

参与的科研项目：

浙江省自然科学基金重点项目(No. Z104252), “移动宽带无线接入(MBWA)系统关键技术研究”。

完成的论文：

Liang Feng, Aiping Huang, “Proportional Fair Scheduling with Rate Constraints for OFDMA Systems”, ICWMMN 2006, 已录用。

冯靛, 黄爱苹, “OFDMA 系统中有速率限制的比例公平调度”, 2006 年 4 月已投浙大报工学版, 稿号 G06193。



## 致谢

两年的研究生生涯虽然短暂,但这两年中结识的师友所给予我的帮助与支持却会让我受益终生。

首先感谢我的导师黄爱苹教授。从本科毕业设计到研究生阶段的学习,黄老师均给我无微不至的指导,深入浅出地为我讲授了许多专业知识,对我的研究工作提出了很多很好的建议。黄老师一丝不苟的治学态度,在理论上深入细致,实践上力求创新的科研精神都给我留下了非常深刻的印象。严谨的治学态度和精益求精的工作态度将使我终生受益,再次向黄老师表示深深的谢意!

感谢张朝阳教授在我攻读学位期间对我的帮助,给了我很多学术上启发和解决实际问题上的帮助。

感谢余官定博士,在我刚刚进入科研领域时,从基本概念的理解到仿真调试上都给予我很大的帮助,并给我的论文提出了很多宝贵意见。

感谢李静博士对我的启发和帮助。感谢朱江力、张伟峰和刘漪同学,和他们在一起进行学术上的讨论交流是愉快且极具启发性的。

感谢室友王婷婷、徐艺和姚伟春,和她们在一起我感到非常开心。

感谢姚天翔博士一直以来对我的关心和帮助。

最后,我要将最衷心的感谢献给我的亲人,在我近二十年的求学生涯中,他们都给予了最大的关心、支持和鼓励。我的父亲是一位辛勤耕耘四十余年的优秀教师,他对我和小弟的成长寄托了无数希望,倾注了大量心血,使我能够进入浙江大学学习,小弟能够进入北京大学学习。养育之恩无以为报,谨以此文献给我敬爱的父亲!

冯 靛

2006年5月于求是园