

无线数据网络中机会调度理论模型的研究

雷 蕾, 林 闯

(清华大学计算机科学与技术系, 北京 100084)

摘 要: 机会通信作为一种对无线通信系统设计具有革命性意义的概念正在受到越来越多的关注. 目前对机会调度算法的研究往往采用不同的模型进行性能评价, 使得不同算法之间很难直接进行性能比较. 本文主要将各种用于机会调度算法研究的理论模型进行分类和比较, 对几种典型的机会调度算法在各类理论模型下的性能评价指标、方法和结果进行讨论, 最后探讨了该领域未来的发展趋势并给出了一些有待研究的方向.

关键词: 无线数据网络; 机会调度; 服务质量; 公平性; 稳定性

中图分类号: TP391.7 **文献标识码:** A **文章编号:** 0372-2112 (2007) 08 1548-10

Research on Opportunistic Scheduling Models in Wireless Data Networks

LEI Lei, LIN Chuang

(Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China)

Abstract: Opportunistic communication is receiving more and more attention as a concept that has recently revolutionized the way engineers think about wireless system design. Currently, opportunistic scheduling algorithms are studied based on many different models. Therefore, comparisons of results are sometimes difficult. This paper classifies and compares the models that have been proposed, and discuss the performance evaluation metrics, methods and results of several typical scheduling algorithms within these models. Finally, we discuss the future trends and present several topics remained to be studied in this area.

Key words: wireless data network; opportunistic scheduling; QoS; fairness; stability

1 引言

无线通信的基本特征之一是无线衰落信道的质量随时间发生随机变化. 对无线衰落信道中通信理论的研究可以追溯到上世纪 60 年代. 一直以来, 信道衰落都被视为一种导致不可靠传输的消极因素, 需要通过分集和功率控制等技术进行补偿. 然而, 自从上世纪 90 年代以来, 在利用信息论对衰落信道容量进行研究的过程中, 人们对无线信道衰落这一客观物理现象产生了一种全新的认识: 衰落所引起的信道质量的随机变化可以作为一种机会被加以利用, 即只有在信道质量变化到最佳状态时才发送信息, 从而使衰落信道的容量得到大幅度的提高, 甚至会超过无衰落信道的容量^[1,2]. 这种机会通信方式主要应用于多用户环境下, 由于各个用户信道质量的随机变化相互独立, 不同用户在同一时刻的信道质量不同, 因而用户数目越多则任一时刻存在某一用户信道质量处于最佳状态的概率越大, 这种现象被称为多用户分集效应^[3].

为了满足高速增长的数据业务和多媒体应用对网络带宽和服务质量 (Quality of Service: QoS) 提出的更高要求, 第三代移动通信系统 (3rd generation: 3G) 的 3GPP 和 3GPP2 两大阵营分别推出了 UMTS HSDPA^[4] 和 CDMA2000 1xEV-DO^[5] 两种空中接口标准, 使得移动互联网用户可以在蜂窝系统中得到较高的数据速率. 这两种系统都采用了机会通信理论, 其下行 (前向) 链路均主要基于时分复用方式, 通过自适应调制和编码技术使得数据传输速率与信道质量相匹配, 并且使用调度机制将下行链路的时隙在各个用户之间进行分配, 从共享同一无线信道的多个用户中选择一个当前信道质量最好的用户发送数据, 使得每个时隙都可以达到最高的用户数据速率和系统吞吐量. 这种调度方式由于在为用户分配时隙时利用了机会通信的思想, 因而被称为“机会调度”. 合理的机会调度算法不仅需要考虑最大化系统的吞吐量, 还要兼顾公平性, 稳定性和用户服务质量要求等.

与有线环境下的分组调度相比, 无线信道的时变特性加大了机会调度算法设计和评价的难度. 近年来对机

会调度算法的研究很多,然而不同的研究往往关注不同的性能指标,并且采用不同的模型进行性能评价,使得不同算法之间很难直接进行性能比较.因此,本文将现有各种用于机会调度算法研究的理论模型归纳分类,介绍和比较几种典型的机会调度算法在各类理论模型下的性能评价方法和结果.通过以上归纳、分析和比较,对该领域有待进一步研究的方向和可以采用的理论工具提出一些建议.

2 理论模型及性能指标

2.1 分组级模型与流级模型

目前研究机会调度算法的理论模型大体可以分为以 Andrew 等为代表进行研究的分组级模型(packet-level model)^[1]和以 Borst 和 Donald 等为代表进行研究的流级模型(flow-level model).本小节对这两种模型进行概述和比较,在第三部分中将进行详细讨论.

分组级模型的时间规模为调度算法工作的频率.其业务模型通常假设用户数目静态固定,即不会有新用户加入,也不会有用户由于业务完成而离开.分组级业务模型可以分为无限积压业务和分组级动态到达两种.在无限积压业务模型中,每个用户总是具有可发送的数据;在分组级动态到达模型中,调度器在每个时隙接收到各个用户到达的数据量是有限的.就服务器模型而言,各个用户的分组级服务速率等于各自信道的瞬时传输速率,它们由于信道质量的波动而随时间发生相互独立的随机变化.根据对信道衰落过程的不同假设,分组级模型中常用的服务器模型也可以分为两类:一类信道传输速率根据平稳随机过程产生,另一类信道传输速率通过敌人模型产生.

流级模型的时间规模为业务到达和离开的频率.流级业务模型假设每个用户具有一定量的业务,用户数目会随着业务的开始和完成而发生动态变化.各个用户的流级服务速率根据流体模型(fluid model)对信道速率进行时间平均后不随时间变化.在流级模型中,一般只能对平稳信道模型进行研究,而且有时还根据所有用户的信道速率统计变化幅度是否独立同分布,将平稳信道模型进一步细分为对称和不对称两种.

利用分组级和流级模型对机会调度算法进行性能分析各有利弊.分组级业务模型对于分组级的时间规模而言是合理的,但却无法描述流级动态,因而流级业务模型能够更准确的反映实际业务状况.另一方面,分组级服务器模型能够描述更复杂的无线信道特性(通过敌人模型等),而且流级模型的服务速率等于信道速率的时间平均值,这只有在信道速率的变化频率远大于业务到达和离开频率时才足够准确,而此条件在用户移动速度较慢时很可能无法满足.机会调度算法在

流级模型中可以等效为速率分配.显然,流级模型由于服务器模型的假设比较简单,因而更容易对速率分配方案的性能进行分析.然而,考虑到实际信道的时变特性,利用分组级的调度算法实现流级模型的速率分配方案往往并不容易.

表 1 机会调度理论模型分类

	分组级模型	流级模型
业务模型	无限积压业务	分组级动态到达
	流级动态到达	
服务器模型	平稳信道模型	对称平稳信道模型
	敌人模型	不对称平稳信道模型

2.2 性能指标

评价机会调度算法的好坏除了需要考虑有线环境下分组调度算法的常用性能指标,如公平性和 QoS 保障等,还包括有效性和稳定性两个重要方面.在有线环境下,任何工作保持调度算法都具有相同的有效性,即吞吐量.然而,无线衰落信道的时变特性导致不同调度算法所能达到的系统吞吐量可能有很大差别.另外,公平性、稳定性和 QoS 保障是机会调度算法设计中非常重要的三个性能指标,它们的进一步分类以及机会调度中常用的分析工具如表 2 所示,下面分别进行介绍.

2.2.1 公平性

对于机会调度算法而言,实现公平性和有效性的合理折中是一个非常重要的问题.调度算法应该能够充分利用信道变化过程中的峰值速率,在一个用户的当前信道质量优于其平均值的时隙选择为该用户服务.另一方面,调度算法还需要考虑公平性从而不会“饿死”信道质量较差的用户.在无线网络中,信道的时变特性导致对“公平性”可以有两种不同的解释:时间公平性和吞吐量公平性.前者是指令所有用户得到相同的服务时间;后者是指令所有用户发送的数据量相同.

关于公平性和有效性的合理折中在有线网络拥塞控制问题中得到了广泛的研究,提出了包括最大最小公平(max-min fairness)和比例公平(proportional fairness)等多种公平性准则.最大最小公平是指是对某一用户服务速率分配量的增加不会引起其它用户(这些用户的分配量已经同该用户一样,或者更小)服务速率分配量下降的情况下最大化该用户的服务速率分配量.在如蜂窝系统这样的单瓶颈网络中,最大最小公平会导致所有用户分配到的服务速率相同,因而它是一种优先考虑公平性的准则. Mazumdar 等指出从博弈理论的角度而言,最大最小公平性准则不如利用纳什交易解(Nash bargaining solution)更能准确描述公平的涵义^[7].以比例公平准则为代表的一类效用公平准则就是根据上述思想试图找到一个效用函数,并通过优化各个用

户服务速率的效用函数之和得到公平性和有效性之间的合理折中. 例如, 比例公平准则的效用函数为对数. 最大最小公平是效用公平的一种极限情况. 由于无线信道的时变性, 对机会调度算法效用公平的分析一般采用随机近似方法等.

最大最小公平是有线环境下调度算法最常用的公平性准则, 而在无线机会调度算法的研究中, 最大最小公平准则一般用于衡量时间公平性, 而以比例公平为代表的效用公平准则一般用于衡量吞吐量公平性. 目前已有研究成果表明, 时间的最大最小公平与吞吐量的比例公平之间具有一定的等效关系^[8].

2.2.2 稳定性

马尔可夫链的稳定性有三种级别递增的定义: ϕ 不可约性、常返性和遍历性. 其中遍历性的应用最为广泛, 它是指不管系统的初始状态如何, 在经历一段时间以后到达平衡状态, 此时系统各状态的概率分布不随时间而变, 亦不依赖于初始状态, 则称系统状态是遍历的. 在分组级模型中, 马尔可夫链的状态空间为队列长度, 或者说缓存中分组的数目; 在流级模型中, 马尔可夫链的状态空间为网络中用户的数目. 因此, 如果一个调度器在任何可能的情况下使得队列大小有界, 则称此调度器或该系统具有分组级稳定性; 如果一个调度器在任何可能的情况下使得网络中的用户数目有界, 则称此调度器或该系统具有流级稳定性. 由于分组级模型和流级模型所假设的业务模型和服务器模型均有所不同, 并且作为稳定性分析对象的马尔可夫链的状态变量所代表的含义也不相同, 所以同一种调度算法不一定同时具有分组级和流级稳定性. 例如, 目前已经证明比例公平算法不具有分组级稳定性, 却具有流级稳定性. 福斯特-李雅普诺夫漂移 (Foster-Lyapunov drift) 和流体极限 (fluid limit) 是两种用来证明调度算法稳定性的重要方法^[9]. 福斯特-李雅普诺夫方法从控制论中经典的李雅普诺夫稳定性方法发展而来, 可以用来对随机模型的稳定性进行分析, 但在复杂情况下可能比较难于求解. 流体极限则在原网络及其流体模型之间建立稳定性的等效关系, 并通过利用李雅普诺夫等方法证明流体模型这种确定性系统的稳定性, 从而达到证明原网络稳定性的目的, 然而这种等效关系需要以系统的马尔可夫特性作为前提条件.

表 2 机会调度算法的主要性能指标及分析工具

	分类		分析工具
	时间公平	最大最小公平	随机近似方法、效用函数
公平性	吞吐量公平	效用公平	
稳定性	分组级稳定	流级稳定	福斯特-李雅普诺夫漂移、流体极限
QoS 保障			有效容量等、随机网络演算

2.2.3 QoS 保障

服务质量主要是指为用户提供性能保证和服务区分^[10]. 体现服务性能的关键参数有带宽、时延、抖动和丢包率等, 对服务性能的保证应该是可预测的; 服务区分意味着为不同的业务需求提供不同的性能保证, 对于一些关键性业务, 即使在高负载的情况下, 也要保证其 QoS 不受影响. 公平性和稳定性对 QoS 保障具有重要的意义, 却都不足以完全提供有效的 QoS 保障. 公平性的实现可以隔离不同的业务流, 让不同的流只享用分配给自己的带宽, 这样即使存在恶意或高突发性业务, 也不致影响到其他的正常业务流. 然而, 它一般不考虑服务区分, 也无法为带宽以外的其他 QoS 参数提供保障. 稳定性的实现可以保证业务具有有限的时延和丢包率等, 但无法得到具体参数值, 因而无法保证满足业务的需求.

排队论、网络演算和有效带宽是目前研究网络服务质量问题的三种典型的理论工具, 分别代表随机模型、演算模型和功能模型, 但它们都存在各自的局限性. 排队论着重于研究随机排队系统在稳定状态下的平均性能指标. 它所得到的大量分析结果需要以业务的马尔可夫或无记忆特性作为前提条件, 而大量统计数据显示真实网络业务在很多情况下不符合此假设条件. 网络演算利用最小加代数和系统理论研究确定排队系统的最差性能指标, 通过流量整形和调度达到对业务的服务质量保障. 虽然网络演算所得到最差性能边界是严格的, 但这种最差情况很少发生. 尤其是考虑聚合业务时, 网络演算不能利用统计复用增益节约网络资源. 有效带宽描述为一定量业务提供期望的服务质量保障所需的最小带宽^[11]. 目前, 已经得出包括自相似等多种业务类型的有效带宽表达式, 并且它可以描述业务的统计复用增益. 然而, 有效带宽无法用来对各种调度算法进行研究. 上述三种理论由于各自的局限性都没有在 QoS 研究中得到十分广泛的应用. 近年来, 一种在网络演算基础上发展起来的随机网络演算理论得到了广泛的研究, 它可以用来定量求解网络性能的统计边界, 得出网络所能提供的统计服务质量保障, 弥补上述三种理论的不足, 被普遍认为是未来网络研究中非常重要的 QoS 理论工具^[12]. 随机网络演算和基于有效带宽发展而来的有效容量等是目前机会调度 QoS 分析的主要工具.

3 基于不同理论模型的算法性能评价

3.1 机会调度算法分类

由于无线机会调度在本质上与有线环境下的分组调度相同, 因此对分组调度算法的各种分类标准对机会调度算法^[11]都适用. 然而, 根据机会调度的特点和目

标, 本文将各种机会调度算法分为贪婪算法、效用公平算法、时间公平算法和队列感知算法等。

贪婪算法是最简单和基本的机会调度算法, 它在每个时隙都选择瞬时信道质量最好的用户发送数据, 因而能够达到最大的系统吞吐量, 但却忽略了算法的公平性、稳定性和 QoS 保障。

效用公平和时间公平算法是以效用公平性和时间公平性作为目标的机会调度算法。轮询算法是最简单的时间公平调度算法, 但它完全不考虑用户信道状态, 与贪婪算法分别代表最大化公平性和有效性的两种极端情况。在第 3.2.1 小节中详细介绍的比例公平调度算法能够在公平性和有效性之间实现较为合理的折中, 是一种效用公平算法, 目前在实际系统中被广泛采用。在对称平稳信道模型下比例公平调度算法也具有长期时间公平性。

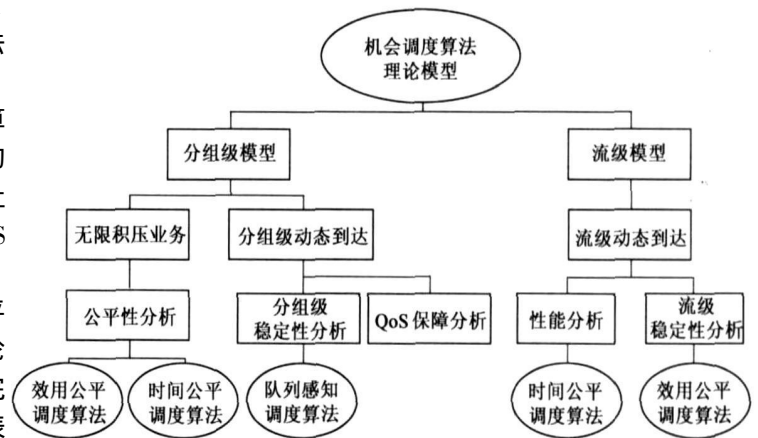
以上机会调度算法在决策时只考虑了信道状态, 因而又称为信道感知调度算法。与此相对应, 一类同时考虑信道状态和队列状态的机会调度算法被称为队列感知调度算法, 将在第 3.2.2 小节进行详细介绍。最大权值算法和跟踪算法分别是在平稳信道模型和敌人信道模型下设计的队列感知调度算法, 前者包括 M-LWDF、M-LWWF 和 EXP 算法等。

不同理论模型对业务和信道过程的假设以及所关注的性能指标往往不同, 因此对各类机会调度算法的评价也不同, 如图 1 所示。

3.2 分组级模型

假设在一个无线小区内 n 个移动数据用户需要由同一个基站进行服务。由于网页浏览等业务的主要数据流都只占用下行链路, 因此这里只考虑下行(基站到移动台)方向。基站为每个移动用户的数据维护一个单独的队列。时间被划分为时隙, 并且基站在每个时隙只能为一个用户传送数据。为了确定在某一时刻内被服务的用户, 基站在所有时隙维护一个向量 $(r_0(t), \dots, r_{n-1}(t))$, 其中 $r_i(t)$ 代表在时隙 t 可以为用户 i 传送的数据量。在 1x EV-DO 系统中, 移动用户 i 在每个时隙测量由基站发送的导频信号强度, 并根据信号的强度计算基站和自身之间的信道质量并确定基站为实现低误码率应发送的速率大小。最后, 用户通过一个控制消息将此速率值告知基站, 从而使位于基站端的调度器获得 $r_i(t)$ 的值。

机会调度问题的正式描述如下: 调度器在每个时隙接收到信道速率向量 $(r_0(t), \dots, r_{n-1}(t))$, 并据此确



定在该时隙接受服务的用户。如果用户 i 被选择, 则从用户 i 的队列中发送 $r_i(t)$ 比特数据。

不同的模型对业务到达过程以及基站和用户之间的信道过程的假设不同。通常考虑的业务模型可以分为两类: 无限积压业务(infinite backlog)模型和分组级动态到达模型。

(1) 无限积压业务模型: 每个用户总是具有可发送的数据。由于实际中并不存在这样的到达过程, 考虑队列长度和时延等指标不具有太大的意义。这种模型中的优化目标一般为最大化某个关于用户可达吞吐量的函数。例如, 设 R_i 代表用户 i 的长期吞吐量, 则一个常见的目标为优化比例公平指标 $\sum_i \log R_i$ 。

(2) 分组级动态到达模型: 调度器在每个时隙接收到一个向量 $(a_0(t), \dots, a_{n-1}(t))$, 其中 $a_i(t)$ 代表用户 i 在时隙 t 到达的数据量。在此情况下, 除了吞吐量之外, 队列大小与时延也具有意义。任何调度算法的一个基本目标为稳定性。

信道过程模型也可以分为两类: 一类信道速率根据平稳随机过程产生, 另一类信道速率通过敌人模型产生。

(1) 平稳信道过程: 信道速率根据平稳随机过程产生, 假设信道状态为一个有限集合 $M = \{1, \dots, M\}$ 。每个状态 $m \in M$ 对应一组信道速率 $(\mu_1^m, \dots, \mu_N^m)$, 每个信道速率 μ_i^m 代表如果在时隙 t 内信道状态为 m , 将此时隙全部分配给用户 i 可以发送的比特数目。信道在 t 时刻所处状态是一个随机过程, 具体而言是一个具有 M 个有限状态的遍历马尔可夫链。当 $m(t) = m$ 时, 信道速率向量为 $r_i(t) = \mu_i^m$ 。在此模型中, 目标一般为得到某种指标的“最优”调度规则。

(2) 敌人信道过程: 假设信道过程不满足平稳性。在每个时隙 t 内, 信道速率向量 $(r_0(t), \dots, r_{n-1}(t))$ 可以由敌人定义的任何向量, 而敌人试图为调度算法设计设置最

多的麻烦. 在敌人模型下, 由于在某一速率向量下的最优调度算法可能随着时间发生变化, 因此通常假设在每个时隙敌人具有自己的调度算法, 可以与信道速率向量相结合产生最好的性能. 因此, 在此类模型下, 优化的目标是与敌人的调度算法尽可能的接近.

3.2.1 无限积压业务模型下的公平性分析

(1) 比例公平调度算法的公平性分析

此模型假设所有用户总是具有待发送的数据, 并且信道速率由一个平稳随机过程产生. 由此模型产生了最为广泛使用的机会调度算法, 即 Tse^[14] 提出的比例公平调度算法. 在每个时隙, 比例公平调度算法为满足下面条件的用户 j 服务:

$$j = \arg \max_i \frac{r_i(t)}{R_i(t)} \quad (1)$$

其中 $R_i(t)$ 为时刻 t 的指数过滤平均服务速率值 (exponentially filtered average service rate), 它根据下面的公式进行更新:

$$R_i(t+1) = \begin{cases} (1-\tau)R_i(t) + \tau r_i(t), & \text{if } i=j \\ (1-\tau)R_i(t), & \text{if } i \neq j \end{cases} \quad (2)$$

其中 τ 为时间常数 (实际中 τ 通常取 1000 倍时隙的数量值). 比例公平算法给予具有较高瞬时信道速率值 ($r_i(t)$) 和较低当前平均服务速率值 ($R_i(t)$) 的用户较高的优先级.

比例公平调度算法具有一个很好的理论特点: 对于所有可能的调度规则, 它最大化比例公平指标 $\sum_i \log R_i$ 的值, 其中 $R_i = \lim_{t \rightarrow \infty} R_i(t)$ 为用户 i 的长期平均服务速率. 之所以将 $\sum_i \log R_i$ 称为比例公平指标, 是由于如果 $(R_0^*, \dots, R_{n-1}^*)$ 是最大化 $\sum_i \log R_i$ 的可行速率向量, 则对于任意其他的速率向量 (R_0, \dots, R_{n-1}) , 公式 $\sum_{i=0}^{n-1} (R_i - R_i^*)/R_i^* < 0$ 总是成立, 即所有用户的相对服务速率改变值之和为负. 其中用户 i 的相对服务速率改变值等于用户 i 的服务速率改变值 $(R_i - R_i^*)$ 除以用户 i 的原有服务速率值 R_i^* . 另一种对比例公平指标的解释是, 将某一用户的速率乘以一个因子 c 与将另一用户的速率乘以一个因子 c 对优化目标的影响相同. 最后, 注意到使用 $\sum_i \log R_i$ 作为指标可以保证不会完全“饿死”任何用户, 这是由于 $\log 0 = -\infty$.

由式(1)和式(2)的调度算法可以实现对比例公平指标 $\sum_i \log R_i$ 的优化, 下面给出不严格的证明.

令 $S(t) = \sum_i \log R_i(t)$, 则 $(\nabla S)(t) = \left(\frac{1}{R_0(t)}, \dots, \frac{1}{R_{n-1}(t)} \right)$. 假设在时隙 t 选择用户 j 接收服

务, 则根据式(2)可以得到:

$$\begin{aligned} & \sum_i \log R_i(t+1) - \sum_i \log R_i(t) \\ & \approx (\nabla S)(t) \cdot (R_0(t+1) - R_0(t), \dots, R_{n-1}(t+1) - R_{n-1}(t)) \\ & = \left(\frac{1}{R_0(t)}, \dots, \frac{1}{R_{n-1}(t)} \right) \cdot (R_0(t+1) - R_0(t), \dots, R_{n-1}(t+1) - R_{n-1}(t)) \\ & = \left(\frac{1}{R_0(t)}, \dots, \frac{1}{R_{n-1}(t)} \right) \cdot (-\tau R_0(t), \dots, \tau r_j(t) - \tau R_j(t), \dots, -\tau R_{n-1}(t)) \\ & = \frac{\tau r_j(t)}{R_j(t)} - \tau n \end{aligned}$$

因此, 为了最大化 $\sum_i \log R_i(t)$ 的变化值, 应该选择能够使得 $r_i(t)/R_i(t)$ 最大化的用户. 而这正是式(1)定义的调度算法. 关于比例公平调度算法优化性的严格证明在参考文献[15~17]中已经给出.

3.2.2 分组级动态到达模型下的稳定性分析

通过上述分析可知, 比例公平算法在每个用户都总是具有需要服务的数据时可以实现公平的速率分配. 然而, 在某些情况下不同用户具有不同数量的数据需要被服务, 而此时的目标是为所有的数据服务. 假设每个用户的业务根据某个平稳随机过程到达, 令 $a_i(t)$ 为用户 i 在时隙 t 到达的数据量, 将由 $(r_i(t), a_i(t))$ 定义的随机过程称为系统的输入过程. 令 $q_i(t)$ 为用户 i 在时隙 t 等待服务的数据量. 队列过程的更新如下. 如果用户 i 在时隙 t 被服务, 则

$$q_i(t+1) = [q_i(t) + a_i(t) - r_i(t)]^+$$

否则

$$q_i(t+1) = q_i(t) + a_i(t)$$

令 λ_i 为用户 i 的平均到达速率. 如果对于某一 $\varepsilon > 0$, 存在一个 SSS 规则使得平均长期服务速率 R_i 大于 $(1-\varepsilon)\lambda_i$, 则称此输入过程“可调度”. 机会调度算法应该是稳定的, 即只要输入过程可调度, 则调度算法应使得排队过程具有平稳的分布, 从而使队列长度不会漂移到无穷大.

(1) 比例公平调度算法的分组级稳定性分析

文献[18]证明, 比例公平调度算法在分组级模型下不稳定. 这可以通过一个两用户的简单例子进行说明. 假设它们的到达过程为一常量, 对所有 t 有 $a_1(t) = 49$ 和 $a_2(t) = 94$. 另外, 假设用户 2 的信道过程为一常量, 对所有 t 有 $r_2(t) = 100$; 而用户 1 的信道过程是一个周期为 10 的周期过程:

$$r_1(t) = \begin{cases} 1000, & \text{if } t \bmod 10 = 0 \\ 100, & \text{otherwise} \end{cases}$$

可以证明此例中业务是可调度的, 只需令 $t \bmod 20 = 0$ 时用户 1 接受服务, 而在其余时隙用户 2 接受服务. 换言之, 将 $r_1(t) = 1000$ 的所有时隙中的一半分配给用户

1, 其余所有时隙分配给用户 2. 这样, 用户 1 的平均服务速率为 50, 用户 2 的平均服务速率为 95, 均大于各自的业务到达速率.

然而, 比例公平调度算法无法进行正确的时隙分配, 用户 2 接受服务的时隙数只占总时隙数的 9/10. 因此用户 2 的平均服务速率只有 90, 小于他的业务到达速率 94, 这意味着用户 2 的队列长度将无限增长.

(2) 队列感知调度算法的分组级稳定性分析

比例公平算法之所以不具有分组级稳定性, 是由于它只考虑了各个用户之间的公平性, 而没有考虑各个用户可能具有不同的业务量. 如果令调度算法在决策时不但考虑信道质量, 而且考虑队列长度, 就可以在用户队列长度过长时采取措施, 这就是队列感知调度算法的基本思想. 一种最简单的队列感知调度算法总是选择具有最大的 $q_i(t) r_i(t)$ 值的用户进行服务, 因此称为最大权重算法. 此算法令具有较大瞬时信道速率和队列长度的用户具有较大的优先级, 可以证明该类算法具有分组级稳定性^[19~22].

根据福斯特-李雅普诺夫漂移 (Foster-Lyapunov drift) 方法, 只要证明最大权重调度算法的队列长度的范数 $\|q(t)\| = \sqrt{\sum_i (q_i(t))^2}$ 具有负漂移, 就可以证明该算法稳定. 为此, 令 $x_i(t)$ 代表在最大权重调度算法下用户 i 在时隙 t 发送的数据量; 令 $y_i(t)$ 代表在最优 SSS 规则下用户 i 在时隙 t 发送的数据量. 根据最大权重算法的定义, 有 $\sum_i q_i(t) x_i(t) \geq \sum_i q_i(t) y_i(t)$. 假设业务可调度, 则根据定义对于任何较大的 w , $E[\sum_{t \leq i \leq t+w} (y_i(t') - a_i(t'))] > \epsilon w \lambda_i$ 有对于所有的 i 和 t 成立. 因此,

$$\begin{aligned} \|q(t+1)\|^2 &= \sum_i (q_i(t) + a_i(t) - x_i(t))^2 \\ &= \|q(t)\|^2 + \sum_i q_i(t)(a_i(t) - x_i(t)) + \sum_i (a_i(t) - x_i(t))^2 \\ &\leq \|q(t)\|^2 + \sum_i q_i(t)(a_i(t) - y_i(t)) + \sum_i (a_i(t) - x_i(t))^2 \end{aligned}$$

由于对于所有 t 和较大的 w , 可以得到不等式 $E[\sum_{t \leq i \leq t+w} (y_i(t') - a_i(t'))] > \epsilon w \lambda_i$. 因此可以利用此不等式证明在较长的时间段内一旦某些队列长度变大则 $\|q(t)\|$ 具有负漂移.

以上对分组级动态业务模型下的稳定性分析均以平稳随机信道模型作为假设条件. 文献[23]对敌人信道模型下的分组级稳定性进行研究, 指出最大权重调度算法在敌人模型下会导致队列长度变大, 并提出一种“跟踪算法 (tracking algorithm)”能够保持敌人信道模型下的分组级稳定性. 跟踪算法将时间分段, 并在每段时间结束时计算该时间段内的最优化调度算法, 并在下

一时间段内运用该调度规则, 从而保证尽可能的接近“敌人”的调度算法.

3.3 流级模型

Robert 和 Massoulié^[24] 引入并研究了互联网拥塞控制的流级模型, 用网络中的一个流代表连续传输的单个文件, 流的数量随机变化并且流之间动态共享网络带宽. 此模型假设“区分时间规模”, 即流级动态的时间规模 (即, 文件的到达和离开) 比分组级动态的时间规模要大得多. 此后, Veciana^[25] 和 Bonald^[26] 分别研究了在不同带宽共享策略下流级模型的稳定性问题. 每种带宽共享策略对应于一种广义的处理器共享 (Processor Sharing) 规则. 在对无线机会调度算法的研究中, Borst 和 Bonald 等人运用上述流级模型的建模分析方法对以比例公平算法为代表的时间公平调度算法和效用公平调度算法的服务性能和流级稳定性进行分析.

(1) 比例公平调度算法的服务性能分析

在分组级模型中将信道过程分为平稳随机模型和敌人模型, 在流级模型中所有的分析都基于平稳随机过程信道模型假设. 令随机向量 (r_0, \dots, r_{n-1}) 代表不同用户的瞬时速率, 它们的概率分布与信道速率过程 $(r_0(t), \dots, r_{n-1}(t))$ 的联合平稳分布相同. 令 $r_i = a_i \times b_i$, $i = 0, \dots, n-1$, 其中时间平均速率 $a_i = E(r_i)$ 为一常量, 一般与用户距离基站的远近有关, 随机变量 b_0, \dots, b_{n-1} 代表不同用户信道速率围绕时间平均速率的瞬时波动. 为了利用流级模型对服务性能进行分析, 需要将平稳信道模型进一步细分:

(a) 对称信道模型: 用户具有不同的时间平均速率 a_i , 但信道过程是对称的, 即不同用户围绕时间平均速率的相对速率变化幅度 b_i 是统计相同的. 这一假设在用户具有瑞利衰落信道并且速率与 SNR 近似呈线性关系时是粗略正确的.

(b) 不对称信道模型: 对称信道模型在 SNR 不太高时是合理的. 实际中, 当不同用户的时间平均速率由于时间平均信噪比 (SNR) 的不同而不同时, 他们的相对速率波动通常不会具有完全相同的概率分布. 一般来说, 由于传输速率在高 SNR 条件下为凹函数, 且 SNR 增加到一定程度后传输速率具有非递增性, 因此相对速率波动将随着 SNR 的增加而减少, 而增益因子将不再独立于用户的时间平均速率, 即 b_i 随着 a_i 的增加而减少.

文献[27]证明, 比例公平算法在对称信道模型下具有长期时间公平性, 即在一个较长的时间段内它分配给每个用户的服务时间相同. 基于比例公平算法的这一性质, 文献[28, 29]对 1x EV-DO 系统的服务性能进行了分析. 假设一个小区中有 n 个移动数据用户, 在对称信道模型下比例公平调度算法为用户分配的时隙数为总时隙数的 $1/n$.

由于 1.67ms 的时隙长度与用户可感知的时间规模相比很小,因而可以假设每个用户被连续服务,其服务速率为总体传输速率的 $1/n$. 这一假设类似于在有线网络中利用经典的处理器共享模型对轮询调度进行抽象. 不同的是,总体传输速率不是一个固定值,而是由具有多用户分集效应的比例公平算法确定,因而与用户数 n 有关. 比例公平算法在每个时隙选择具有最大相对速率的用户,即经过指数过滤平均服务速率值 $R_i(t)$ 归一化后的最大瞬时速率. 不失一般性的假设所有用户的衰落过程统计相同(不一定独立),因此不同用户的 $R_i(t)$ 应该也具有相同的分布. 另外,当指数过滤的时间常数 τ 足够大时,指数过滤平均服务速率值随时间的波动不大. 将上述两点结合,可以得出所有用户的指数过滤平均服务速率值应该在同一个常数附近进行极小的波动. 因此,比例公平算法在每个时隙选择具有最大瞬时速率的用户, n 个用户的总体传输速率为 $H(n) = E\{\max\{r_1, \dots, r_n\}\}$. 由于假设所有用户的信道过程统计相同,可以令 $r_i = a \times b_i$, $i = 0, \dots, n-1$, 因此 $H(n) = H(1)G(n)$, 其中 $H(1) = a$ 为时间平均速率. 例如,如果用户具有独立瑞利衰减信道,且瞬时速率与瞬时信噪比成线性关系,则 b_0, \dots, b_{n-1} 为具有单位均值的独立指数分布随机变量. 在此情况下, $G(n) = \sum_{m=1}^n 1/m$. 当 n 趋于无穷大时 $G(n)$ 趋于 $\log(n)$. 在实际的 1xEV-DO 系统中,瞬时速率从一组有限的离散值中选择. 将 $G(n)$ 称作增益因子,它代表比例公平算法通过机会调度所得到的相对于时间平均速率 a 的吞吐量增益.

假设用户的数目随业务的到达和完成而动态变化,当有 n 个用户时每个用户都以 $H(n)/n$ 的速率被连续服务,此速率为时间平均速率的 $G(n)/n$ 倍. 此时,根据服务器共享模型就可以得到比例公平调度算法的用户数目、业务时延等参数的概率分布函数,从而对 1xEV-DO 系统的性能进行分析. 文献[30]利用相似的方法对 HSDPA 系统的性能进行分析.

上述分析假设对称信道模型,在不对称信道模型下,PF 算法由于不满足时间公平性,只能使用区分服务器共享(Discriminatory Processor Sharing: DPS)模型进行分析,无法得出各种参数的概率分布函数. 因此,只有具有时间公平性的机会调度算法才能利用上述方法进行性能分析.

(2) 比例公平调度算法的流级稳定性分析

在第 3.1.2 小节中已经证明比例公平算法不具有分组级稳定性,然而文献[31]采用流体极限(fluid limit)方法证明包括比例公平在内的一类效用公平调度算法具有流级稳定性. 首先,将网络中的用户按照离基站的距离分类,同一类用户具有几乎相同的时间平均速率. 令 $n(t) := (n_1(t), \dots, n_K(t))$, 其中 $n_k(t)$ 代表第 k 类

用户在时隙 t 的总用户数. $n_k(t)$ 的流体极限为:

$$N_k(t) = \lim_{\omega \rightarrow \infty} \frac{n_k(\omega t)}{\omega}$$

其中 $\sum_{k=1}^K n_k(0) = \omega$.

定义如下形式的李雅普诺夫函数:

$$F(x_1, \dots, x_K) = \sum_{k=1}^K \frac{\theta_k}{\mu_k} \rho_k^{-1} \frac{x_k^{1+\eta}}{1+\eta}$$

根据控制论中证明系统稳定性的李雅普诺夫方法,只要证明网络中各类用户数目的李雅普诺夫函数的导数为负,就可证明流级稳定性. 即证明:

$$\frac{d}{dt} F(N(t)) = \sum_{k=1}^K \theta_k \rho_k^{-1} N_k(t)^\eta (\rho_k - S_k(t)) < 0 \quad (3)$$

文献[31]证明采用效用公平调度算法和队列感知调度算法后,式(3)都成立,因而这两类调度算法都具有流级稳定性.

效用公平调度算法之所以具有流级稳定性,是由于调度算法会给予具有较多用户数目的那些类用户较高的优先级. 这一点与队列感知调度算法优先考虑队列长度较长的用户,从而具有分组级稳定性类似.

3.4 用于 QoS 分析的新模型

上述基于分组级和流级模型的研究主要对机会调度算法的公平性和稳定性进行分析,一般无法得出算法所能提供的具体 QoS 指标. 虽然在流级模型中,利用服务器共享模型能够对调度算法进行性能分析,但要求算法必须具有时间公平性. 上述研究的这一局限性主要源于对无线信道模型的描述较为粗略,至多假设信道速率变化符合平稳随机过程,而对该随机过程的一阶和二阶特性不作任何假设.

无线信道的衰落过程虽然具有很大随机性,但目前仍然存在一些简单有效的模型可以对无线信道特性进行抽象,用于调度算法和传输控制等上层无线网络技术的研究. 如 Elliot-Gilbert 模型和多状态马尔可夫信道模型等. 因此,一些研究利用这些信道模型,构建基于有效带宽、随机网络演算的无线衰落信道 QoS 模型. 虽然目前这些模型大多基于单用户系统假设,然而近年来开始出现一些文献运用该类 QoS 模型对机会调度算法进行研究. 按照上述分组级和流级模型分类,这类 QoS 模型一般假设分组级动态到达业务模型.

(1) 基于有效带宽的 QoS 分析

首先介绍有效带宽的定义. 将业务到达过程表示为 $\{A(t), t \geq 0\}$, 其中随机过程 $A(t)$ 代表在时间段 $[0, t)$ 内数据源发出的比特数,则 $A(t)$ 的有效带宽定义为:

$$q_1(s, t) = \lim_{t \rightarrow \infty} \frac{1}{t} \log E[e^{sA(t)}] \quad (4)$$

在无线网络中,可以将信道过程表示为 $\{C(t), t \geq 0\}$,

其中随机过程 $C(t)$ 代表在时间段 $[0, t)$ 内通过无线信道传输的比特数. 目前有三种基于有效带宽发展而来的无线网络 QoS 模型: 有效容量^[32], 无线有效带宽^[33, 34]和有效信道容量^[35].

文献[32]利用业务模型与服务器模型的对偶性, 将有效带宽理论进行扩展, 定义有效容量为:

$$\alpha^{(c)}(s, t) = - \lim_{t \rightarrow \infty} \frac{1}{st} \log E[e^{-sC(t)}] = \alpha_c(-s, t) \quad (5)$$

有效带宽理论研究具有不同统计特性和 QoS 要求的业务所需要的网络传输速率, 而有效容量则研究具有某种统计特性的无线信道可以为具有某种 QoS 要求的业务所提供的网络传输速率. 文献[36]利用有效容量概念对贪婪机会调度算法和队列感知机会调度算法进行分析. 有效容量虽然可以对机会调度算法所提供的 QoS 进行研究, 但必须假设业务到达率固定, 因而无法描述真实的业务特性.

无线有效带宽则可以描述业务的统计特性, 它分别求出业务到达过程和信道过程的有效带宽 $\alpha_d(s, t)$ 和 $\alpha_c(s, t)$, 并根据定理 1 求出积压业务和时延的近似性能边界.

定理 1 如果存在 $s^* > 0$, 使得 $\alpha_d(s^*, t) + \alpha_c(-s^*, t) \rightarrow 0, t \rightarrow \infty$, 则 $\Pr(W > x) \sim e^{-s^*x}, x \rightarrow \infty$.

其中, W 代表积压业务的比特数. 定理 1 根据渐进方法 (Asymptotic Approximation) 得到, 它提供的性能边界不够准确. 为此, 文献[35]提出了有效信道容量的概念:

$$S_{\Delta}^e(t) = \sup_{P\{C(t) \leq X\} \leq \epsilon} X \quad (6)$$

并得到有效信道容量与有效带宽的定量关系式:

$$S_{\Delta}^e(t) = \sup_{s < 0} \{t\alpha_c(s, t) - \frac{\log \epsilon}{s}\} \quad (7)$$

根据有效信道容量的定义, 可以得到积压业务和时延的较为严格的边界. 对无线有效带宽和有效信道容量的研究目前都只考虑了单用户系统的情况, 而没有对多用户系统中机会调度算法的 QoS 问题进行研究.

(2) 基于随机网络演算的 QoS 分析

如前所述, 利用有效容量虽然可以对机会调度算法所提供的 QoS 进行分析, 却无法描述真实的统计业务特性. 另一方面, 无线有效带宽和有效信道容量两种模型具有类似的思路, 都通过分别求出业务到达和信道速率这两个随机过程的有效带宽, 得到网络提供的统计 QoS 保障, 如积压业务和时延等. 这种思路与近年来兴起的另一种 QoS 理论工具-随机网络演算具有很大的-一致性. 随机网络演算通过随机到达曲线和随机服务曲线分别描述业务和无线信道的统计特性, 它作为一种系统化的理论工具比无线有效带宽和有效信道容量具有更强的严密性和普适性.

近年来, 随机网络演算已经成为网络演算领域的一个重要研究方向. 然而, 随机网络演算对随机到达曲

线和随机服务曲线的定义必须保持网络演算的原有基本特征. 在采用最小加代数的网络演算中这些属性全部成立, 但直到 2006 年, Y Jiang 和 M Fidler 才分别在 Sigcomm 和 IWQoS 会议上发表文章使得随机网络演算在最小加代数下可以满足上述所有五个基本属性. 前者定义了最大(虚拟)积压业务中心(maximum (virtual)-backlog centric: m. b. c) 随机到达曲线和随机服务曲线^[38], 后者提出了一种基于矩母函数(Moment Generating Function: MGF)的随机网络演算^[39]. 文献[38]证明了两随机网络演算之间具有一定的联系.

长期以来, 随机网络演算作为一种数学工具自身不够完善, 在很大程度上阻碍了它的应用. 然而, 2006 年发表的两篇论文^[38, 39]对它进行了极大的完善. 因此, 如何利用这一崭新的数学工具对无线机会调度算法所提供的统计服务质量进行研究, 是一个充满挑战又急待解决的问题. 目前在这一领域的研究还不多. 文献[40]虽然最早利用随机网络演算对无线信道链路模型进行研究, 但当时随机网络演算本身还不够完善. 文献[41]和[42]分别利用文献[39]和[38]提出的较为完善的随机网络演算对无线机会调度算法进行研究, 然而只针对最简单的开关信道模型和贪婪算法进行了分析.

4 总结与展望

目前针对无线机会调度算法的公平性和稳定性已经进行了大量的研究, 公平性和稳定性虽然是衡量调度算法的重要性能指标, 却不足以为用户提供有效的 QoS 保障. 另一方面, 朗讯贝尔实验室在 2006 年 5 月发布的 IPTV/VOD 研究报告中指出, 未来五年 IPTV/VOD 的业务量将增长一个数量级, Web 浏览、FTP 等尽力而为的数据业务在网络总业务量中所占的比重将由 2005 年的 90% 下降至 2010 年的 50%^[43]. 多媒体应用的迅猛发展要求必须充分利用有限的无线频率资源为用户提供包括性能保证和服务区分在内的 QoS 保障. 目前十分缺乏有效的理论工具对无线机会调度算法进行 QoS 研究. 虽然也有一些文献对调度算法的性能和 QoS 进行分析, 但所使用的理论工具都带有很大的局限性. 例如, 利用服务器共享模型只能够对具有时间公平性的调度算法进行性能分析; 利用有效容量进行 QoS 分析则对业务模型具有极大的限制. 近年来随机网络演算的发展为机会调度算法的研究提供了新的机遇与挑战. 随机网络演算理论在 2006 年取得了若干突破性的成果. 虽然各种随机网络演算模型还有待进一步的融合和统一, 但目前已经为机会调度算法的 QoS 研究提供了较为完善的理论基础. 由于随机网络演算是一项崭新的理论, 利用它对机会调度算法进行 QoS 研究的

文献刚刚开始出现,还处于初期的探索阶段.

上述关于分组级和流级模型的研究均以 1x EV DO 或 HSPDA 等 3G 系统作为研究对象. 另一方面,未来超三代(Beyond 3G: B3G)系统的实现也必然采用机会通信理论. 未来无线系统将采用多入多出(Multiple Input Multiple Output: MIMO)、正交频分复用(Orthogonal Frequency Division Multiplexing: OFDM)等宽带无线技术. 因此,如何在采用 MIMO、OFDM 技术的无线系统中设计机会调度算法也是目前研究的热点问题^[44~46]. 另外,虽然目前对机会调度的绝大多数研究都面向蜂窝网络,但针对多跳网络中机会调度算法的研究^[47~49]也正在开始受到关注.

总之,对机会调度算法的研究一定要立足于满足高速增长的数据业务和多媒体应用对网络带宽和服务质量所提出的更高要求. 目前,机会调度算法的统计 QoS 问题,以及在 MIMO、OFDM 系统和多跳网络中的应用问题等方面还存在较大的研究空间,需要更多具创新性的研究.

参考文献:

- [1] D Tse, S Hanly. Multi access fading channels: Part I: polymer structure, optimal resource allocation and throughput capacities[J]. IEEE Trans on Information Theory, 1998, 44(7): 2796–2815.
- [2] S Hanly, D Tse. Multi access fading channels: Part II: Delay limited capacities[J]. IEEE Trans on Information Theory, 1998, 44(7): 2816–2831.
- [3] R Knopp, P Humblet. An information capacity and power control in single cell multiuser communications[A]. Proc of IEEE ICC[C]. Seattle, 1995. 331–335.
- [4] 3GPP TS 25. 848, Physical layer aspects of ultra high speed downlink packet access[S]. v4. 0.0, Release 4, 2001.
- [5] 3GPP2 C. S0024 Version 4. 0, CDMA 2000 high rate packet data air interface specification[S]. October 2002.
- [6] M Andrews. A survey of scheduling theory in wireless data networks[A]. Proc of the 2005 IMA summer workshop on wireless communications[C]. University of Minnesota, 2005.
- [7] F P Kelly, A K Maulloo, D K H Tan. Rate control for communication networks: shadow prices, proportional fairness and stability[J]. Journal of the Operational Research Society, 1998, 49: 237–252.
- [8] S C Liew, Y J Zhang. Proportional fairness in multi channel multi rate wireless networks[A]. Proc of IEEE Globecom[C]. San Francisco, CA, USA, 2006. 1–6.
- [9] G Song. Cross layer resource allocation and scheduling in wireless multicarrier networks[D]. Atlanta: Georgia Institute of Technology, 2005.
- [10] 林闯, 单志广, 仁丰原. 计算机网络的服务质量[M]. 北京: 清华大学出版社, 2004.
Lin C, Shan Z G, Ren F Y. Quality of Service of Computer Networks (QoS) [M]. Beijing: Tsinghua University Press, 2004. (in Chinese)
- [11] F P Kelly. Notes on effective bandwidths[J]. ser. Royal Statistical Society Lecture Notes. Oxford University, 1996, (4): 141–168.
- [12] J Liebeherr. Post Internet QoS research[A]. IWQoS Panel Talk[C]. Montreal, Canada, 2004.
- [13] 王重钢, 隆克平, 等. 分组交换网络中队列调度算法的研究及其展望[J]. 电子学报, 2001, 29(4): 553–559.
Wang C, Long K. The study and perspective of queue scheduling algorithms in packet switching networks[J]. Acta Electronics Sinica, 2001, 29(4): 553–559. (in Chinese)
- [14] D Tse. Multiuser diversity in wireless networks[Z/OL]. <http://www.eecs.berkeley.edu/~dtse/stanford416.ps>.
- [15] R Agrawal, V Subramanian. Optimality of certain channel aware scheduling policies[A]. Proc of the 40th Annual Allerton Conference on Communication, Control, and Computing[C]. Monticello, Illinois, USA, 2002. 1532–1541.
- [16] H Kushner, P Whiting. Asymptotic properties of proportional fair sharing algorithms[A]. Proc of 40th Annual Allerton Conference on Communication, Control, and Computing[C]. Monticello, Illinois, USA, 2002. 2051–2059.
- [17] A Stolyar. On the asymptotic optimality of the gradient scheduling algorithm for multiuser throughput allocation[J]. Operations Research, 2005, 53: 12–25.
- [18] M Andrews. Instability of the proportional fair scheduling algorithm for HDR[J]. IEEE Trans on Wireless Communications, 2004, 3(5): 1422–1426.
- [19] L Tassiulas, A Ephremides. Dynamic server allocation to parallel queues with randomly varying connectivity[J]. IEEE Trans Inf Theory, 1993, 30: 466–478.
- [20] M Andrews, et al. Scheduling in a queueing system with asynchronously varying service rate[J]. Probability in the Engineering and Informational Sciences, 2004, 18: 191–217.
- [21] S Shakkottai and A Stolyar. Scheduling for multiple flows sharing a time varying channel: The exponential rule[J]. American Mathematical Society, 2002, 207: 185–202.
- [22] A Eryilmaz, R Srikant, J R Perkins. Stable scheduling policies for fading wireless channels[J]. IEEE/ACM Trans on Networking, 2005, 13(2): 411–424.
- [23] M Andrews, L Zhang. Scheduling over nonstationary wireless channels with finite rate sets[J]. IEEE/ACM Trans on Networking, 2006, 14(5): 1067–1077.
- [24] J Roberts, L Massoulié. Bandwidth sharing and admission control for elastic traffic[J]. Telecommunication Systems, 2000, 15(1–2): 185–201.

- [25] G De Veciana, T J Lee, T Konstantopoulos. Stability and performance analysis of networks supporting services with rate control Could the Internet be unstable[A]. Proc of IEEE INFOCOM[C]. New York, USA: 1999. 802– 210.
- [26] T Bonald, L Massoulié. Impact of fairness on Internet performance[A]. Proc of ACM Sigmetrics[C]. Cambridge, MA, USA, 2001. 81– 92.
- [27] J Holtzman. CDMA forward link waterfilling power control [A]. Proc of the IEEE VTC[C]. Tokyo, Japan, 2000. 1663– 1667.
- [28] S C Borst. User-level performance of channel aware scheduling algorithms in wireless data networks[A]. Proc of INFOCOM[C]. San Francisco, USA, 2003. 321– 331.
- [29] S C Borst, et al. User level QoS and traffic engineering for 3G wireless 1xEV-DO systems[J]. Bell Labs Technical Journal, 2003, 8: 33– 47.
- [30] R Litjens. HSDPA flow level performance and the impact of terminal mobility[A]. Proc of IEEE WCNC[C]. New Orleans, LA, USA: 2005. 1657– 1663.
- [31] S C Borst, M Jonckheere. Flow-level stability of channel aware scheduling algorithms[A]. Proc. of WiOpt[C]. Boston, MA, USA, 2006. 272– 277.
- [32] D Wu, R Negi. Effective capacity: A wireless link model for support of quality of service[J]. IEEE Trans Wireless Commun, 2003, 2(4): 630– 643.
- [33] M M Krunz, J G Kim. Fluid analysis of delay and packet discard performance for QoS support in wireless networks[J]. IEEE J Sel Areas Commun, 2001, 19(2): 384– 395.
- [34] M Hassan, M M Krunz, I Matta. Markov based channel characterization for tractable performance analysis in wireless packet networks[J]. IEEE Trans Wireless Commun, 2004, 3(3): 821– 831.
- [35] C Li, H Che, S Li. A wireless channel capacity model for quality of service[J]. IEEE Trans Wireless Commun, 2007, 6(1): 356– 366.
- [36] D Wu, R Negi. Utilizing multiuser diversity for efficient support of quality of service over a fading channel[J]. IEEE Trans on Vehicular Technology, 2005, 54(3): 1198– 1206.
- [37] J Y Le Boudec, P Thiran. Network Calculus: A Theory of Deterministic Queueing Systems for the Internet[M]. Heidelberg, Germany: Springer Verlag, 2001.
- [38] Y Jiang. A basic stochastic network calculus[A]. ACM SIGCOMM[C]. Pisa, Italy: 2006.
- [39] M Fidler. An end-to-end probabilistic network calculus with moment generating functions[A]. Proc of IWQoS[C]. New Haven, CT, USA: 2006. 261– 270.
- [40] F Aghahabparast, V C M Leung. Link-layer modeling of a wireless channel using stochastic network calculus[A]. Proc. Canadian Conference on Electrical and Computer Engineering [C]. Ontario, Canada, 2004. 1923– 1926.
- [41] M Fidler. A network calculus approach to probabilistic quality of service analysis of fading channels[A]. Proc. of IEEE GLOBECOM[C]. San Francisco, CA, USA, 2006. 1– 6.
- [42] Y Jiang, P J Emstad. Analysis of stochastic service guarantees in communication networks: A server model[A]. Proc of IWQoS[C]. University of Passau, Germany, 2005. 233– 245.
- [43] K Sabnani. QoS enablers for next generation networks[A]. IWQoS Keynote Address[C]. New Haven, CT, USA, 2006.
- [44] W Ajib, D Haccoun. An overview of scheduling algorithms in MIMO based fourth generation wireless systems [J]. IEEE Network, 2005, 19(5): 43– 48.
- [45] C Antori Haro, et al. Cross layer scheduling for multi-user MIMO systems[J]. IEEE Communications Magazine, 2006, 44(9): 39– 45.
- [46] G Song, Y Li. Utility based resource allocation and scheduling in OFDM-based wireless broadband networks[J]. IEEE Communications Magazine, 2005, 43(12): 127– 134.
- [47] X Lin, N B Shroff, R Srikant. A tutorial on cross layer optimization in wireless networks[J]. IEEE J Sel Areas Commun, 2006, 24(8): 1452– 1463.
- [48] X Lin, N B Shroff. The impact of imperfect scheduling on cross layer rate control in wireless networks[J]. IEEE/ACM Trans on Networking, 2006, 14(2): 302– 315.
- [49] N Hegde, A Proutiere. Packet and flow level performance of wireless multihop data networks[A]. Proc of IEEE GLOBECOM[C]. San Francisco, CA, USA, 2006. 1– 5.

作者简介:



雷 蕾 女, 1980 年生于河北唐山, 博士后. 主要研究方向为网络性能评价、无线网络、互联网体系结构. Email: leil@buptnet.edu.cn



林 闯 男, 1948 年生于辽宁沈阳, 教授, 博士生导师. 主要研究领域为计算机网络和系统性能模型及评价. E-mail: chlin@tsinghua.edu.cn