

文章编号: 1001-0920(2007)02-0121-06

网格系统的服务质量保障与控制综述

梁 泉¹, 杨 扬¹, 梁开健^{1,2}

(1. 北京科技大学 信息工程学院, 北京 100083; 2. 湖南工程学院 计算机系, 湖南 湘潭 411101)

摘 要: 针对类型多样的异构资源、动态协作的复杂网格环境, 需要有效的服务质量 (QoS) 控制机制保障网格应用的需求, 网格 QoS 需引入新的概念和机制, 实现 QoS 需求的描述与映射、动态资源聚合与共享及任务间协作, 灵活有效地进行网格应用流的聚合、区分和控制. 在此基础上, 介绍了网格 QoS 目前的研究情况, 分析了网格 QoS 控制的特点、难点及其基本需求. 结合实际研究, 对网格 QoS 控制的关键技术——网格 QoS 管理、控制策略、资源预留、自适应控制、QoS 协商进行了较深入的剖析, 同时讨论了 QoS 评价的问题. 最后总结了网格 QoS 的基本研究目标和未来的研究方向.

关键词: 网格服务质量; 网格 QoS 管理; 资源预留; 自适应控制; 网格 QoS 评价

中图分类号: TP393

文献标识码: A

Guarantee and control of quality of service on grid system : A survey

LIANG Quan¹, YANG Yang¹, LIANG Kai-jian^{1,2}

(1. School of Information Engineering, University of Science and Technology of Beijing, Beijing 100083, China;

2. Department of Computer Science, College of Hu 'nan Engineering, Xiangtan 411101, China. Correspondent:

LIANG Quan, E-mail: liangquan_lq@163.com)

Abstract: Under dynamic and collaborative environment with varieties of heterogeneous resources, effective control mechanisms for quality of Service are needed to guarantee requirements of the grid applicatons. It is essential to introduce new concepts and mechanisms for grid QoS, and to realize QoS description, QoS mapping, dynamic resource aggregation and collaboration among various tasks. It is also necessary to aggregate and distinguish application flows in the grid. A summary of research on grid QoS is introduced and the characteristics of grid QoS with its basic requirements are discussed. Referencing to actual researches, the key technologies, such as grid QoS management, control policies, resource reservation, adaption control and QoS negotiation, are analyzed. Grid QoS evaluation is also discussed. Finally, the research directions and basic contents of grid QoS are presented.

Key words: Grid QoS; Grid QoS management; Resource resevation; Adaption control; Grid QoS evaluation

1 引 言

网格技术是当前国内外研究的一个热点领域^[1]. 基于网格技术建立一个区域或全球合作的虚拟科研和应用环境, 可以改变和提高目前科学研究工作和实际应用的方式与效率^[2,3]. “提供非凡的服务质量 (QoS)”^[4] 是网格技术的关键内容. 实际的网格应用需要使用异构的、类型多样的资源, 任务间需要协作, 这样使得资源的调度分配与执行变得复杂, 因此引入 QoS 机制. 网格 QoS 要为应用提供端到端 QoS 保障, 目前的网格系统大多只能提供部分 QoS 功能.

网格 QoS 目前属于前沿性课题, 国内外的专项

论文不多. 本文致力于网格环境下资源管理与调度、QoS 控制的基础理论研究^[5-9], 对网格 QoS 的保障与控制进行综述和分析, 并为之提供有益思路.

2 网格 QoS 及其需求

2.1 QoS 属性

QoS 属性包括质和量两个方面: 质是指服务的可靠性和用户的满意度; 量^[10] 包括网络 QoS 属性、计算资源 QoS 属性^[11]、存储资源 QoS 属性、经济考量等, 还有其他特定需求的 QoS 参数. QoS 属性描述应用的 QoS 需求, 确定相应 QoS 参数.

2.2 网格 QoS 控制的分层

网格 QoS 涉及诸多方面, 目前网格实现实际上

收稿日期: 2005-11-11; 修回日期: 2006-01-10.

基金项目: 国家自然科学基金重点项目 (90412012).

作者简介: 梁泉 (1972 -) 男, 湖南洞口人, 博士生, 从事网格环境下的资源组织、管理与服务质量 (QoS) 控制等研究; 杨扬 (1958 -) 男, 河北隆化人, 教授, 博士生导师, 从事多媒体与网络通信、网格计算等研究.

建立在 Internet 上,应该算是中间件.低层需要网络 QoS 的保证,高层是高端的网格应用.所以网格 QoS 应从应用层、中间件层、网络层 3 方面来考虑^[12],使用分层发现的策略,确定 QoS 测量参数.

在应用层,网格用户参与协作和分布式的应用,访问和使用分布式资源,必须使用 QoS 管理方法协调与确保应用和服务之间的及时交互.应用层 QoS 有两个不同视角,其一是关心可以感知的响应时间,其二是安全问题,二者必须予以综合考虑.

中间件层负责创建描述 QoS 度量的执行环境以及随后的管理、控制和实施 QoS.目前很多中间件,包括 Globus 等,通常没有考虑 QoS.

网络层有两种 QoS 协议支持高端的应用^[13]: Int Serv 和 Diff Serv.但网格协作环境需要各种流的混合、变化的服务等级需求和相互依赖、多管理域和多类型资源的聚合等,需要引入 QoS 协议和较高层的体系结构,并建立自适应机制,提供端到端的解决方案.

此外,操作系统层对 QoS 控制会产生影响^[5],部分诸如请求、调度等处理需要在操作系统的内核中完成;上层的 QoS 描述也需要操作系统的正确理解.因此,网格 QoS 比网络 QoS 复杂得多,后者只是前者的一部分.

2.3 网格 QoS 的需求

网格 QoS 面临体系结构层和网络层的双重挑战.在体系结构层,需要引进新的概念和构建方式处理异构资源;在网络层,改善 QoS 体系结构,将高层的 QoS 需求正确地映射到网络层.

总之,网格 QoS 控制应满足如下需求^[14,15]:1) 支持不同特征混合流,提供数据传输保证;2) 实现形式统一的提前、立即预留和多资源联合分配;3) 有利于预留管理与使用的预留策略^[16];4) 保证用户期望的资源质量,并进行监控;5) 遵从网格系统的安全机制;6) 保障网格系统整体性能;7) 网格 QoS 机制应该能够适应动态变化.

3 目前网格 QoS 的研究概况

目前实现的一些网格项目对 QoS 控制可以分为 3 类:1) 无 QoS 保证,如欧洲 Data Grid^[17];2) 软 QoS 保证,如 Globus^[18] 和 Legion^[19] 等;3) 硬 QoS 保证,如 Darwin^[20] 等.软 QoS 支持能够为资源请求提供显示的 QoS 特征,但不能管制 SLA^[16],硬 QoS 可以管制 SLA.

网格 QoS 的目标是提供访问资源的保证,同时保持域间的安全等级.与多媒体和网络 QoS 不同的是,网格 QoS 需要集中的信息服务^[21,22]、保持关于资源可用性的最新信息、SLA^[17] 协商用户必须接收

而提供者必须提供的服务等级,使用资源属性作为 QoS 参数并得到监控.Sahai 等^[23] 提出了在商业网络背景下支持 QoS 的 SLA 管理实体,遵从 OGSA,但普遍适用性有待验证.SNAP^[24] 提出了一个通用的资源协商、管理模型,在分布式环境中描述复杂服务需求,具有普遍适用性.Keahey 等^[25] 提出了虚拟应用服务 VAS 体系结构,管理计算网格中的 QoS,使实时服务更易执行.虚拟资源管理器 VRM^[26] 能够在不同机群的众多调度器之间进行交互,充当协调器,聚合不同子系统间的 SLA 协商^[27].Darwin^[20] 能够基于 QoS,支持/管理对复杂网络服务的请求,创建“等级分组”,用户以任务图的形式提出请求,资源管理器根据 QoS 需求定位合适资源,运行任务.

在 QoS 管理结构上,出现了一些新的概念和机制,比较典型的有 GARA^[28] 和网络资源调度器 NRS^[29].NRS 在某种程度上类似 GARA,在提供区分服务 QoS 上也很有效果,但它只支持 cisco 路由器,不能明确地定义应用编程接口,如何使用是个问题.GQoSM^[12,30] 是针对面向服务网格而提出的,目前得到了广泛支持.GRAAP 的建议草案讨论将基于 OGSi 的协定模型^[21] 纳入支持 QoS 管理的体系结构中,并把重点放在修改 OGSi 上,使之遵从 Web Service,这更符合实际应用的要求.

4 网格 QoS 结构及其管理活动

只有把 QoS 控制和资源管理框架集成,才能最终解决网格 QoS 问题,管理框架应该支持资源动态发现和预留,实现异构资源、控制和管理的相互独立,也应有利于 QoS 控制策略的管理与实施^[7].

4.1 GARA 的 QoS 结构

GARA^[18,28] 框架赋予了编程人员指定端到端 QoS 需求的能力,提供提前预留,以通用的方式处理网络、计算、存储等各类资源,为资源定义了统一的描述机制,独立地进行创建、监控和管理,简化了异构资源的管理.GARA 也可以构建应用层联合预留和联合分配库,使应用程序既能根据应用的 QoS 需求,也能根据本地管理策略,动态地汇集、装配资源,同时提供应用程序接口和相应操作.GARA 实际上是一个分层结构,如图 1 所示.

尽管 GARA 被广泛接受,但面对当前的应用需求仍存在技术上的局限:1) 不兼容 OGSA,面向服务的应用不能直接使用 GARA 的服务;2) 不支持 SLA 概念,无法进行资源和 QoS 协商;3) QoS 监控和自适应是重要且成功的机制^[31],GARA 本身不具备这些功能.



图 1 GARA 体系结构

4.2 面向服务的 QoS 结构 G-QoSM

G-QoSM^[12,30] 遵从 O-GSA, 支持最新的 GGF 标准^[32], 提供面向服务的 QoS 管理模型, 有 3 个基本功能: 1) 基于 QoS 属性的网络服务发现; 2) 支持在中间件层和网络层协商 SLA, 实施 QoS; 3) 基于 SLA 提供对已分配资源的 QoS 管理, 如自适应。G-QoSM 框架包括 QoS 网络服务(QGS)、策略服务、预留管理等组件, 支持提前或立即预留和联合预留, 支持任意资源类型的融合。如图 2 所示, QGS 存在于每个域中, 捕获有 QoS 约束的服务请求, 协商 SLA, 获取最佳服务的信息和必要的策略验证请求进行预留通信, QoS 选择服务 UDDIe^[33] 提供服务信息。G-QoSM 设计不同接口控制异构资源^[34], 任何资源集成到 G-QoSM 都只需相应资源类型接口充当来往资源管理器的网关, 将资源连接到分配管理器。

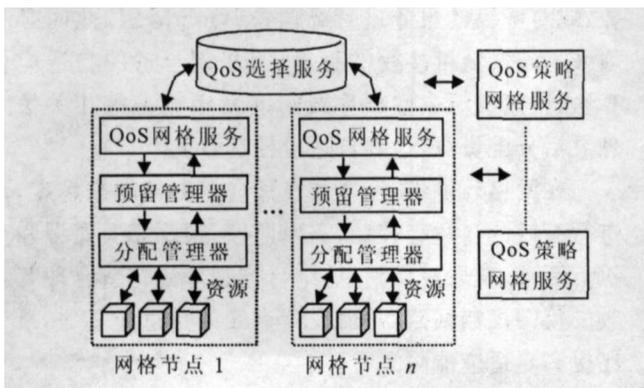


图 2 G-QoSM 体系结构

G-QoSM 的优势在于它顺应了面向服务网络的方向, 通过扩充 WSDL & UDDI^[33], 加进 QoS 属性类, 使服务发现可以基于 QoS 属性来进行, 确保用户的 QoS 需求。另外, 利用 UDDI API 还可以开发基于 QoS 属性搜索的机制。但 G-QoSM 在很大程度上只是一个概念, 还有许多关键技术需要完成。

4.3 网络的 QoS 管理

在网格环境下, QoS 管理活动包括资源选择、分配、资源释放、自适应、监控、重协商等。G-QoSM

将 QoS 管理作为一次 QoS 会话^[30,35], 有 3 个阶段的活动: 建立阶段包括 QoS 描述、映射、协商及资源预留、记帐; 活动阶段包括 QoS 监控、自适应、记帐、协商和重协商; 终止阶段则负责结束 QoS 会话, 预留终止、协定违规、服务完成和资源释放都可导致 QoS 会话终止。

请求区分、接纳控制、分配策略、负载均衡和经济考量等系统策略性机制是网格 QoS 综合控制的组成部分。本文提出基于请求内容和优先级区分的控制思想^[6], 考虑负载均衡和 QoS 控制结合的综合策略和方案^[7,8]; 对请求队列的处理, 基于 Web QoS 提出多服务器多队列系统模型^[9]; 对资源的质量状况, 研究了整体质量特性的衡量方法, 并在此基础上引入网格的经济模型, 进行资源质量的博弈, 以获得网格资源市场的公平和稳定, 有利于实际网格应用的 QoS 控制^[36], 这些研究保证了 QoS 控制目标的一致性和协调性。

针对网格系统的特点, 网格 QoS 管理应具备如下基本功能: 1) 描述 QoS 需求; 2) 根据资源能力映射 QoS 需求; 3) 进行 QoS 协商; 4) 能够与客户建立相应的 SLA; 5) 预留和分配资源; 6) 监控 QoS 参数; 7) 进行自适应控制; 8) 处理 QoS 终止。

5 网格系统 QoS 保障与控制的关键问题

5.1 网格资源预留

预留和自适应是 QoS 控制的两种基本技术。预留是对客户服务质量的承诺, 包括提前、立即和“按需”预留。很多应用研究过预留^[16,37], 大多基于 RSVP 协议实现。

GARA 引入了通用的资源对象和联合预留代理的新实体^[28], 分为预留和分配两个阶段, 能够进行提前预留。联合预留代理负责发现满足 QoS 需求的资源集, 预留并返回预留句柄集, 随后传递给联合分配代理进行分配。除此之外, 标准信息服务动态收集资源信息, 满足应用 QoS 的需求。由于采用 GRAM 的资源管理, 能支持异构资源的联合分配。G-QoSM 的办法^[12] 是先将资源临时预留, 由 QGS 生成 SLA, 包含所需的资源说明参数, 若客户同意, 则形成委托契约, 由 QGS 传递 SLA 参数, 发送给分配管理器, 并由它控制资源管理器分配/取消资源。预留管理器中定义了由 API 实现的一系列操作, 可以提供立即或提前预留以及处理预留。

进行资源预留会出现不少问题, 如预留时间重叠。为避免这种情况, 应拒绝无确定时间段的预留, 但需要预测服务运行的时间。也有人建议将立即和提前预留的资源加以区分, 这样预留的资源可能变为不可用。预留了资源而不用, 会导致资源浪费或其

他请求被拒绝,需要进行预留容错,而且预留会对网络系统调度产生影响. 研究人员发现,将超级计算调度系统引入预留后^[38],提交给队列的应用程序平均等待时间增加,请求预留时间与实际预留时间的平均偏移量也随不同的预留技术而有所不同,因此需要优化预留技术,减小预留对系统的影响. 预留必须考虑接纳控制来处理联合预留. 目前网络资源预留存在较多困难,如:处理的不止是单个资源,而是复杂资源集合;资源类型广泛、异构,且分属不同管理域,遵循不同的安全和使用策略;大部分资源不支持预留等.

5.2 网络中的 QoS 协商

QoS 协商是保障服务质量必不可少的一环, QoS 控制通过 QoS 协商得到实施,为此要建立 QoS 协商的相应模型和协议.

QoS 协商的目标是在用户和服务提供者之间关于预留、调度、服务参数等达成某种协定. 协商包含服务协商和 QoS 协商,两者结合可增进系统的可用性和灵活性. 可用性是指关于接纳服务的数目,而灵活性是对不同请求的自适应性. QoS 协商需要在用户的 QoS 约束和服务提供之间进行匹配,保证请求不超过资源的能力. 另外,若要遵从已达成的协定,要么减少请求量,要么提供额外资源量,两者都会导致重新协商. 若重新协商提高了 QoS 等级,则还需运行接纳控制校验,看作新的协商. QoS 协商模型需要有服务协商阶段、协商资源特征或服务等级.

协商协议指定协商过程中各个主体间所交换信息的语法、语义,目标是使各个主体包括用户、QoS 代理、服务提供者等能互相达成协定. 各个主体间交换 XML 消息进行协商,协调资源的发现、预留、获取和监控等,并形成约定,保障服务质量. 图 3 以 G-QoSM 为原型,演示了这种 QoS 协商过程. QGS 是中心主体,负责在用户和服务提供者之间协调协商过程. 协商模型中的 4 种操作分别是“查询”、“预留”、“更新”和“取消”,定义了协商的协议.

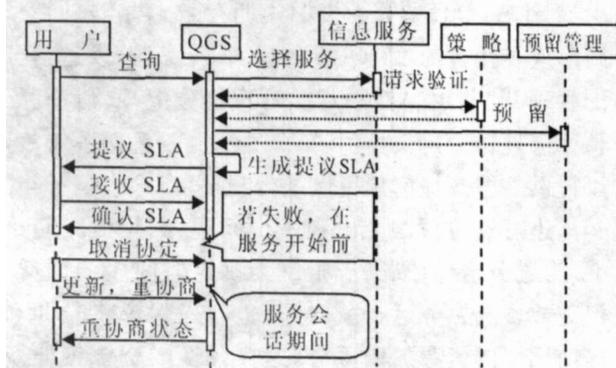


图3 QoS协商过程示意

QoS 协商还有许多问题需要研究和解决. 对于 SLA 违规问题,需要建立相应的机制;内容多样、网络的经济模型还需进行价格折中、交易、服务结算等协商.

5.3 网络 QoS 自适应

QoS 自适应一般基于反馈原理做出反应,在通信网络、动态多媒体应用和实时系统等领域均有广泛应用^[7,39,40]. 但固定量的预留常会造成资源浪费,资源管理策略随分配情况变化,应用的性能不能平滑降级,拥塞发生后容易造成自适应技术失败. 网络应用常常是动态流、多个流和多种流的混合,需要更为稳健和有效的自适应机制来保障.

GARA 为此引入 3 个自适应控制组件^[18]:制动器实时掌握资源特性,这是自适应的第 1 个先决条件;传感器监控资源分配或应用行为,确定系统组件的状态和检测状态的改变,这是自适应的第 2 个需求;决策过程根据传感器返回的信息,调用制动器作出反应. 例如,若传感器检测到路由器发生丢包,则发出丢包非零信号,它与应用关联的决策过程决定是否要减少发送包的速率,或者向资源管理器发出请求,创建或增加相应流的带宽,由制动器实施控制. 面向服务的网络中,有以下几种自适应情况:1) 新到服务请求. 在遵从当前的 SLA 情况下,可释放部分资源给新服务,或对某些服务降级;2) 服务终止. 可增加其他服务的资源,提高资源的使用利润,也可升级其他服务的 QoS 等级,使之得到最好服务,或对降级的服务进行补偿;3) QoS 降级. 此时必须恢复到最低可接受的程度,需采用合适的自适应算法,根据资源失败概率或阻塞发生概率等相关信息灵活分配资源,达到自适应控制目的^[35].

在网络环境中,反馈和自适应将变得越来越难,进行有效的自适应控制,需要监控、跟踪相关资源和执行活动,获取有用信息,结合控制理论构造合理有效的反馈控制回路,从而设计有效的自适应算法,最终提高自适应能力.

5.4 QoS 效用模型和 QoS 评价

效用在经济领域是指消费者消费某种商品的满意程度,满意程度高则效用大,反之,则效用小. 网络 QoS 引入效用的概念和模型很重要,有利于达到目标的优化. 如定义 $(SLA_1, SLA_2, \dots, SLA_n)$, QoS 管理可分为验证 SLA 的有效性(批准或拒绝)和优化服务的使用费用. 若要付的总费用为 $i\text{cost}(q_i(t), r_i)$, 其中 $\text{cost}(q_i(t), r_i)$ 表示使用资源 r_i 所对应的质量等级 $q_i(t)$ 需付出的费用. 分别定义效用函数 $\max_i \text{cost}(x, r_i)$ 和 $\min_i \text{cost}(q(t), r_i)$ 给提供者和请求者,把需求转化为数学描述和优化问题.

网络 QoS 评价,质的方面是对主观满意程度和服务可用性、可靠性的认识,量的方面是取得 QoS 参数可接收值或可接收值范围.对 QoS 性能评价应反映多个 QoS 目标^[10],并能体现其内在联系.关键在于找到有效的评价模型,通过对模型的研究提出网络 QoS 综合性能评价的标准.确定评价标准是另一个关键,不同的评价标准会导致不同的评价结果,从而导致 QoS 控制方案的不同^[7].目前对网络 QoS 评价的研究较少.网络 QoS 分层对比如表 1 所示,总结了网络 QoS 的层次目标和相应参数.

表 1 网络 QoS 分层对比

内容目录	QoS 参数	参数说明			
应用层 QoS	参与资源、服务的操作与协调,进行 QoS 管理的实时约束与交互、测量、记录参数,自适应、QoS 协商、接纳控制等	服务响应时间 服务执行时间 服务成本 服务可达性/可用性	请求发起到完成的时间间隔 开始执行到收到结果的时间间隔 获得或提供服务所需的经济费用 可用请求服务的成功概率来衡量		
	中间件层 QoS	提供 QoS 描述的执行环境,基于 QoS 的调度环境和资源管理环境,支持 QoS 参数映射和协商机制	资源访问速率 存储能力、带宽 资源可达(用)性 CPU 比率或数目	单位时间内访问资源数据量 存储量及 I/O 访问速率 用访问资源成功概率来衡量 CPU 占用百分比或 CPU 个数	
		网络层 QoS	提供端到端网络流保障,支持网络应用流的聚合、区分、拥塞机制和方法	传输延迟 延迟 jitter 丢包率 吞吐量	传输请求到完成的时间间隔 同一路由包延迟的变化量 单位时间内丢包的比率 单位时间内传输最大字节数

6 结 语

本文对目前网络 QoS 研究作了比较详尽的综述,并对其中的关键问题,如 QoS 结构、资源预留、QoS 协商、自适应控制和 QoS 评价等进行了剖析.

为实现网络系统的服务质量保障与控制,网络 QoS 未来应该研究的内容包括:1)可应用、融合多类型资源的 QoS 管理结构;2)网络 QoS 需求描述和映射机制;3)预留及联合预留与分配;4)网络 QoS 自适应方法和算法;5)网络 QoS 综合控制机制:协商与重协商、接纳控制、负载平衡、监控机制及扩展机制等;6)网络应用网络流的聚合、区分和控制;7)基于 QoS 的网络服务描述,未来要解决服务组合、服务 workflow 等涉及的 QoS 控制问题;8)网络 QoS 评价.这些问题的解决,最终将有助于 QoS 保障与控制的实现.

参考文献(References)

- [1] Ian Foster, Carl Kesselman. The grid2: Blueprint for a new computing infrastructure [M]. San Francisco: Morgan Kaufmann Publishers, 2004.
- [2] Fran Berman. Grid computing: Making the global infrastructure a reality[M]. San Diego: John Wiley and Sons Ltd, 2003.
- [3] Foster I, Kesselman C, Nick J, et al. The anatomy of the grid: Enabling scalable virtual organizations[J]. Int J Supercomputer Applications, 2001, 15(3): 78-91.
- [4] Foster I. What is the grid? A three point checklist[EB/OL]. <http://www-fp.mcs.anl.gov/~Foster>, 2005-07-20.
- [5] 杨扬,单志广,林闯,等. Web QoS 控制研究综述[J]. 计算机学报, 2004, 27(2): 145-156. (Yang Y, Shan Z G, Lin C, et al. Control of Web QoS: A survey[J]. J of Computer, 2004, 27(2):145-156.
- [6] Yang Y, Shan Z G, Lin C. Modeling and performance analysis of QoS-aware load balancing of Web server clusters [J]. Computer Networks, Elsevier Science, 2002, 40(2): 235-256.
- [7] 单志广. Web 服务质量(QoS)控制的策略、模型及性能分析[D]. 北京:北京科技大学, 2002. (Shan Z G. Policy, model and performance — Analysis for Web QoS [D]. Beijing: University of Science and Technology of Beijing, 2002.)
- [8] 杨扬,单志广,林闯,等. Web 请求分配和选择的综合方案与性能分析[J]. 软件学报, 2001, 12(3): 355-366. (Yang Y, Shan Z G, Lin C, et al. The composite method with analysis for Web request [J]. J of Software, 2001, 12(3): 355-366.)
- [9] Shan Z G, Lin C, Yang Y, et al. A multiserver multiqueue network: Modeling and performance analysis [J]. J of USTB, 2002, 9(5): 389-395.
- [10] 林闯,单志广,任丰原,等. 计算机网络的服务质量(QoS)[M]. 北京:清华大学出版社, 2004. (Lin C, Shan Z G, Ren F Y, et al. Quality of service (QoS) for computer network[M]. Beijing: Publisher of Tshinghua University, 2004.)
- [11] Bent J, Venkataramani V, LeRoy N, et al. Flexibility, manageability and performance in a grid storage appliance [C]. Proc of the 11th IEEE Symposium on High Performance Distributed Computing. Scotland, 2002.
- [12] Al A R, Amin K, Laszewski von G, et al. An OGSA-based quality of service framewok[C]. Proc of the 2nd Int Workshop on Grid and Cooperative Computing (GCC2003). Shanghai, 2003.
- [13] Deora V, Shao J, Gray W A, et al. A quality of service management framework based on user expectations

- [C]. 1st Int Conf on Service-oriented Computing. Trento, 2003.
- [14] Foster I, Markus Fidler, Alain Roy, et al. End-to-end quality of service for high-end applications [J]. *Computer Communications*, 2004, 27(14): 1375-1388.
- [15] Al A R. Analysis and provision of QoS for distributed grid applications [J]. Kluwer Academic Publishers, 2004, 25(4): 59-68.
- [16] Karsten M, Berier N, Wolf L, et al. A policy-based service specification for resource reservation in advance [C]. *Int Conf on Computer Communications (ICCC)*. Toronto, 1999.
- [17] Hoscheck W, Jaen Martinez J, Samar A, et al. Data management in an international data grid project [R]. 1st IEEE/ACM Int Workshop on Grid Computing, 2000.
- [18] Foster I, Roy R, Sander V. A quality of service architecture that combines resource reservation and application adaptation [C]. 8th Int Workshop on Quality of Service. Pittsburgh, 2000.
- [19] Chapin S, Karpovich J, Grimshaw A. The legion resource management system [C]. 5th Workshop on Job Scheduling Strategies for Parallel Processing. Nevada, 1999.
- [20] Chandra P, Fisher A, Kosak C, et al. Darwin: Customizable resource management for value-added network services [J]. *IEEE Network*, 2001, 15(1): 102-135.
- [21] Czajkowski K. Agreement-based grid service management (OGSF Agreement) [EB/OL]. <http://www.gridforum.org>, 2003.
- [22] Czajkowski K, Fitzgerald S, Foster I, et al. Grid information services for distributed resource sharing [C]. *Proc of the 10th IEEE Int Symposium on High-performance Distributed Computing (HPDC-10)*. San Francisco, 2001.
- [23] Sahai A, Graupner S, Machiraju V, et al. Specifying and monitoring guarantees in commercial grids through SLA [C]. *Proc of the 3rd IEEE/ACM CCGrid*. Berlin, 2003.
- [24] Czajkowski K, Ian Foster, Carl Kesselman, et al. SNAP: A protocol for negotiating service level agreements and coordinating resource management in distributed systems [C]. *The 8th Workshop on Job Scheduling Strategies for Parallel Processing*. Edinburgh Scotland: Springer-Verlag, 2002.
- [25] Keahey K, Motawi K. The taming of the grid: Virtual application service [J]. *Argonne National Laboratory Technical Memorandum*, 2003, 26(2): 267-285.
- [26] Burchard L, Hovestadt M, Kao O, et al. The virtual resource manager: An architecture for SLA-aware resource management [C]. *Proc of IEEE CCGrid*. Chicago, 2004.
- [27] Fawaz W, Daheb B, Audouin O, et al. Service level agreement and provisioning in optical networks [J]. *Communications Magazine*, 2004, 42(1): 36-43.
- [28] Foster I, Kesselman C, Lee C, et al. A distributed resource management architecture that supports advance reservation and co-allocation [C]. *Proc of the Int Workshop on QoS(IWQoS)*. Seattle, 1999.
- [29] Bhatti S. Network QoS for grid systems [J]. *Int J of High Performance Computing Applications, Special Issue on Grid Computing: Infrastructure and Applications*, 2003, 17(3): 178-200.
- [30] Al A R, Rana O. A model for quality of service provision in service oriented architectures [J]. *Int J of Grid and Utility Computing*, 2005, 1(3): 72-86.
- [31] Al A R, Hafid A, Rana O, et al. An approach for QoS adaptation in service-oriented grids [J]. *Concurrency and Computation: Practice and Experience J*, 2004, 16(5): 401-412.
- [32] The global grid forum Web page [EB/OL]. <http://www.gridforum.org>, 2005-07.
- [33] ShaikhAli A, Rana O, Al A R, et al. UDDIe: An extended registry for web services [C]. *Proc of Workshop on Service-oriented Computing: Models, Architectures and Applications SAINT2003*. Los Alamitos: IEEE CS Press, 2003.
- [34] Al A R, Sohail S, Rana O, et al. Network QoS provision for distributed grid applications [J]. *J of Simulation*, 2004, 5(5): 66-79.
- [35] Al A R, Hafid A, Rana O, et al. QoS adaptation in service-oriented grids [C]. *Proc of MGC*. Saint Paul, 2003.
- [36] 刘丽, 杨扬. 基于 QoS 的社区公共服务网格资源调度 [J]. *北京科技大学学报*, 2004, 26(5): 560-563. (Liu L, Yang Y. QoS-based grid resource scheduling of community services [J]. *J of USTB*, 2004, 26(5): 560-563.)
- [37] Kim K, Nahrstedt K. A resource broker model with integrated reservation scheme [C]. *IEEE Int Conf on Multimedia and Expo (ICME2000)*. Shanghai, 2000.
- [38] Smith W, Foster I, Taylor V. Scheduling with advanced reservations [C]. *Proc of the IPDPS Conf*. Cancun, 2000.
- [39] Xu D, Nahrstedt K, Wichadakul D. QoS-aware discovery of wide-area distributed services [C]. *Proc of the 1st IEEE/ACM CCGrid2001*. Berlin, 2001.
- [40] Steenkiste P. Adaptation models for network-aware distributed computations [C]. *Proc of CANPC (99)*. New York, 1999.