

网格环境下的服务质量(QoS)研究^{*}

梁 泉¹ 张宏正² 梁开健³ 杨 扬¹

(北京科技大学信息工程学院 北京 100083)¹ (山东中创软件工程有限公司 济南 250014)²

(湖南工程学院计算机系 湘潭 411101)³

摘 要 面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。该文分析了网格 QoS 控制的特点、难点和基本需求,介绍了网格 QoS 目前的研究情况,并从 QoS 管理结构和控制策略、资源预留、QoS 协商、自适应方法和 QoS 评价等几个方面进行阐述,最后分析和总结了网格 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,网格 QoS 评价

Research on Grid Quality of Service(QoS)

LIANG Quan¹ ZHANG Hong-Zheng² LIANG Kai-Jian³ YANG Yang¹

(School of Information Engineering, University of Science and Technology of Beijing, Beijing 100083)¹

(CVIC Software Engineering Co. Ltd., Jinan 250014)² (Computer Science Department of Hunan Engineering College, Xiangtan 411101)³

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 applications. 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's also necessary to aggregate and distinguish application flows in the grid. The characteristics of Grid QoS with its basic requirements are discussed and a summary of research on Grid QoS is presented in this paper. Analysis and comparative discussions are done on the aspects such as QoS Architectures with their Control Policies, Resource Reservation, QoS Negotiation, Adaptive Methods and QoS Evaluation. Finally, research directions and basic contents of Grid QoS are also discussed.

Keywords Grid Quality of Service(QoS), Grid QoS management, Grid QoS evaluation

1 前言

随着计算机科学的发展,网格技术已经成为当前国内外研究的一个热点和前沿领域^[1]。基于网格技术建立一个能够实现区域或全球合作的虚拟科研和实验环境,从而支持以大规模计算和数据处理为特征的科学研究活动,改变和提高目前科学研究工作的方式与效率^[1,3]。

“提供非凡的服务质量(QoS, Quality of Service)”是判断网格的三个准则之一^[4],但是实际的网格应用都需要使用异构的、类型多样的资源,特别是要对这些资源共享使用,各个任务间更多的是需要协作,使得资源的调度分配、任务执行变得复杂,而目前的网络基础设施又都是建立在“尽力服务”的基础之上,因而服务质量得不到保障,在多个服务协作使用资源时,性能和效率也容易造成降级,而且用户的需求也具有多样性。解决这些问题,需要引入 QoS 机制。网格 QoS 问题的焦点在于网格能够提供端到端 QoS 的程度,不只是提供网络 QoS,而目前的大多数网格实现都只能提供部分的 QoS 功能。

目前,网格 QoS 还属于一个前沿性的关键课题,国内外尤其是国内,还很少见到这方面的专门研究论文。杨扬教授领导的网格课题组,由国家自然科学基金重点项目资助,致力

于网格计算环境下资源管理与调度、服务质量(QoS)控制和性能保证的基础理论研究,我们在 QoS 控制方面有丰富的工作经验和成果^[5,6,20],在此背景下,本文针对还处于起步阶段的网格服务质量(QoS)研究进行了综述和分析比较,并归纳和总结了其中亟待研究与解决的问题,为网格 QoS 的研究提供了有益的思路。

2 网格服务质量及其需求

2.1 服务质量(QoS)属性

QoS 属性包括两个方面:质和量。质的方面是指,服务的可靠性和用户的满意度。量的方面比较广泛^[7],网络 QoS 属性有:延迟、延迟 jitter(同一路由包延迟的变化量)、丢包率,吞吐量;计算 QoS 指定计算资源 CPU 如何使用、分配,可以把占用 CPU 的比例或数目作为 QoS 属性;存储资源 QoS 属性有访问速率、存储能力等;经济上使用成本或费用(Cost),也是一个重要的 QoS 属性;还有其他特定需求的可能 QoS 参数,等等。必须描述应用的 QoS 需求,确定网格服务质量的 QoS 参数。

2.2 网格应用需求

高端网格应用的 QoS 需求^[8]要充分考虑:(1)异构流:涉

^{*} 国家自然科学基金重大研究计划重点项目(No. 90412012)。梁 泉 博士生,研究方向为网格资源组织、管理与质量(QoS)控制;梁开健 博士,副教授,研究方向为网格计算、数据挖掘;杨 扬 归国博士,教授,博士生导师,研究方向为多媒体与网络通信、网格计算。

及到多种流的变化特性,如:带宽、延迟等;(2)高带宽流:流的特点决定了需要两个层面上的考虑,QoS 机制允许流的共存,而策略层需要解决多个域资源的协同管理;(3)端到端 QoS:对网络及其他资源进行协同管理;(4)应用层的控制:能发现可用资源、监视取得的服务、修改 QoS 请求、动态调整应用的行为;(5)预留:设计与预留有关的调度器,提高整体性能。

保证网格应用的服务质量,面临网络层和体系结构层双重挑战,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,改善网络层的 QoS 体系结构,能把高层的 QoS 需求正确地映射到网络层。目前都还没有得到满意的解决。

2.3 网络 QoS 研究的分层观点

网络 QoS 可以从三个层次上来考虑^[17]。在应用层,网络用户参与协作和分布式应用,需要访问和使用分布式资源,必须满足相应的 QoS 需求,使用 QoS 管理方法协调与确保应用和服务之间的及时交互。例如,面向服务的网格应用层可能进行下列操作:1)发送控制信息到远程服务,2)向远程服务传送数据,3)执行远程服务,4)返回数据结果。应用层 QoS 管理,就要干预到上述的每一个操作当中,测量和记录关联操作的相应 QoS 参数,并根据具体情况灵活应变。中间件如 Globus^[9]等,提供描述 QoS 度量的服务执行环境,及随后的管理和控制,实施 QoS。在网络层有两种 QoS 协议支持高端的应用,IntServ 和 DiffServ,在网格环境下的网络层 QoS 更加复杂,需要实现不同管理域、不同资源的聚合。网络资源的聚合,过去都是建立在 Internet 的“尽力服务”基础上,现在不同,需要软实时约束和更加严格的传输保障。如何管理这些聚合资源的 QoS 需求,如何监控和测量应用的网络流,并建立自适应机制,是需要解决的问题。操作系统层对网络 QoS 控制也会产生影响^[2],很多诸如请求、调度等处理,部分需要在操作系统的内核中完成,如果这个层次的处理不精确,会带来负面效应。

综上所述,从分层的观点看,网络 QoS 和网络 QoS 是不同的两个问题,前者要复杂得多,网络 QoS 只是网络 QoS 中的一个部分。

2.4 网络服务质量(QoS)的需求

总结来说,网络 QoS 应满足以下需求^[8]:

- (1)支持不同特征混合流,提供可靠传输保证;
- (2)统一的资源预留和分配,实现联合分配;
- (3)预留策略,提高预留的灵活性;
- (4)协定协议:保证用户的提前预留状态及期望的资源质量,形成协定并得到监控;
- (5)安全性,遵从网格系统的安全机制;
- (6)精简性,不影响网络系统的整体性能;
- (7)可扩展性,网络 QoS 机制应能适应动态变化。

3 网络 QoS 研究的国内外现状

目前实现的一些网格项目,对 QoS 控制可以分为三类:无 QoS 保证、软 QoS 保证和硬 QoS 保证。能够为资源请求提供显式的 QoS 特征但不能通过管制手段保证服务等级协定 SLA(Service Level Agreement),这是软 QoS 支持;硬 QoS 支持是指网格的所有节点都可以管制由资源管理系统保证的 SLA。例如,Globus^[9]、Legion 等提供的是软 QoS 保证,Darwin^[10]等提供的是硬 QoS 保证,Condor、欧洲 DataGrid 等尚无 QoS 保证机制。

与多媒体和网络 QoS 不一样的是,网络 QoS 需要集中的信息服务^[11],保持关于资源可用性的最新信息。SLA 协商用户必须接收提供者必须提供的服务等级,集中了用户需求和资源拥有者的能力。Sahai 等提出了在商业网络的背景下支持 QoS 的 SLA 管理实体^[12],这是个不错的设想,但仍处于很初步的阶段,普遍适用性有待验证。服务协商和获取协议(SNAP:Service Negotiation and Acquisition Protocol)^[13]提出了一个通用的资源协商、管理模型,在分布式系统环境中描述复杂服务需求,具有普遍适用性,切合面向服务的架构,满足了管理需求的扩展。Keahey 等提出了虚拟应用服务(VAS:Virtual Application Service)的体系结构^[14],管理计算网格中的 QoS,VAS 是扩展了的网格服务,有助于协商 QoS 等级和服务需求的附加接口。虚拟资源管理器(VRM:Virtual Resource Manager)可以进行 SLA 管理^[15],VRM 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。

Globus 项目最先设想了一种 QoS 体系结构 Qualis^[9],它把应用层 QoS 需求映射成中间件、操作系统、网络层的 QoS,分配和调度各类型资源,处理 QoS 违规。Darwin 项目是面向服务网格的管理系统^[10],能够基于 QoS,支持/管理对复杂网络服务的请求。GARA^[16]提供了 QoS 支持的资源配置、发现、选择和分配,支持接纳控制和自适应。NRS(NRS:Network Resource Scheduler)采用 P2P 模型,某种程度上有些类似 GARA,在提供区分服务 QoS 上也很有效果,但没有清楚地定义应用编程接口,不便使用。网络 QoS 最基本和最有效的技术仍然是进行资源预留。GARA 是支持资源预留的;面向服务网格的 QoS 管理框架 G-QoSM^[17],也充分考虑了资源的预留;全球网格论坛 GGF 的 GRAAP 工作组提出了一份“state of art”文档,讨论了预留问题。

国内也开始重视网络 QoS,许多网格平台开始加入对 QoS 的支持,更多的研究也正在进行。2003 年,GRAAP 推出一份建议草案,讨论修改 OGSi,结合 Web Service,因此把网络 QoS 的研究集中在面向服务的框架更符合发展潮流。

4 网络 QoS 管理框架

只有把 QoS 控制和资源管理框架集成,才能最终解决 QoS 问题,管理框架应该支持资源动态发现、预留,实现异构资源、控制和管理的相互独立,管理框架也应有利于 QoS 控制策略的管理与实施^[6]。

4.1 GARA 的 QoS 框架结构

在网格中支持 QoS,最为广泛接收的框架是 GARA^[9,16],它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、存储等各类型资源。GARA 把预留和计算元素如进程、网络流、内存块都看成实体,可以独立地进行创建、监控和管理。为资源定义了统一的描述机制,简化了异构资源的管理。GARA 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 实际上是分层的结构,如图 1 所示,高层是 GARA 外部接口组件,负责鉴定和分派接收的请求,提供远程进程注册,传播、发布资源信息;低层的本地资源管理器(LRAM),负责提供基本对象和预留服务。引入时间槽表及其管理器,跟踪当前的资源分配和预留,保证要求的资源不超过限制。GARA 目前实现的预留资源类型有

网络资源、计算资源、存储资源,对底层资源的控制管理主要还是利用 GRAM。

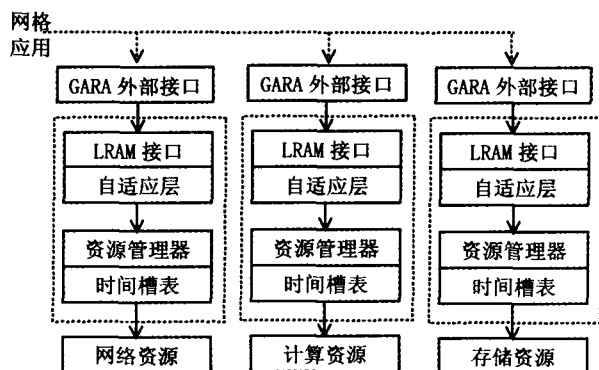


图1 GARA 体系结构

尽管 GARA 被广泛接受,但面对当前的应用需求却存在局限:(1)GARA 不兼容 OGSA,面向服务的应用不能直接使用 GARA 的服务;(2)GARA 不支持 SLA 的概念;(3)QoS 监控和自适应,是一种最重要且成功的机制^[18],GARA 本身不具备这些功能。

4.2 面向服务的 QoS 管理框架 G-QoS M

R. Al-Ali 等提出了 G-QoS M 框架^[17],G-QoS M 遵从 OGSA,提供面向服务的 QoS 管理模型,有三个基本管理功

能:(1)基于 QoS 属性的网格服务发现;(2)支持在中间件层和网络层提供 QoS,建立 SLA 增强 QoS;(3)在预先协商的 SLA 基础上,提供对已分配资源的 QoS 管理,如:自适应。框架支持最新的 GGF 标准,也与 OGSA 的最新规范匹配。G-QoS M 框架包括几个主要组件:QoS 网格服务(QGS),策略服务,通用资源预管理器,等等。支持提前或立即预留及联合预留;支持任意资源类型的融合;运行时,底层的资源特性使用面向对象技术,增强了可扩展性和灵活性。如图 2 所示,QGS 是此体系结构的焦点,存在于每个域中。QGS 与应用程序交互,捕获有 QoS 约束的服务请求,协商 SLA;与 QoS 选择服务交互,获取最佳服务的信息,通常,选择服务会返回一个匹配服务的列表,需要 QGS 从中选择一个最佳的,UD-DIe^[19]被认为是这样的 QoS 选择服务;与预留管理器交互,选择了一个网格服务以后,支持预留的功能需求被抽象和形式化为资源说明,提交到预留管理器进行预留,如果预留成功,返回一个预留“句柄”,以后可以用来申明或使用预留;与策略网格服务交互,可以让 QGS 能捕获到必要的策略以验证服务请求,例如,去发现服务在资源效用上有没有什么限制,或者去发现请求服务的类别。QoS 分配管理器的主要作用是和资源管理器交互,进行资源分配或取消分配,查询资源状态。对底层资源,设计不同接口,把不同类型资源灵活地集成 G-QoS M。

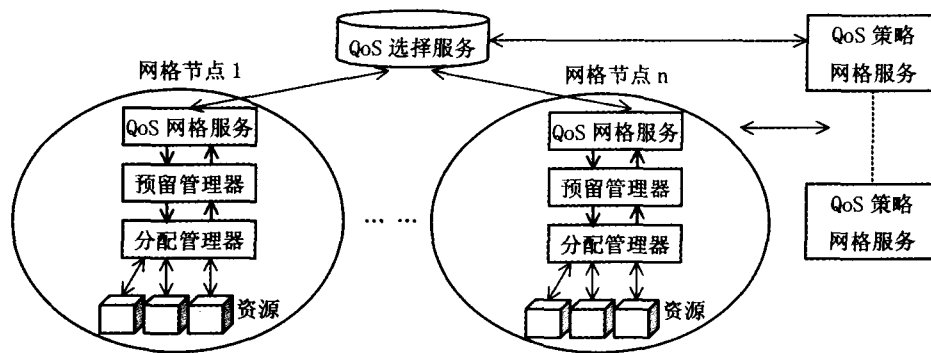


图2 G-QoS M 体系结构

G-QoS M 框架的优势在于它顺应了面向服务网络的方向,扩展了 WSDL&UDDI^[19],服务发现基于 QoS 属性来进行,从而提供 QoS 保证;还可以利用 UDDI APIs 开发基于 QoS 属性搜索的机制,为 QoS 保证提供了更广阔、灵活的空间。但目前 G-QoS M 很大程度上还只是一个概念,许多关键技术需要完成。

4.3 网络 QoS 管理

在网格环境下,QoS 管理活动有许多内容,如:资源选择、分配、资源释放、自适应、监控、重协商,等等。G-QoS M 框架把 QoS 管理分成一系列行为,组成一次 QoS 会话^[18],QoS 会话包含三个阶段,每个阶段都有特定的 QoS 管理活动。在建立阶段,包括 QoS 描述、QoS 映射、QoS 协商、资源预留和记帐;活动阶段的行为包括:资源分配、QoS 监控、QoS 协商和重协商、自适应和记帐;终止阶段,负责结束 QoS 会话,资源预留终止、协定违规、服务完成和资源释放,都可以导致 QoS 会话终止。

网络 QoS 管理还需要服务请求的区分、接纳控制、负载均衡等系统策略性机制,是 QoS 综合控制的组成部分。网络

应用的请求不能都一视同仁地看待,建立接纳控制机制,区分不同请求,并考虑请求分配策略及负载均衡,引入经济考量等等,都是网格 QoS 管理与控制的重要内容。当前综合考虑这些 QoS 控制理念的研究还很少,因而有广阔的开放研究空间。

网络 QoS 管理系统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)能与资源拥有者进行 QoS 协商;(4)能与客户建立相应的 SLA;(5)预留和分配资源;(6)监控 QoS 参数;(7)根据变化的资源质量特征进行自适应;(8)终止 QoS 会话。只有实现上述管理活动,QoS 控制和保障才算比较成熟,但目前还处在探索阶段。

5 网络环境下的资源预留

保证 QoS 的基本技术之一是资源预留。预留是对客户服务质量的承诺,预留包括提前、立即或“按需”预留。GARA 引入了通用的资源对象和联合预留代理的新实体^[16],如图 3 所示,并有 GRAM 提供的标准接口,以解决资源集的异构问题。

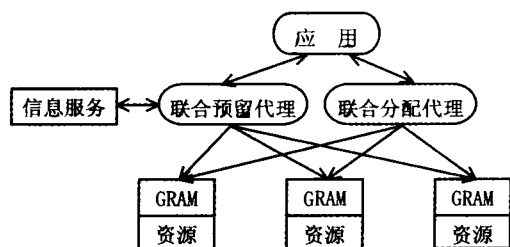


图3 GARA引入的联合预留代理

联合预留代理负责发现能满足 QoS 需求的资源集,但不是分配这些资源,而是预留并返回预留句柄集,随后传递给联合分配代理,进行对应资源集的分配。信息服务动态收集独立管理的资源,满足应用 QoS 的需求。由于采用 GRAM 的资源管理,因此能支持异构资源的联合分配,启动、监控和控制特定资源。G-QoS 的办法^[17]是先把资源临时预留,由 QoS 代理生成 SLA,包含所需的资源说明参数,若客户同意,则形成委托契约,否则资源释放预留状态,以便为其他请求所用;预留管理器定义了一系列操作,提供立即或提前预留及处理预留。但预留的时间段不确定,会导致资源的浪费,即使预留成功,但真正要用时,资源可能会失败,变为不可用;虽然预留了资源,但是却不用,导致其他需要资源的服务被拒绝。因此,应该提出“预留容错”概念,处理预留出错的问题。

引入预留,会对网络系统产生影响。研究人员发现,在超级计算调度系统引入预留后^[21],提交给队列的应用程序平均等待时间(MWT)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。若 10% 的应用程序使用了预留,将使 MWT 增加 9 个百分点,平均偏移量为 211 分钟;若 20% 使用预留,则 MWT 增加 10 个百分点,平均偏移量为 278 分钟。这表明,引入预留会产生系统开销,需要优化预留技术,减小对系统的影响。目前网络资源预留存在不少困难,预留和分配的不止是单个资源,而是潜在的复杂资源集合;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,等等。

6 网络 QoS 协商

QoS 协商是保证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间关于预留、调度、服务参数等达成某种协定,但不是任何时候都需要这种协商,特别是当客户请求能立即满足时。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,可用性是关于接纳服务的数目,而灵活性是对不同请求的自适应性。另外,如果协商失败或者资源的使用发生变化,都会导致重新协商,若重新协商提高了 QoS 等级,则还必须运行接纳控制校验,被看成新的协商。QoS 协商模型还应包括服务协商阶段,协商资源特征或服务等级,管理 SLA。协商协议指定协商过程中的各个主体间所交换信息的语法、语义,目标是使各个主体能互相达成协议,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。QoS 协商还有许多问题需要研究和解决,包括:SLA 违规问题、协商内容多样性、网络经济模型中交易与服务结算,等等。

7 网络 QoS 自适应

预留和自适应是 QoS 控制的两种有效技术。QoS 自适

应一般是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有广泛应用。在处理高带宽、动态流时,固定量的预留常会造成资源浪费,当资源管理策略随分配情况变化时,应用的性能不能平滑降级,拥塞发生后,自适应技术就可能失败。网络应用常常是多个流、多种流混合,需要更为稳健和有效的自适应机制来保障。GARA 引入了三个自适应控制组件^[9]:制动器(Actuator),在线控制资源特征,这是自适应的第一个先决条件;传感器(Sensor),监控资源分配或应用行为,确定系统组件的状态及检测状态的改变,这是自适应的第二个需求;决策过程(Decision procedure),实体根据传感器返回的信息,调用制动器做出反应。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,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,然后调用制动器进行控制。

网络服务的质量若发生改变,就应做出自适应反应,恢复 SLA 达成的 QoS,或者重新协商 QoS,对重大的 QoS 降级,可以终止服务。若有新到服务请求,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,并做出相应的补偿;若有服务执行完毕,则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;最常见的情况是 QoS 降级,此时需恢复其到最低可接受的程度,达到自适应控制目的^[18]。

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,从而设计有效的自适应算法和反馈机制。反馈控制方法涉及到各种控制理论,需要结合控制理论构造合理有效的反馈控制回路,最终提高自适应能力。但在多域环境中,反馈和自适应将变得越来越难,需要研究更多的实际应用以改善反馈和自适应技术。

8 研究的效用模型及 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是必要的,有利于达到优化的目标。譬如,执行某个服务要付出的总费用为 $\sum_i \text{cost}(q_i(t), r_i)$,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。从提供者角度定义效用函数 $\max \sum_i \text{cost}(x, r_i)$;从服务请求者角度定义 $\min \sum_i \text{cost}(q(t), r_i)$;这样就把质量要求转化成形式化的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。QoS 控制机制和方法的可行性、有效性需要证明,只是主观上的认可是不行的。对 QoS 性能评价应能反映多个 QoS 目标^[7],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,从而可导致 QoS 控制方案的不同。研究网络的形式化性能模型,进而研究网络 QoS 评价的模型和建模技术,确立科学有效的评价标准和方法,是有效的思路。目前,网络 QoS 性能评价模型及其标准还是少有人涉足的领域。

总结和展望 本文对目前网络 QoS 的研究做了比较详尽的阐述,网络 QoS 需解决网络应用对底层网络平台的需求,还需面对资源类型的复杂性、资源的联合调度和分配、多

管理域任务的协作、多层次的 QoS 需求,考虑网络的经济模型和网络安全(如鉴定、授权)等等。应该明确 QoS 需求和目

标,表 1 基于面向服务的网络总结了 QoS 各个层次的目标和相应参数。

表 1 网络 QoS 分层对比

QoS 层次	内容或目标	QoS 可能参数举例	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	发起服务请求到服务完成的时间间隔
		服务执行时间	服务开始执行到收到执行结果的时间间隔
		服务成本	获得或提供服务所需的经济费用
		服务可用度	可以利用请求服务的成功概率来衡量
中间件 QoS	提供 QoS 描述的执行环境、基于 QoS 的调度环境和基于 QoS 的资源管理环境,支持 QoS 参数映射和协商机制	资源访问速率	可以用单位时间内访问资源所获数据量大小来衡量
		存储能力、带宽	存储资源的存储量及 I/O 访问速率
		资源可达性/可用性	可以用访问资源时成功的概率来衡量
		CPU 分配比率或数目	计算资源被占用 CPU 的百分比或 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、拥塞等控制机制和方法	传输延迟	传输请求到确认完成的时间间隔
		延迟 jitter	同一路由包延迟的变化量
		丢包率	单位时间内传输包中丢失包的比率
		吞吐量	单位时间内网络传输的最大字节数
		带宽	单位占用传输通道的比特数

为增强网络性能、保障端到端网络应用,未来需要研究网络服务质量(QoS)的如下内容:

(1)可应用的 QoS 管理结构和集成、融合多类型资源的方法;

(2)网络 QoS 的需求描述和映射机制,把粗粒度的用户需求变成网络能理解的细粒度参数描述;

(3)资源的预留及联合预留与分配,考虑诸如负载均衡和资源协商问题;

(4)网络 QoS 自适应方法和算法,自适应和透明的 QoS 机制在网络环境中是必须的;

(5)服务质量(QoS)的管理控制机制,譬如,QoS 协商与重协商、服务接纳控制、QoS 的监控机制及 QoS 的扩展机制等;

(6)面向网络应用网络流的聚合、区分、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 管理容错,譬如,预留容错,保障 QoS 管理的可靠性;

(9)网络 QoS 评价和反馈机制,建立一套评价标准和相应模型;

通过对上述内容的研究和解决,相信能促进网络服务质量保证的最终实现,对网络技术的发展具有重要意义。

参 考 文 献

- Foster I, et al. The Grid2: Blueprint for a new Computing Infrastructure. Morgan Kaufmann Publishers, 2004
- 杨扬,等. Web QoS 控制研究综述. 计算机学报, 2004, 27(2)
- Foster I, Kesselman C, Nick J, Tuecke S. The Anatomy of the Grid: Enabling Scalable Virtual Organizations. International J. Supercomputer Applications, 2001, 15(3)
- Foster I. What is The Grid? A Three Point Checklist. <http://www-fp.mcs.anl.gov/~Foster>, July 20, 2005
- Yang Yang, et al. Modeling and Performance Analysis of QoS-aware Load Balancing of Web Server Clusters. Computer Networks, Elsevier Science, 2002, 40(2): 235~256
- 单志广. Web 服务质量(QoS)控制的策略、模型及性能分析: [博士学位论文]. 北京: 北京科技大学, 2002
- 林闯, 单志广, 任丰原. 计算机网络的服务质量(QoS)[M]. 清华大学出版社, 2004
- Foster I, Roy A, Sander V, Winkler L. End-to-End Quality of Service for High-End Applications. Computer Communications, 2003
- Foster I, Roy A, Sander V. A Quality of Service Architecture that combines resource reservation and application adaption. In: 8th Int'l Workshop on Quality of Service(IWQoS'00), June 2000
- Chandra P, Fisher A, Kosak C, et al. Darwin: Customizable Resource Management for Value-added Network Services. IEEE Network, 2001, 15(1)
- Czajkowski K, Fitzgerald S, Foster I, Kesselman C. Grid Information Services for Distributed Resource Sharing. In: Proc. of the Tenth IEEE International Symposium on High Performance Distributed Computing(HPDC-10), 2001
- Sahai A, Graupner S, Machiraju V, Moorsel A. Specifying and Monitoring Guarantees in Commercial Grids through SLA. In: Proc. of the 3th IEEE/ACM CCGrid2003, 2003
- Czajkowski K, Foster I, Kesselman C, et al. SNAP: A Protocol for Negotiating Service Level Agreements and Coordinating Resource Management in Distributed Systems. In: Proc. of the 8th Workshop on Job Scheduling Strategies for Parallel Processing, 2002
- Keahey K, Motawi K. The Taming of the Grid: Virtual Application Service. Argonne National Laboratory Technical Memorandum, 2003(262)
- Burchard L, et al. The Virtual Resource Manager: An Architecture for SLA-aware Resource Management. In: Proc. of IEEE CC-Grid 2004, Chicago, US, 2004
- Foster I, Kesselman C, Lee C, et al. A Distributed Resource Management Architecture that Supports Advance Reservation and CoAllocation. In: Proc. of the International Workshop on QoS(IWQoS), 1999. 27~36
- Al-Ali R, Rana O, et al. A Model for Quality of Service Provision in Service Oriented Architectures. International Journal of Grid and Utility Computing, 2004
- Al-Ali R, Hafid A, Rana O, Walker D. An Approach for QoS Adaptation in Service Oriented Grids. Concurrency and Computation: Practice and Experience Journal, 2004, 16(5): 401~412
- ShaikhAli A, Rana O, Al-Ali R, Walker D. UDDIe: An extended registry for web services. In: Proc. of Workshop on Service Oriented Computing: Models, Architectures and Applications SAINT2003, IEEE CS Press, Orlando FL, USA, 2003. 85~90
- 杨扬,等. Web 请求分配和选择的综合方案与性能分析. 软件学报, 2001, 12(3): 355~366
- Smith W, Foster I, Taylor V. Scheduling with Advanced Reservations. In: Proc. of the IPDPS Conf., May 2000