

QoS 约束的云经济资源管理模型研究

钱琼芬^{1,2} 李春林¹ 张小庆¹

(武汉理工大学计算机科学与技术学院 武汉 430063)¹ (空军雷达学院四系 武汉 430019)²

摘要 提供用户满意的、具有 QoS 约束的云计算应用是云计算面临的一大难题。提出了以商品市场为原型的云计算经济资源管理模型,其通过云用户与供应商的 SLA 协商,实现应用服务层 QoS 到资源设备层 QoS 的映射,最后利用效用函数的管理策略实现资源的优化调度。

关键词 云计算, QoS, 经济模型, 资源调度, SLA

中图法分类号 TP393 **文献标识码** A

Research on Cloud Economic Resource Management Model with QoS Constrains

QIAN Qiong-fen^{1,2} LI Chun-lin¹ ZHANG Xiao-qing¹

(School of Computer Science and Technology, Wuhan University of Technology, Wuhan 430063, China)¹

(NO. 4 Department, Air Force Radar Academy, Wuhan 430019, China)²

Abstract It is a major challenge for the cloud computing system to provide satisfied cloud application with QoS constraints. We proposed a cloud computing economic management model based on commodity market prototype, which implements the mapping from application service layer QoS to resource device layer QoS over the SLA negotiations between the users and providers in cloud computing market. Finally, the utility function as management strategy was used to optimize resource scheduling.

Keywords Cloud computing, QoS, Economic model, Resource scheduling, SLA

1 引言

云计算是目前信息领域研究的新热点。云计算端用户可以省去大量的信息基础设施建设投入费用,直接通过互联网按需获取远程计算、存储等多元化的、商业化的信息服务,云计算中“一切皆是服务”^[1,2]。云计算将大量虚拟的、动态的资源包装为互联网的服务提供给用户的同时,如何为各种各样的用户应用程序提供所需的服务质量保障是云计算面临的关键问题之一。服务等级协议 SLA 是实现用户与服务提供商之间服务质量的重要手段^[3]。

云计算是虚拟化技术、效用计算、网格计算的综合发展,是一种计算资源的商业实现。云计算模式可以视为生产者-消费者模式,即云用户根据自己的需求按照相应的付费方式从云计算市场购买云计算服务,云计算服务商提供云计算服务,并将其作为商品出售给消费者。采用市场经济管理模型,考虑在满足用户 QoS 需求的情况下,通过 SLA 协商,实现资源的最优分配具有重要意义。

典型的云计算资源提供策略主要有基于多级资源池的租借理论、基于随机整数规划的最优资源提供策略、基于经济学的资源提供策略。文献[4]采用虚拟化技术,将资源归类形成多级资源池,结合不同策略(共享策略、私有策略、收回策略和

借还策略)完成资源的调度。该方法可有效减少资源空闲,提高资源利用率,但没有从用户花费的角度进行考虑,也没有从资源的动态性方面进行考虑,资源提供的过程中缺乏灵活性。文献[5-7]使用随机整数规划优化资源提供方法,采用预订和即付即用两种方式给用户提供服务。该策略能提供动态的资源分配方法,考虑用户各个阶段的资源花费,能达到最小化客户资源花费的目的,但在不确定资源的预测精确性上还有待提高。文献[8]从经济学原理的角度出发提出了云计算经济学架构,设计了基于 SLA 的云资源管理经济模型。该策略为云消费者和供应商提供有关经济激励的反馈,提高资源利用率,避免了信息系统的重复建设,有助于实现云环境下资源的高效管理、优化配置,最大限度地满足用户服务质量需求。虽然此策略是从经济学角度出发,但是并没有为节省用户花费提供相关的策略,也同样没有考虑资源和价格的动态性。

文献[5]提出了一种公平的网格资源调度策略,在满足用户 QoS 需求的条件下,将任务分配给最合适的资源,在资源分配时,考虑最早时间竞争策略,每个任务按照调整过的公平竞争时间进行排序选择。文献[6]在本地资源管理系统 HPC4U 和 AssessGrid 中引入了 SLA 感知的概念,采用运行时间估计的资源管理策略,提供协作的监测协议,支持故障容

本文受国家自然科学基金项目(60970064),教育部新世纪优秀人才计划项目(NCET-08-0806),武汉市科技攻关项目(201010621207),霍英东教育基金项目(121067),中央高校基本科研业务费专项资金(2010-YB-19),软件开发环境国家重点实验室开放课题(SKLSDE-2011KF-01)资助。

钱琼芬(1970—),女,博士生,讲师,主要研究方向为网格计算、云计算, E-mail: qqfld@hotmail.com; 李春林(1974—),女,教授,博士生导师,主要研究方向为网格计算、云计算。

错,实现 QoS 管理。文献[7]提出了基于策略的最优时间任务调度资源管理,采用拍卖正比例共享模型,用户通过更高的出价获得更多共享份额,维护策略的一致性,保证用户配置的服务等级。文献[9]对基于 QoS 的资源管理系统进行了研究,提出了 QoS 层次模型、QoS 参数的映射和转换等在资源管理系统中的重要作用。

上述工作从不同的方面研究了云资源分配策略,但对 QoS 在云计算体系架构中的意义、QoS 管理与 SLA 管理的区别以及 QoS 对资源分配的影响没有做更多的深入刻画。

本文首先提出了 QoS 约束的云计算经济资源管理模型,分析了模型中引入 QoS 约束和 SLA 协商的意义;然后分析了 QoS 管理和 SLA 管理之间的关系;最后通过 SLA 协商实现应用层 QoS 到资源层 QoS 的映射,采用运行时间最小策略,实现云计算虚拟机资源的最优分配。

2 QoS 约束的云经济资源管理模型

云计算用户采用即付即用的模式从云供应商获得服务,云计算的系统结构设计可以采用生产者-消费者模型,使用人类经济市场的经济原理,实现云计算处理过程中的资源分配与管理。从云用户对资源的使用角度看,用户通过付费,从服务商处购买所需的资源,得到相应的服务,将商品市场作为云计算经济管理原型是自然的。因此,可以把云计算看作一个面向产品的计算市场环境^[10]。借鉴网格计算云经济市场中,可运用不同的经济模型来解决服务与用户需求的问题^[4-6]。

2.1 模型描述

基于 QoS 的云计算经济管理简化模型主要包括云用户、经纪人、云供应商、云服务市场^[4,14,15],如图 1 所示。模型中引入 QoS 管理机制和 SLA 管理机制。

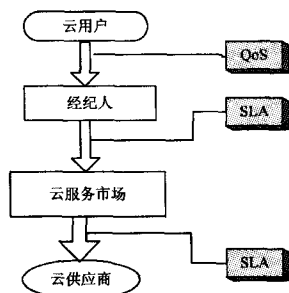


图 1 QoS 云计算经济管理简化模型

云用户:云用户通过经纪人向云服务市场提出自己的要求,包括所需完成任务的描述、服务质量要求 QoS。用户应用的 QoS 参数包括时间、成本、可靠性和信任系数等。

经纪人:经纪人接收云用户请求,根据用户的任务需求和 QoS 服务质量要求访问云服务市场,查找并选择适合云用户应用需求的资源。经纪人含有 SLA 谈判模块,此模块根据资源的当前条件和当前需求做出决定。

云供应商:云供应商可以集成各种资源提供特定的服务,也可以是一个具体的资源。云供应商将资源包装为服务,通过注册,进入云服务市场,退出时从市场撤消注册。

云服务市场:云服务市场提供云环境中云用户和云供应商的交易场所,是进行资源管理和交易的基础设施。云服务市场可以联接不同的云,具有服务信息目录、服务发现、服务

管理、注册、撤销、订阅、服务交易等管理功能。云服务市场含有 SLA 谈判模块。

引入 QoS 模块和 SLA 模块可以在云计算经济模型中分别实现 QoS 管理和 SLA 管理。云市场要实现资源监测、存储、网络、虚拟机、服务迁移和容错性等功能,必然面临 QoS 问题。服务质量 QoS 提供了服务性能保证、可用性保证,以及安全性、可靠性等其他方面的服务质量。QoS 要求与云服务提供商和云用户有关联,但云用户通过代理向云市场获取所需云服务 QoS 时,并不关心云市场和云服务提供商如何来达到服务 QoS,因此,SLA 在商定服务供应商和最终用户间的 QoS 发挥重要的作用。

服务等级 SLA 协议(Service Level Agreement, SLA)是由用户与服务提供商签订的关于提供商如何向用户提供服务的协议。SLA 包括技术部分和商业法律部分,技术部分被称为服务等级规范。SLA 管理的主要功能包括 SLA 咨询、SLA 创建、SLA 管理、违约检测、合约补偿、处罚等。由于这些内容是事先建立并写入合约的,不同的参与者有不同的理解,因此必然会发生冲突。为了维护自身的利益,每个参与者都必须对合约进行协商。所以 SLA 协商的主要任务是消除冲突,达成一致。SLA 协议分为 3 类:任务服务等级协议 TSLA,表明任务的性能;资源服务等级协议 RSLA,承诺对用户提供服务的服务质量;绑定服务等级协议 BSLA,将 RSLA 承诺的资源用于 TSLA 的任务。

2.2 应用层 QoS 到虚拟机 QoS 映射

QoS 是对一个或多个对象的集合行为的需求质量的描述,显然,在不同的应用环境中,QoS 有不同的含义。在传统网络应用、多媒体应用、网格应用和云计算应用中,QoS 对象及其描述是不一样的。正由于 QoS 的复杂性,QoS 参数表达能力的局限性,QoS 机制分为 QoS 提供机制、QoS 控制机制和 QoS 管理机制。QoS 提供机制提供 QoS 映射、资源预约、QoS 测试以及 QoS 协商/重协商功能。

为保证云用户的 QoS,云市场管理简化模型中,我们主要关注 QoS 的提供机制,关注如何通过 SLA 实现 QoS 的映射、协商。对 QoS 的研究可分为 3 个层次:应用层、系统层、资源层^[16],而且具有普遍的意义。在不同的层次,QoS 表达的方式不同,侧重不同。例如云用户通过虚拟桌面播放一段视频的应用场景中,应用层的服务质量关心画面质量的高、中、低,声音质量高、中、低。系统层关心虚拟机 CPU 占用时间、内存大小、缓冲区使用等,资源层关心具体物理资源的访问策略、资源管理算法。针对云计算中资源虚拟化的特点,可以将 QoS 分为 3 层:应用层 QoS、虚拟层 QoS、资源设备层 QoS,其中虚拟层具有承上启下的作用,可以将应用层的 QoS 需求映射到某种类型的虚拟层 QoS,同时屏蔽物理资源的异构性。根据虚拟层 QoS 参数所表达的 QoS 特性,可以将虚拟层 QoS 分为逻辑资源类、系统类、安全类、信任类和记账类。其中逻辑资源 QoS 和系统 QoS 对云服务的服务能力产生决定性的影响,云用户在申请服务时,可以选择不同的系统 QoS 和逻辑资源 QoS 以达到 QoS 需求。安全 QoS、信任 QoS 和记账 QoS 这 3 种类型的 QoS 参数都属于服务属性方面的 QoS,它们对云计算服务的服务能力不会产生决定性的影响,只是向用户提供一些必须的 QoS 信息,用户没必要也不能决定这些 QoS 参数的具体值,这类 QoS 参数在 QoS 监控、维护等方面

发挥重要作用。资源设备层的 QoS 分为资源环境 QoS、网络 QoS、物理资源 QoS 及其它的相应模块。

QoS 参数的映射转换过程就是不同 SLA 建立和维护的过程。由于云计算应用资源的虚拟化和分布性,对完成一个应用或服务所需的所有资源建立一次 SLA 协商是不现实的,因此,可以将 SLA 分解,在不同的资源组之间实现协作管理,本文提出了基于 QoS 约束的二次 SLA 协商资源分配方法。当云用户通过经纪人向云市场提出带 QoS 约束的服务要求时,经纪人与云市场为每个服务进行第一次 SLA 协商,服务被分解为不同的任务,每个任务建立相应的 TSLA,即实现从应用到虚拟资源的映射。QoS 协商成功时,形成一组 TSLA 映射关系。

假设云应用为一组云服务的集合,云应用 P 与服务 S 的关系为:

$$P = \{S_1, S_2, \dots, S_i\} \quad (1)$$

每个服务 S_i 对应一个 TSLA 协商过程,包含不同的 QoS 需求,那么,可以将 TSLA 协商过程看作是应用服务层 QoS 到云市场虚拟层 QoS 的映射转换,则云服务可以转换为一组 QoS 需求的组合:

$$S_i = \{Q_{i1}, Q_{i2}, \dots, Q_{ij}\} \quad (2)$$

于是,一个云应用可以表示为 QoS 需求的集合:

$$P = U\{S_i\} = U\{U\{Q_{ij}\}\} \quad (3)$$

另一方面,资源设备层可以抽象为一组逻辑资源的集合,即:

$$R = \{R_1, R_2, \dots, R_m\} \quad (4)$$

尽管资源层由物理资源构成,但当资源设备层向云市场提供逻辑资源时,需要综合考虑物理资源本身的整体服务能力、资源本地任务的负载情况以及资源的共享策略等因素,抽象出逻辑资源的 QoS 参数。云计算环境中,物理机通过虚拟化技术,以虚拟机的方式提供逻辑资源,抽象逻辑资源组合为一个虚拟机资源,虚拟机可以看作是一组 QoS 需求的集合,即:

$$R_m = \{q_1, q_2, \dots, q_k\} \quad (5)$$

云资源管理的最终目的是将云服务分解为一组最小调度单位的任务集合,并将任务按照一定的目标函数进行调度,实现用户的满意度。在由逻辑资源向物理资源映射的过程中,需要进行第二次 SLA 协商(RSLA),比如实现系统 QoS 到网络 QoS 和环境 QoS 的映射,逻辑资源 QoS 到物理资源 QoS 的映射,安全 QoS 到本地安全策略模块的映射,可用性 QoS 到本地可用性管理模块的映射,计费 QoS 到本地计费模块的映射。

3 QoS 约束的资源分配算法

3.1 SLA 的协调机制

根据第 2 节的分析,可以得到具有 QoS 约束的二次 SLA 协商机制,如图 2 所示。

步骤 1 云用户在云市场注册,提出云服务 QoS 需求。

步骤 2 用户经纪人向云市场发出包含服务请求的 TSLA 合同。

步骤 3 云市场解析 TSLA 合同,将服务 QoS 映射为一组 QoS 需求的组合,通过云市场管理监督模块进行 TSLA 协商。若 TSLA 协商失败,交易结束。

步骤 4 TSLA 协商成功,实现服务层 QoS 到虚拟层 QoS 的映射,同时实现服务到任务的分解。

步骤 5 云市场向供应商发送包含虚拟层 QoS 请求的 RSLA 合同,通过云市场管理监督模块进行 RSLA 协商,若 RSLA 协商失败,交易结束。

步骤 6 RSLA 协商成功,实现虚拟层 QoS 到资源设备层 QoS 的映射,将任务调度到资源上。

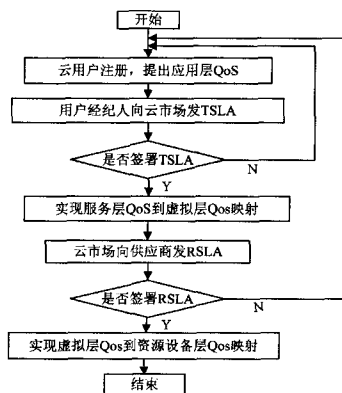


图 2 QoS 约束的二次 SLA 协商机制

3.2 基于效用函数的虚拟机资源分配

从上面的分析可以看到,二次 SLA 协商的过程可以简单抽象为从 QoS 服务到 QoS 资源的映射,在实现 QoS 映射的同时,完成资源的调度,即式(3)到式(5)的匹配过程,匹配策略为:

$$\begin{cases} F_{map} = P \rightarrow R \\ s. t. \phi(F_{map}) = \text{Min} \end{cases} \quad (6)$$

通过 RSLA 协商进行资源调度算法的实质就是找到匹配函数 F_{map} ,使得目标函数 $\phi(F_{map})$ 效用最优,效用函数可以根据资源管理策略目标的不同而有不同的定义,本文采用最小的执行时间。

云计算中,逻辑资源为虚拟机资源。设用户任务集 $T = \{T_1, T_2, \dots, T_n\}$,任务 T_i 用 MI 表示,即任务的指令数大小, $Time_T$ 为用户任务的完成时间;虚拟机资源集合 $V = \{VM_1, VM_2, \dots, VM_m\}$,其中,具体的虚拟机 VM_j 用 MIPS 表示,即虚拟机的内核处理速度,一般情况下, $n > m$ 。

效用函数定义为: $U = \text{Min}(Time_T)$ 。基于该效用函数最优化的算法如表 1 所列。

表 1 基于最小运行时间的资源调度算法

算法:最小运行时间虚拟机分配	
Step1	获取任务列表中的任务数 n
Step2	获取虚拟机列表中的虚拟机数 m
Step3	DownSort(T) //按 MI 降序排列任务集
Step4	UpSort(V) //按 MIPS 升序排列虚拟机资源集合
Step5	For($i=1; i \leq n; i++$) For($j=1; j \leq m; j++$) $E(i, j) = T_i / V_j$ //此时 T, V 为排序后的任务集和虚拟机集
Step6	从矩阵中行号为 0 的任务开始,每次都尝试分配给最后一列对应的 VM,若该虚拟机是最优的,就完成分配;否则将任务分配给使当前结果最优的 VM。
Step7	重复 Step6 到 Step7,直到将所有任务分配到虚拟机上。
Step8	End

值得注意的是,该算法中用 MIPS 描述虚拟机,任务执行时间的选值不能太小,必须高于最小门限,否则不能找到最优

(下转第 214 页)

其中,IsPacketArrived()函数判断是否有数据到达,有则调用ReadRfdData()函数将数据存放在数组 rxBuff 中;然后由 frameCorrect()判断接受到的数据是否符合发送端定义的数据结构,即数据帧是否包括前导码、定界符、结束符以及 CRC 校验码是否正确,如果符合,则在数码管显示温度,并将结果发送给 PC 机。

4.3 实验结果验证

通过对硬件进行调试,能正确地接收到传感器采集到的温度数据。图 6 为数码管实时温度显示结果,图 7 为 PC 机串口接收的数据显示效果。

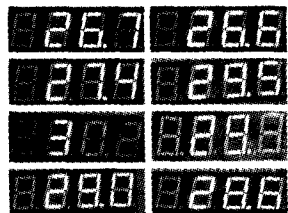


图 6 数码管显示效果

结束语 基于 CC1100 无线数据传输的单片机通信系统,可实现短距离的无线数据通信,通过实验验证该无线数据传输系统运行良好。本文详细讲述了系统设计方案,并给出了相关程序流程。可以将其应用在短距离无线通信环境中,如果把本设计扩展为具有串口传输给上位机系统的功能,则可以实现具有无线短距离温度监控系统,其将具有更大的应用价值。

(上接第 197 页)

的虚拟机。另外,任务数应远远大于虚拟机数,才能充分体现算法的意义和优越性。

结束语 云计算提供了一种基于互联网的全新计算模式,受到学术界高度关注,众多厂商提供的服务模式在商业上获得了巨大的成功和广泛的应用。本文的云经济资源管理模型中,用户从云市场获取所需服务,通过 QoS 约束和 SLA 协商保证云服务质量,同时实现资源的优化调度。对云计算中 QoS 管理模块和 SLA 管理模块的设计、QoS 参数性能的测量评估是下一步需要研究的内容。

参考文献

- [1] Rimal B P, Jukan A, Katsaros D, et al. Architectural Requirements for Cloud Computing System; An Enterprise Cloud Approach[J]. Grid Computing, 2010, 9(1): 3-26
- [2] Armbrust M, Fox A, Griffith R, et al. Above the clouds: a Berkeley view of cloud computing[R]. UCB/EECS-2009-28, Electrical Engineering and Computer Sciences, University of California at Berkeley, 2009
- [3] Maurer M, Ivona Brandic V C E, et al. Cost-benefit analysis of an SLA mapping approach for defining standardized cloud computing goods [J]. Future Generation Computer Systems, 2011, doi:10.1016/j.future.2011.05.023
- [4] Buyya R, Abramson D, Giddy J. et al. Economic models for resource management and scheduling in grid computing[J]. Concurrency and Computing, 2002, 14: 1507-1542
- [5] Neumann D, Christof Weinhardt J S. Bridging the Adoption Gap Developing a Roadmap for Trading in Grids[J]. Electronic Markets, 2008, 18(1): 65-74
- [6] Nimis J, Anandasivam A, Borissov N, et al. SORMA-Business Cases for an Open Grid Market; Concept and Implementation

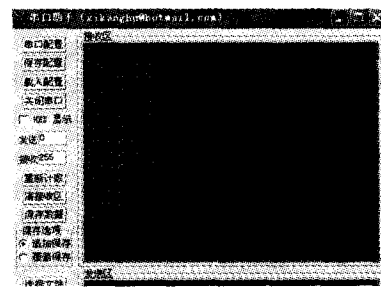


图 7 PC 机接收数据的效果

参考文献

- [1] 刘志东, 陈彦明. 基于 Zigbee 的无线传感器网络节点的设计[J]. 桂林电子科技大学学报, 2008
- [2] 王红玲, 白政民. 基于 AT89C51 的多点温度检测系统设计[J]. 现代电子技术, 2010, 01
- [3] 王伟, 景博, 赵卉. 一种无线传感器节点及通讯协议的研究[J]. 微计算机信息, 2008, 10
- [4] CC1100 的技术手册
- [5] 赵宏图. 基于单片机的温度控制系统的设计与实现[J]. 微计算机信息, 2008
- [6] 彭伟. 单片机 C 语言程序设计实训 100 例[M]. 北京: 电子工业出版社, 2009
- [7] 杨绍恭, 伍彩琴, 董国芳. 基于 WSN 节点定位与道路方位信息的地图匹配改进算法[J]. 云南民族大学学报, 2011
- [8] [C]//Proceedings of the 5th international workshop on Grid Economics and Business Models(GECON'08). 2008: 173-184
- [7] Younge A J, von Laszewski G, Wang Li-zhe, et al. Efficient resource management for Cloud computing environments [C] // Green Computing International Conference. 2010: 357-364
- [8] 孙瑞锋, 赵政文. 基于云计算的资源调度策略[J] 航空计算技术, 2010, 40(3): 103-105
- [9] 高宏卿, 邢颖. 基于经济学的云资源管理模型研究[J]. 计算机工程与设计, 2010, 31(19): 4139-4212
- [10] Doulamis N, Litke A D A, et al. Adjusted fair scheduling and non-linear workload prediction for QoS guarantees in Grid computing[J]. Computer Communications, 2007, 30(3): 499-515
- [11] Battre D, Hovestadt M, Keller A, et al. Planning-based scheduling for SLA-awareness and Grid Integration [C] // Proc. of Workshop of the UK Planning and Scheduling. Special Interest Group, University of Paderborn. 2007
- [12] Sulistio A, Buyya R. A Time Optimization Algorithm for Scheduling Bag-of-Task Applications in Auction-based Proportional Share Systems[C] // Proc. of the 17th International Symposium on Computer Architecture and High Performance Computing (SBAC-PAD'05). 2005: 235-242
- [13] 伍之昂, 罗军舟, 宋爱波. 基于 QoS 的网格资源管理[J]. 软件学报, 2006, 17(11): 2264-2276
- [14] Buyya R, Sudharshan Vazhkudai. Compute Power Market: Towards a market-oriented grid[C] // Proceedings of First IEEE/ACM International Symposium on Cluster Computing and the Grid. 2001: 574-581
- [15] Lai K, Rasmusson L, Adar E, et al. Tycoon: An implementation of a distributed market-based resource allocation system[J]. Multi agent and Grid Systems, 2005(3): 169-182
- [16] Sabata B, Chatterjee S, Davis M, et al. Taxonomy for QoS specifications[C] // Proc of the 3rd Int'l Workshop on Object-Oriented Real-Time Dependable Systems. 1997: 100-107