第22卷 第8期
2012年8月

计算机技术与发展
COMPUTER TECHNOLOGY AND DEVELOPMENT

Vol. 22 No. 8
Aug. 2012

# 云计算环境中 SaaS 的接入控制和调度策略研究

李 波,杨从有,武 浩,裴以建
(云南大学 信息学院,云南 昆明 650091)

**摘 要**:SaaS 是一种基于网络的软件应用模式,是服务提供商将应用软件统一部署在自己的服务器上,用户根据自己的实际需要,通过互联网向服务提供商订购并支付自己所需的服务。在未来,SaaS 模式是占主导地位的云服务模型。文中阐述 SaaS 的基本概念,介绍了 SaaS 的参考结构以及服务流程,分析概括了不同类型的服务要求的接入控制策略,总结了不同性能要求作业的调度策略,最后结合已有的云计算环境下的 SaaS 接入控制和调度策略研究成果,展望了未来的研究方向和亟待解决的关键问题。

**关键词**:云计算;软件即服务 SaaS;接入控制;调度策略
**中图分类号**:TP31 **文献标识码**:A **文章编号**:1673-629X(2012)08-0009-04

# Survey of Admission Control and Scheduling Mechanisms for Software-as-a-Service in Cloud Computing Environments

LI Bo, YANG Cong-you, WU Hao, PEI Yi-jian
(School of Information Science and Engineering, Yunnan University, Kunming 650091, China)

**Abstract**: SaaS is a kind of network-based software application paradigm that service providers deploy their application software on their servers. Users order and pay for their actual services via the internet. In the future, the SaaS model will be the dominant cloud service model. It introduces the concept of SaaS, its architecture and its service processes, and analyzes the types of admission control and scheduling algorithms for different service and performance requirements. It also presents a summary of the current state-of-the-art of the admission control and scheduling algorithms for SaaS in cloud computing environments, a discussion on the future work and some crucial problems should be solved pressingly.

**Key words**: cloud computing; soft as a service (SaaS); admission control; scheduling algorithms

## 0 引 言

云环境是一种商业模式,是一个极具发展前景的网络环境。随着云技术的发展及相关标准的逐渐成熟,在可以预见的未来里,SaaS 模式是占主导地位的云服务模型[1],云将像当年的互联网一样,从根本上改变人们的生活方式。云计算提供了最可靠、最安全的数据存储中心,用户不用再担心数据丢失、病毒入侵等麻烦;云计算对用户端的设备要求低,使用起来方便快捷;云计算易于实现不同设备间的数据与应用共享;云计算为人们使用网络提供了几乎无限多的可能等显著特点;云环境下的 SaaS 具有成本、性能、功能集成、安全、个性化应用和互联网等优势[2,3]。

软件即服务 SaaS(Software-as-a-Service),也称为软件服务化,是随着互联网技术的快速发展和应用软件的逐渐成熟,在二十一世纪初兴起的一种完全创新的通过 Internet 提供软件的应用模式[2,3]。SaaS 是服务提供商将应用软件统一部署在自己的服务器上,用户根据自己的实际需要,通过互联网向服务提供商订购所需的服务,按订购服务的多少、服务类型、使用时间长短向服务提供商支付费用,用户享有软件的使用权和升级权,可随时随地使用软件,而终端无需维护相应软件。目前,Google Apps 企业应用套件和 Salesfore 公司提供的在线客户关系管理 CRM(Client Relationship Management)服务属于这类服务。

目前,云环境下 SaaS 的研究主要集中在:体系结构、服务发现、安全性问题、接入控制和调度策略等方面。体系机构主要从用户、管理者和开发商三个不同的角度进行剖析,设计更完善的结构[4];服务发现是用户通过因特网找到相关的云服务的系统;安全性问题是涉及到特权用户的接入、可审查性、数据位置、数据隔离、数据恢复、调查支持和长期生存性等七大风险的

**收稿日期**:2011-12-27;**修回日期**:2012-03-30
**基金项目**:国家自然科学基金(60663009);云南大学中青年骨干教师培养计划
**作者简介**:李 波(1976-),男,云南曲靖人,副教授,博士,主要研究方向为网络和分布式计算领域的资源管理和调度。

相关问题[5]；接入控制是SaaS提供商用来解释和分析用户的服务质量QoS(quality of service)参数，并依据系统当前资源的计算能力和可用性来决定是否接收或拒绝该请求；调度策略是基于接入控制的决定，SaaS提供商的平台层给作业请求分配资源的方案，决定了作业的执行顺序，并根据虚拟机的价格、动态服务的初始化时间和数据传输时间来决定在哪儿初始化以及初始化什么类型的虚拟机，还负责管理减少当用户共享资源处理动态服务要求时的违规处罚。

接入控制和调度策略是SaaS研究的核心内容，是SaaS提供商是否接收或拒绝作业请求以及如何给作业分配资源的依据，会严重影响到用户和服务提供者的性能。文中对此问题进行了研究，介绍了参考的体系结构以及服务流程，归纳了不同参考因素的接入控制策略，总结了不同性能要求的作业的调度策略，并对未来的研究方向进行了展望。

## 1 体系结构及服务流程

文中研究的云计算体系结构中的SaaS层，参考的结构是文献[6]提到的SaaS层结构的系统模型。该结构从上到下依次是应用层、平台层、基础设施层，应用层负责管理服务提供商提供给用户的服务，平台层负责把用户的QoS映射到基础设施层和调度、配置虚拟机，基础设施层负责控制虚拟机的初始化和销毁。

在上述结构体系中，服务流程如下：

(1)服务发现，用户和提供商签订服务等级协议。目前，有基于语义的服务发现[7]、基于代理的服务发现[8,9]、采用非的P2P技术的服务发现[10]等系统；

(2)用户根据服务等级协议，按QoS要求向SaaS提供商递交作业请求；

(3)SaaS提供商在接到作业请求后，依据接入控制条件来决定是否接纳该作业，满足控制条件就接纳，否则拒绝；

(4)作业接纳后，为实现效益最大化目的，在调度列表上对该作业进行调度，使之处在最合适的位置；

(5)按调度列表的顺序执行该作业；

(6)监测该作业的整个处理过程，若出现违规，进行惩罚处理；

(7)最后，将执行结果递交给用户。

在上述服务流程中，任务的接入控制和调度策略是SaaS研究的核心内容，是SaaS提供商是否接收或拒绝作业请求以及如何给作业分配资源的依据，会严重影响到用户和服务提供者的性能。

## 2 接入控制策略

接入控制的目的是避免资源过载和能满足用户的性能要求、服务提供商的利益要求。其主要作用是判断是否接收一个新作业，满足接入控制条件就接收，否则拒绝该作业。

控制条件可分为两类：一类是从资源可用性方面来考虑是否接纳作业[11-14]；另一类是在资源能满足用户提交的QoS的条件下考虑所获得利润的多少来决定是否接纳作业[15,16]。

文献[11,12]从用户作业的属性的角度，侧重于用户QoS的接入控制策略。文献[11]是针对抢占的、独立的、实时的截止时间不变的周期性任务，提出了一种要求截止时间小于等于任务周期的连续时间的接入控制策略。文献[12]针对采用截止时间单调的调度任务，提出了一种近似最优的恒定时间接入控制策略，采用可配置的线性分段数量来近似任务的执行时间要求。

文献[15,16]从商业的角度，根据作业的大小、截止时间、预算(Budget)以及执行该作业的成本(Cost)计算出SaaS提供商的收益参考值来进行接入控制。文献[15]依据投资回报率，在已有的初始化一台虚拟机策略、等待策略、插入策略和惩罚延迟策略的基础上，提出了在最小化虚拟机数量、重调度作业列表和执行惩罚因子三种情况下最大化提供商利益的接入控制策略。文献[16]依据执行一个作业的潜在获利值，提出了一种只有当潜在的获利值大于预设的阈值时才接纳的接入控制策略。目前研究的调度策略主要是面向市场的、商业化的，一般都是基于截止时间和成本约束的任务。

在接入控制方法方面，文献[13,14]是基于控制理论提出的接入控制策略，都采用了一个预测器来预测状态的变化，以及一个自适应的反馈控制器来修正当前系统的偏差。文献[13]采用了一个卡尔曼滤波器来预测未来负载的变化、一个前馈控制模型识别预测器控制系统状态使之处在动态的平衡操作点附近，自适应反馈来纠正处在不准确的系统模型中带来的误差，目的是防止超载、保证作业的响应时间的一种自治接入控制策略。文献[14]采用了由分析预测模块和自适应预测模块两部分组成的预测系统的接入控制策略。该预测系统负责预测为了接收一个新作业，在满足该作业SLA条件下需要的最少资源，并修正系统存在的偏差。

在上述接入控制策略中，基于控制理论提出的策略侧重接入控制的方法，从用户作业的属性角度提出的策略侧重用户的QoS，从商业角度提出的策略侧重SaaS提供商的利益。但是，不管侧重点是什么，所有的策略都是为了避免资源过载和能满足用户的性能要求、服务提供商的利益要求。

## 3　调度策略

调度策略依据用户递交作业的大小、截止时间、响应时间、执行时间、惩罚率、预算等参数以及 SaaS 提供商的目标给作业分配资源,并决定了作业的执行顺序,其目的主要包括提高资源利用率、满足用户作业的 QoS、实现 SaaS 提供商的目标等。

依据作业性能要求,现有调度策略可分为尽快完成的、只有截止时间限制的、只有成本限制的以及有截止时间和成本双重限制的四种类型。其中,成本相关的调度策略在调度过程中需要考虑用户的成本和服务提供商收益。尽快完成的调度策略主要解决的问题是没有 QoS 保证的作业,性能目标主要是作业等待时间、响应时间、完成时间最小化和资源利用率最大化等。

已有策略主要包括:先来先服务 FCFS(First Come First Serve)、最短作业优先 SJF(Shortest Job First)、最大作业优先 LJFS(Largest Job First Served)、随机调度(Random scheduling)、回填调度(Backfill scheduling)等策略[17,18]。其中,FCFS 是最基本的调度策略,也是最早采用的策略;SJF 执行时间短的优先级高;LJFS 有利于大作业的调度;随机调度是随机选择作业调度;Backfill 调度利用了在各调度中产生的时间碎片,提高了资源的利用率。

有截止时间限制的调度策略主要解决的问题是有截止时间限制的作业,性能目标主要是作业的 SLA 的违规率最小化、资源利用率最大化等。已有策略主要包括:速率单调策略 RMS(Rate Monotonic Scheduling)、最小截止时间调度策略 EDF(Earliest Deadline First)、最小松弛度优先策略 LLFS(Least Laxity First Scheduling)等策略[19~22]。其中,RMS 是一种适用于可抢占的硬实时周期性作业调度的静态优先级调度策略;EDF 是按作业的截止时间进行排序,截止时间小的作业优先执行的策略;LLFS 是结合作业执行的缓急程度(空闲时间)来给作业分配优先级,作业的松弛度越小,越需要尽快执行的策略。

依据 SaaS 提供商的利益要求,文献[6,23]研究了基于成本限制的调度策略。文献[6]依据已初始化的虚拟机的可用空间大小提出了将新作业安排到具有最大、最小可用资源的虚拟机上的两种最小化基础设施成本的调度策略。文献[23]在动态的环境中处理不同用户递交的作业,在最大作业优先、最短作业优先、重塑的价钱优先、重塑的报酬优先的调度策略基础上,提出了利润优先、机会优先和机会比率优先三种实现 SaaS 提供商利益最大化的调度策略。

依据是能满足截止时间和 SaaS 提供商的利益要求,文献[14,24,25]是基于截止时间和成本限制下的调度策略。文献[14]依据当系统处理一个作业时可能带来利润或是惩罚的作业模型[26],提出了一种在线的云计算服务环境下新颖的获利调度策略。根据文中定义的最好、最差的执行时间,截止时间,执行时间的概率密度函数以及利益时间实用函数和惩罚时间实用函数,计算出执行一个作业的潜在获利值,最后根据该值来决定如何调度该作业。文献[24]依据服务提供商重点考虑的参数不同,提出了成本和时间两个面向市场的调度策略,采用了向 IaaS 提供商租借资源来扩大本地资源的计算能力,来达到满足服务请求的目的。文献[25]在截止时间和预算限制下提出了成本和时间优化策略,对服务请求的资源配置考虑了多个云,并引入了 RAINBOW 结构的概念,满足用户的服务要求,同时最小化了成本。

在上述调度策略中,尽快完成的和只有截止时间限制的基本调度策略是基本、常见常用的策略;只有成本限制的调度策略是依据 SaaS 提供商的利益要求;有截止时间和成本限制双重限制的调度策略,其依据是能满足截止时间和 SaaS 提供商的利益要求。不管依据何种参数、何种分类,所有的调度策略都是为了实现提高资源利用率,降低资源成本,满足用户和 SaaS 提供商的要求。目前研究的调度策略基本都是面向商业的,一般都是基于截止时间和成本限制的作业。

依据作业属性,调度策略可分为在线调度策略和离线调度策略两种。在云环境下,一般都是在线调度,很少有离线调度。离线调度在调度之前,系统就知道作业的 QoS 及各种相关信息,而在线调度却不知道。

## 4　结束语

文中对接入控制和调度策略进行了研究,介绍了参考的体系结构以及服务流程,归纳了不同参考因素的接入控制策略,总结了不同性能要求的作业的调度策略,并对未来的研究方向进行了展望。在未来,随着商业模式下的云计算的服务由单一、简单服务向复杂的综合服务转变,而且要求服务动态化、自动化,对算法的研究也从单层向跨层过渡[27],这对接入控制和调度策略的研究提出了严谨的挑战。

一是随着作业的复杂多样化、服务提供商的价钱策略的多变性,接入控制和调度策略变得越来越复杂,对策略进一步改进和创新适应未来的发展变化;

二是目前的接入控制和调度策略很少考虑时间复杂度,很多都是遍历策略,策略的时间复杂度有待改善;

三是惩罚机制的选择。惩罚机制很大程度上影响着接入控制和调度策略,一个合理的惩罚机制不仅能提高服务提供商的信誉同时也是最优化性能要求。

**参考文献：**

[1] Brodkin J. Gartner: Seven cloud-computing security risks [M]//Infoworld. [s. l. ]:[s. n. ],2008:1-3.

[2] 刘 鹏. 云计算[M]. 第 2 版. 北京:电子工业出版社, 2011.

[3] 吴朱华. 云计算核心技术剖析[M]. 北京:人民邮电出版社,2011.

[4] Carstoiu B. Cloud SaaS infrastructure[J]. UPB Scientific Bulletin, Series C: Electrical Engineering, 2011, 73(2): 89-102.

[5] Popovic K, Hocenski Z. Cloud computing security issues and challenges[C]//33rd International Convention on Information and Communication Technology, Electronics and Microelectronics. [s. l. ]:[s. n. ],2010:344-349.

[6] Wu L, Garg S K, Buyya R. SLA-based Resource Allocation for Software as a Service Provider (SaaS) in Cloud Computing Environments[C]//CCGrid. [s. l. ]:[s. n. ],2011:195-204.

[7] Chen F, Bai X, Liu B. Efficient service discovery for cloud computing environments[C]//International Conference on Advanced Research on Computer Science and Information Engineering. [s. l. ]:[s. n. ],2011:443-448.

[8] Han T, Sim K M. An ontology-enhanced cloud service discovery system[C]. International MultiConference of Engineers and Computer Scientists 2010, 2010. 644-649.

[9] Han T, Sim K M. An agent-based cloud service discovery system that consults a cloud ontology[C]//International Conference on Advances in Intelligent Control and Computer Engineering. [s. l. ]:[s. n. ],2011:203-216.

[10] Zhou J, Shi Z. Unstructured P2P-enabled service discovery in the cloud environment[C]//6th IFIP International Conference on Intelligent Information Processing. [s. l. ]:[s. n. ], 2010:173-182.

[11] Masrur A, Chakraborty S, Farber G. Constant-time admission control for deadline monotonic tasks[C]//Design, Automation and Test in Europe Conference and Exhibition. [s. l. ]: [s. n. ],2010:220-225.

[12] Masrur A, Chakraborty S. Near-optimal constant-time admission control for DM tasks via non-uniform approximations[C]//17th IEEE Real-time and Embedded Technology and Applications Symposium. [s. l. ]:[s. n. ],2011:57-67.

[13] Leontiou N, Dechouniotis D, Denazis S. Adaptive admission control of distributed cloud services[C]//2010 International Conference on Network and Service Management. [s. l. ]: [s. n. ],2010:318-321.

[14] Ventura G R, López J A, Fernández J G. Deadline constrained prediction of job resource requirements to manage high-level SLAs for SaaS cloud providers[C]//NCA, 2010. [s. l. ]:[s. n. ],2010:224-231.

[15] Wu L, Garg S K, Buyya R. SLA-based Admission Control for a Software-as-a-Service Provider in Cloud Computing Environments[C]//CCGrid. [s. l. ]:[s. n. ],2011:132-146.

[16] Liu S, Quan G, Ren S. On-line scheduling of real-time services for cloud computing[C]//2010 6th World Congress on Services. [s. l. ]:[s. n. ],2010:459-464.

[17] Li W, Shi H. Dynamic load balancing algorithm based on FCFS[C]//2009 4th International Conference on Innovative Computing, Information and Control. [s. l. ]:[s. n. ],2009: 1528-1531.

[18] Jiang H, Ni T. PB-FCFS-A Task Scheduling Algorithm Based on FCFS and Backfilling Strategy for Grid Computing [C]//2009 Joint Conferences on Pervasive Computing. [s. l. ]:[s. n. ],2009:507-510.

[19] Han S, Park M. Predictability of least laxity first scheduling algorithm on multiprocessor real-time systems[J]. Emerging Directions in Embedded and Ubiquitous Computing, 2006, 4097:755-764.

[20] Liu C L, Layland J W. Scheduling algorithms for multiprogramming in a hard-real-time environment[J]. Journal of the ACM (JACM), 1973, 20(1):46-61.

[21] Saleh M, Dong L. Comparing FCFS EDF scheduling algorithms for real-time packet switching networks[C]//2010 International Conference on Networking, Sensing and Control. [s. l. ]:[s. n. ],2010:698-703.

[22] Stankovic J, Ramamritham K, Spuri M, et al. Deadline scheduling for real-time systems: EDF and related algorithms [M]. [s. l. ]:Springer, 1998.

[23] Popovici F I, Wilkes J. Profitable services in an uncertain world[C]//SC '05 Proceedings of the 2005 ACM/IEEE Conference on Supercomputing. Washington: IEEE Computer Society, 2005.

[24] Salehi M A, Buyya R. Adapting market-oriented scheduling policies for cloud computing[C]//10th International Conference on Algorithms and Architectures for Parallel Processing. [s. l. ]:[s. n. ],2010:351-362.

[25] Chintapalli V R. A deadline and budget constrained cost and time optimization algorithm for cloud computing[C]//1st International Conference on Advances in Computing and Communications. [s. l. ]:[s. n. ],2011:455-462.

[26] Yu Y, Ren S, Chen N, et al. Profit and penalty aware (pp-aware) scheduling for tasks with variable task execution time [C]//SAC'10 Proceedings of the 2010 ACM Symposium on Applied Computing. [s. l. ]:[s. n. ],2010:334-339.

[27] Theilmann W, Yahyapour R, Butler J. Multi-level sla management for service-oriented infrastructures[M]//Towards a Service-Based Internet. [s. l. ]:[s. n. ],2008:324-335.

# 无线传感器网络的 RPL 路由协议研究

作者：朱琳， 高德云， 罗洪斌
作者单位：北京交通大学电子信息工程学院,北京100044
刊名：计算机技术与发展
英文刊名：Computer Technology and Development
年，卷(期)：2012(8)

## 参考文献(27条)

1. Brodkin J Gartner:Seven cloud-computing security risks 2008
2. 刘鹏 云计算 2011
3. 吴朱华 云计算核心技术剖析 2011
4. Carstoiu B Cloud SaaS infrastructure 2011(02)
5. Popovic K;Hocenski Z Cloud computing security issues and challenges 2010
6. Wu L;Garg S K;Buyya R SLA-based Resource Allocation for Software as a Service Provider(SaaS) in Cloud Computing Environments 2011
7. Chen F;Bai X;Liu B Efficient service discovery for cloud computing environments 2011
8. Han T;Sim K M An ontology-enhanced cloud service discovery system 2010
9. Han T;Sim K M An agent-based cloud service discovery system that consults a cloud ontology 2011
10. Zhou J;Shi Z Unstructured P2P-enabled service discovery in the cloud environment 2010
11. Masrur A;Chakraborty S;Farber G Constant-time admission control for deadline monotonic tasks 2010
12. Masrur A;Chakraborty S Near-optimal constant-time admission control for DM tasks via non-uniform approximations 2011
13. Leontiou N;Dechouniotis D;Denazis S Adaptive admission control of distributed cloud services 2010
14. Ventura G R;López J A;Fernández J G Deadline constrained prediction of job resource requirements to manage highlevel SLAs for SaaS cloud providers 2010
15. Wu L;Garg S K;Buyya R SLA-based Admission Control for a Software-as-a-Service Provider in Cloud Computing Environments 2011
16. Liu S;Quan G;Ren S On-line scheduling of real-time services for cloud computing 2010
17. Li W;Shi H Dynamic load balancing algorithm based on FCFS 2009
18. Jiang H;Ni T PB-FCFS-A Task Scheduling Algorithm Based on FCFS and Backfilling Strategy for Grid Computing 2009
19. Han S;Park M Predictability of least laxity first scheduling algorithm on multiprocessor real-time systems 2006
20. Liu C L;Layland J W Scheduling algorithms for multiprogramming in a hard-real-time environment 1973(01)
21. Saleh M;Dong L Comparing FCFS EDF scheduling algorithms for real-time packet switching networks 2010
22. Stankovic J;Ramamritham K;Spuri M Deadline scheduling for real-time systems:EDF and related algorithms 1998
23. Popovici F I;Wilkes J Profitable services in an uncertain world 2005
24. Salehi M A;Buyya R Adapting market-oriented scheduling policies for cloud computing 2010
25. Chintapalli V R A deadline and budget constrained cost and time optimization algorithm for cloud computing 2011
26. Yu Y;Ren S;Chen N Profit and penalty aware(pp-a-ware) scheduling for tasks with variable task execution time 2010
27. Theilmann W;Yahyapour R;Butler J Multi-level sla management for service-oriented infrastructures 2008