

一种分布式科技资源池架构及其交互行为分析

张妮妮^{1,2}, 李苗钰³, 石梅³, 朱省吾³, 黄位³, 郭军⁴

(1. 西北大学 物联网研究中心, 陕西 西安 710127;

2. 西安思源学院, 陕西 西安 710019;

3. 西北大学 信息学院, 陕西 西安 710127;

4. 西北大学 京东人工智能与物联网联合研究院, 陕西 西安 710127)

摘要:科技资源服务是云计算环境下数据服务领域的重要业务之一。但是,由于科技资源归属单位不同,条块分割,实现跨平台服务与共享存在很大障碍。为了解决资源空间分散、业务并发等问题,需要构建基于分布式虚拟化存储技术的科技资源服务业务虚拟化资源池,统一管理资源并提供可靠的数据访问和交互功能。针对这一应用背景,基于云计算理论,在软硬件虚拟化技术的基础上,该文提出了一种分布式虚拟化科技服务资源池多层次模型架构,该模型将资源池应用、汇聚、虚拟和物理层进行了统一描述;在此基础上,利用Pi演算这一描述移动并发系统的数学工具,对所构建的集成化科技资源服务模型的外部行为进行描述,通过Pi演算的推演规则分析系统的交互行为,形式化验证了模型系统数据汇聚和交互行为的安全性和可靠性。提出的模型和取得的研究结果可为实际系统的设计开发提供重要的理论依据和参考。

关键词:分布式系统;科技资源池;Pi演算;交互行为;数据汇聚

中图分类号:TP391

文献标识码:A

文章编号:1673-629X(2021)12-0105-05

doi:10.3969/j.issn.1673-629X.2021.12.018

A Distributed Scientific and Technological Resource Pool Architecture and Its Interact Behavior Analysis

ZHANG Wei-wei^{1,2}, LI Miao-yu³, SHI Mei³, ZHU Sheng-wu³, HUANG Wei³, GUO Jun⁴

(1. Internet of Things Research Center, Northwest University, Xi'an 710127, China;

2. Xi'an Siyuan University, Xi'an 710019, China;

3. School of Information Science and Technology, Northwest University, Xi'an 710127, China;

4. Northwest University-Jingdong Wisdom Cloud Joint Research Center for AI & IoT, Xi'an 710127, China)

Abstract: Scientific and technological resource service is one of the important businesses in the field of data service in cloud computing environment. However, there are still many obstacles to cross-platform service and sharing because scientific and technological resources belong to different units. In order to solve the problems such as resource dispersion and concurrent business, a virtualized scientific and technological resource pools needs to be built based on distributed virtualized storage technology, which can provides unified management, reliable data access and interaction. With the consideration of these cases, we propose a distributed virtualized resource pool model and a technology service framework based on cloud computing theory and hardware/software virtualization technology, which provides a unified description of application service, data collection, virtual and physical layers. As an excellent tool for describing mobile concurrent systems, Pi calculus is employed to describe the external behavior of the proposed scientific and technological resource service model. By analyzing the interaction behavior of the system through Pi deduction rules, the reliability and security of the system is formally verified. The model and the research results can provide valuable theoretical reference for the system design and development.

Key words: distributed system; scientific and technological resource pool; Pi calculus; interact behavior; data collection

0 引言

随着人类社会进入信息化时代,科技信息的数量规模呈现出爆炸式增长,并通过互联网广泛传播。科

技信息资源种类繁多、数量巨大,包括常见的技术专利、科技文献、科学仪器数据、技术报告等,也包括一些行业领域内部的科技资源,比如电力、电信、机械电子、

收稿日期:2020-12-09

修回日期:2021-04-13

基金项目:国家重点研发计划项目(2017YFB1400301)

作者简介:张妮妮(1978-),女,硕士,副教授,CCF会员(20147M),研究方向为大数据与云计算技术。

金融、气象等行业特有的科技资源^[1-2]。为了有效利用这些科技资源,各种科技信息服务平台和资源库随之大量涌现,如国内比较著名的有万方数据、维普期刊、超星数字图书、知网数据等,国外的数字科技资源种类更加丰富,著名的有 Elsevier、Springer、Wiley 在线图书、IEEE 数据库、ACM 数据库等,此外一些专业机构也提供部分科技资源,如国家专利局。相比于传统的纸质文献资料,这些基于云计算和虚拟化技术构建的资源库,可以让人们通过网络快速检索需要的资料,为科学研究和技术开发提供了极其便利的条件。但是,由于科技资源的业务服务比较分散,加之各种科技资源库相互独立,并且科技项目、科技成果、技术专利归属不同部门,造成科技资源条块分割,共享和协同服务难度很大。例如,为了申请一个科研或开发项目,人们需要了解和这个项目相关的科技资料,包括项目相关的中英文文献、专利、相关领域的专家学者的著作等。有时,这些资料没有相应的资料库,人们需要通过一些通用的搜索引擎工具,例如 Google 和百度进行搜索,然后过滤相关性不高的信息,获得有价值的信息;有些资源虽然有专业的资料库,例如常见的 CNKI、IEEE、ACM 等文献资料库,但是,由于这些资料库相互独立,人们需要分别登录访问,查找感兴趣的信息,很不方便,而且要完成以上这些工作,需要消耗人们大量的时间和精力。为了节省查找资料的时间和精力,让科技人员快速准确地找到所需要的信息,构建一个有效的分布式科技资源库和业务服务系统是工业界非常关心的问题,也是学术界面临的一个挑战课题。

基于云计算技术和虚拟化资源池理论,该文提出一种分布式科技资源池服务架构,并从形式化软件开发的角度出发,利用 Pi 演算作为描述工具,讨论这种分布式科技资源池的外部交互行为,给出科技资源库的外部交互行为的描述,以便进一步通过 Pi 演算的规约规则预测系统的行为,为系统平台的设计开发提供理论依据。

1 相关工作

1.1 云计算与虚拟化技术

云计算是近年来信息技术领域出现的一种新技术,它融合了计算机领域中的并行计算、分布式计算、虚拟化技术、网络存储、负载均衡等先进技术,能够高效地管理和调度地理分布广、数据类型多的信息资源,并提供有效的服务,是当前学术界和工业界研究的热点之一^[1-8]。云计算环境中资源管理包括对云环境中的资源组织、数据存取等系统建模。云计算资源调度包括资源调度算法以及相关的资源发现、描述、定位、组织、分配、监测、更新等。Google 公司作为云计算领

域的领导者,拥有最为完整和先进的云计算技术。Google 通过创建多级分布式的数据中心,使用分布式文件系统 GFS (Google File System) 较好地解决了数据的存储和访问难题。Google 的资源调度系统会根据用户的地理位置查询距离当前位置最近的资源并且考虑当前资源的服务能力,选择合适的资源分配给用户使用;Amazon 公司在云计算领域也拥有独特的技术,它把各类资源组建成一个完全分布式、去中心化的云计算平台,其资源调度算法依据用户的特征、需要使用资源的种类、资源数量、资源使用时间等信息为用户合理安排所需要的资源;另一个重要的云计算参与者是 VMware 公司,其主要考虑通过对物理资源的虚拟化来提高资源利用率,VMware 的数据中心虚拟化管理软件可以提供虚拟化基础构架、资源管理和应用程序等多种服务。目前,分布式、去中心化和虚拟化技术为构建云服务基础设施 IaaS 提供了基本的指导原则^[9-11],也是该文构建科技信息服务资源池的主要理论依据。

1.2 Pi 演算理论

Pi 演算理论起源于 20 世纪 80 年代,由图灵奖得主 Robin Milner 参照物理学大统一理论提出^[12-13],用演算中的归约表示由进程间的相互通信形成的动态演化。Pi 演算最初是一种描述和分析通信系统的并发性以及移动性的计算模型,用动态演化结构表示过程间的间歇性相互作用。由于 Pi 演算对于动态并发行为具有很好的描述分析能力,因而被广泛应用于并发系统的分析验证。

关于 Pi 演算的研究工作主要分为两大类:一类是理论研究,另一类是应用研究。其中,早期的工作以理论研究为主^[12-14],国内外许多学者都开展了相关研究,文献[14]用 Pi 演算的通道和表编程概念,分析同构分布式环境下的多任务调度算法;文献[15]则对复杂分支和同步模式进行了 Pi 演算描述和理论分析。Pi 演算的应用研究也十分活跃,文献[16]基于 Petri 网建模跨组织业务过程模型的内部视图,进而使得跨组织业务过程协同的验证转换成 Pi 演算的推演;文献[17]应用 Pi 演算对医院信息系统中的耗材计划管理流程建模,并应用 MWB 对流程进行验证;文献[18]应用 Pi 演算对服务交互流程进行建模描述,但其模型相对比较简单;文献[19]运用 Pi 演算对移动通信服务机制进行研究,通过模型检测验证服务的有效性。这些理论和应用研究工作表明,对于分布式并发系统动态行为的分析描述,Pi 演算仍然是一个非常有效的分析工具。该文正是利用 Pi 演算这种独特优势,描述科技服务资源池的交互行为,进而验证交互服务业务的有效性和可靠性。

2 分布式科技资源池应用架构

在传统中心数据库和 C/S 和 B/S 网络架构中,用户请求服务时,服务器都会分配固定的资源节点。但是,随着用户数量增加,节点的负载能力就会出现不足。而且,一个节点失效就可能停止整个服务。所以,这种中心节点(服务器)网络架构难以适应大规模服务计算的需求。另一方面,考虑到网络中存在大量的空闲节点,其计算资源、存储资源并没有得到充分有效的利用,如果将用户请求的各类资源节点形成一个资源池,在响应用户服务请求时,根据资源池中的资源情况,统一协调地分配给用户,替代传统的中心节点和分散孤立节点的模式,就可以平衡资源节点的负载压力,提高资源的服务效率。

云计算和虚拟化技术为分布式科技资源服务系统构建提供了成熟可靠的技术基础,通过云计算技术可以将不同地域分布的各种科技资源进行统一的管理,而虚拟化技术屏蔽了繁琐复杂的内部访问数据和调用资源细节,为系统提供统一的读写操作、分布式文件存储、弹性资源服务等^[9]。依据云计算思想和虚拟化技术,本文提出了一种分布式科技资源池应用系统架构,如图1所示,自下而上描述为四层模型:物理层、虚拟(资源池)层、汇聚层、应用层,下面将对各层功能做详细的描述。

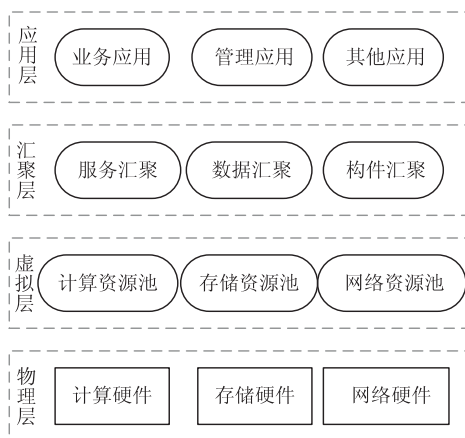


图1 分布式科技资源池应用系统架构

物理层:物理资源层是科技资源服务管理系统的基础设施层,是承载各种科技资源的物理实体,由分布在不同地区、不同规格的物理设备组成,主要包括计算机、存储设备、网络设备以及基础软件等。

虚拟层:实际上也是资源池层,通过采用虚拟化技术,将各种资源抽象为一个容量的池化模型,资源的分布与组织相对上层是透明的,资源池按需为服务分配所需的资源。

汇聚层:为上层用户提供一个高效的资源汇聚平台,汇聚层通过判断用户所需资源特性选择相应的汇聚策略,执行相关的汇聚算法,可以实现服务资源汇

聚、数据汇聚和软件构件的汇聚。

应用层:主要面向科技资源服务的应用需求,通过虚拟的云计算门户对外界提供各种业务应用、管理应用和其他交互应用等。

3 资源池数据汇聚行为分析

在分布式科技资源池应用系统中,汇聚层是整个系统的关键环节,资源的汇聚是最重要的任务,其中数据汇聚是最复杂的工作。因为在跨区域的分布式资源汇聚时,既要考虑资源的空间分布关系,还要考虑资源的时间并发关系。为了达到时间空间有效汇聚,需要对汇聚过程和方法进行描述分析,获得理论上的性能验证。形式化方法通常具有良好的定义和表达,是模型验证的首选方法。因此,该文采用 Pi 演算这一形式化工具,对科技服务资源池的数据汇聚行为进行分析描述。首先给出 Pi 演算的定义,然后模型描述验证数据汇聚方法在时间空间上的合理性和有效性。

3.1 Pi 演算的代数定义

Pi 演算是图灵奖得主 Milner 在 20 世纪 90 年代提出的一种进程演算,具备完备的数学体系,对并行行为具有很强的建模表达能力。在 Pi 演算理论中,名称是基本的概念,值、变量和管道通过名称引用;管道不仅可以传递信息,也可以将管道名作为一种信息传递。图2是 Pi 演算的一种基本图形表示,其中 P 、 Q 表示过程, \bar{b} 表示端口(一个过程可有若干端口), b 表示动作, \rightarrow 表示管道。在图中,过程 P 通过端口 \bar{b} 沿管道发送消息,过程 Q 通过管道接收消息。

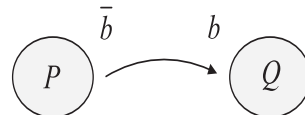


图2 Pi 演算的图形表示

定义:设 N 是一个表示名称的可数无限集合,用 a, b, x, y, \dots 表示 N 中的元素; A 是一个代理(或过程)的集合,用 P, Q, R 表示 A 中的元素,则 Pi 演算的语法可定义如下:

$$P ::= 0 \mid \bar{x}y. P \mid x(y). P \mid \tau. P \mid (y)P \mid P \mid Q \mid P + Q \mid A(x_1, x_2, \dots, x_n) \mid \mid P$$

这里, $x(y). P$ 和 $(y)P$ 中的 y 被称为约束名称, P 是它们的辖域。在 P 中出现的非约束名称称为自由名称, P 中的自由名称的集合表示为 $fn(P)$, 约束名称的集合表示为 $bn(P)$, P 中名称的集合用 $n(P)$ 表示。

公式中操作符的含义如下:

- (1) 0 表示这个过程不做任何动作;
- (2) $\bar{x}y. P$ 表示名称 x 发送名称 y 后,开始执行 P ;
- (3) $x(y). P$ 表示名称 x 收到任何名称后,执行 P 并用收到的名称替换 y ;

(4) $\tau.P$ 表示从 $\tau.P$ 不可见的演化到 P , 即可认为 τ 是一个内部动作;

(5) $(y)P$ 表示名称 y 的作用范围是 P ;

(6) $P \mid Q$ 表示 P 和 Q 独立执行并且可通过共有的名称相互通信;

(7) $P + Q$ 表示 P 和 Q 只能有一个执行;

(8) $A(x_1, x_2, \dots, x_n)$ 表示过程 A 中的自由名称;

(9) $!P$ 表示可无限重复执行 P , 即认为 $!P = P \mid P \mid P \dots$ 或 $!P = P \mid !P$ 。

3.2 数据汇聚行为的 Pi 演算描述

在描述数据的汇聚行为时,主要考虑从已经建立好的独立的科技资源库中提取数据进行集成,其业务流程框架如图 3 所示。在图 3 中,空间分布的独立资源节点由资源汇聚调度中心统一管理,数据资源需要经过前期清洗之后才能进入虚拟资源池,以保证较低的数据重复率和错误率,资源汇聚过程既要考虑空间高效性,也要考虑时间准确性。用户对资源的访问可以是直接的,也可以是间接的资源链接地址,资源调度算法会给出一个优化的资源服务方案。

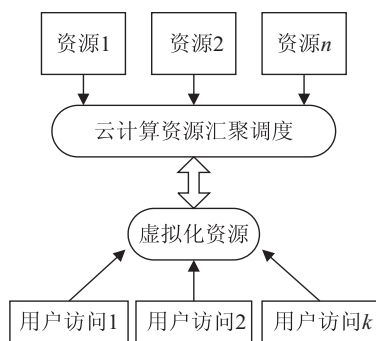


图 3 数据资源汇聚基本框架

假设目前已经有 n 个独立的科技资源库,分别用 P_1, P_2, \dots, P_n 表示,即把每一个独立资源库看做一个进程,则 P_1, P_2, \dots, P_n 的 Pi 演算描述如下:

$$P_1(t_1, c_1) = \bar{x}. P_2(t_2, c_2)$$

...

$$P_{n-1}(t_{n-1}, c_{n-1}) = \bar{x}. P_n(t_n, c_n)$$

$$P_n(t_n, c_n) = \bar{x}. 0$$

其中, t_i, c_i 表示从时刻 c_i 开始的一段时间 t_i ,也就是当从时刻 c_i 开始在时间段 t_i 产生新数据时,进程 P_i 发起和进程 Q 的通信(Q 是集成的资源池),并把时间段 t_i 内的数据汇聚到 Q 中,然后把通信权转入下一个进程。

集成后的资源库也称为资源池,用 Q 表示,把集成后的资源池 Q 也看作一个进程,则 Q 的 Pi 演算描述如下:

$$Q = x. Q$$

最后,集成资源池与独立资源库交互行为可用 Pi

演算描述为:

$$S = Q \mid P$$

其中, $P = \bar{x}. P_1(t_1, c_1)$ 。

则系统的交互行为利用 Pi 演算的规约规则可推演如下:

$$S = Q \mid P = x. Q \mid \bar{x}. P_1(t_1, c_1) = Q \mid P_1(t_1, c_1) =$$

$$x. Q \mid \bar{x}. P_2(t_2, c_2) = P_2(t_2, c_2) = \dots =$$

$$Q \mid P_n(t_n, c_n) = x. Q \mid \bar{x}. 0 = Q \mid 0.$$

3.3 客户访问集成资源的 Pi 演算描述

客户访问数据是资源池服务的基本交互业务。服务过程通常由用户进程发起请求,供应商服务进程提供响应,业务流程如图 4 所示,图中箭头代表了信息的流向。为了保证不同类型数据传输的正确性,一般可以开辟不同的传输通道,图中把发送通道和接收通道分开表示。以上资源访问交互活动可以用 Pi 演算描述如下:

$$I = UP \mid SP$$

$$UP = \overline{\text{req}} \langle \text{msg} \rangle . \text{res}(\text{msg}). UP$$

$$SP = \text{req}(\text{msg}). \tau. \text{res} \langle \text{msg} \rangle . SP$$

这里, UP 是用户进程, SP 是服务商进程, res 和 req 是逻辑独立的用户进程通道和服务进程通道, msg 代表产生的消息, $\langle \rangle$ 表示有方向地传递消息。

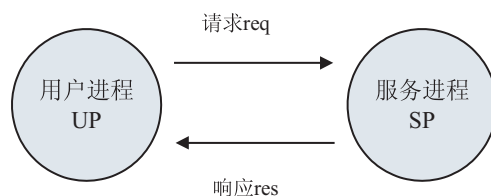


图 4 基本资源池服务交互流程

在构建和集成分布式科技资源池服务系统时,考虑到系统的负载及用户的访问速度,需要分别在不同区域建立集成资源的备份。当用户访问中心资源库(资源池)时,可以根据用户的位置把用户的访问定位到不同区域的备份上,从而提高服务效率。例如,当客户进程访问中心资源库 Q 时, Q 可以把 Q_1 的连接发给客户进程,让客户进程访问 Q_1 ,如图 5 所示,也就是需要把图 5(a) 的结构变为图 5(b) 的结构。在图 5(a) 中,进程 Q 和 Q_1 之间的连接是私有的,也就是 Q 和 Q_1 之间的连接名称的作用范围只限于 Q 和 Q_1 ,而进程 C 一开始并不知道这个链接。可以把这个链接名考虑为 Q_1 的网址,在刚开始,客户进程 C 并不知道进程 Q_1 的网址,但进程 Q 知道,可以让进程 Q 把 Q_1 的网址通过它和 C 的管道发给 C ,这样进程 C 就可以获得 Q_1 的网址,因而进程 C 可以和进程 Q_1 实现通信。

图 5(a) Pi 演算描述为:

$$S = (z) (Q \mid Q_1) \mid C$$

图5(b)Pi演算描述为:

$$S' = Q' \mid C' \mid Q_1$$

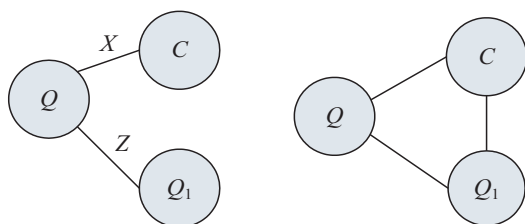
其中, $Q = \bar{x}(z). Q'$, $C = x(y). y. C'$, $Q_1 = \bar{z}. Q_1$ 。

通过Pi演算的演化规则,可以得到:

$$S = (z) (\bar{x}(z). Q' \mid \bar{z}. Q_1) \mid x(y). y. C' =$$

$$Q' \mid \bar{z}. Q_1 \mid z. C' = Q' \mid C' \mid Q_1$$

也就是说,通过Pi演算的演化规则可以从图5(a)到达图5(b),交互过程是可行的,不会出现死锁等问题,以上分析也体现了Pi演算对移动行为的描述能力。



(a) 初始访问结构

(b) 变迁访问结构

图5 客户访问中心数据库时的变迁图

4 结束语

针对科技资源服务业务需求,提出了一种科技资源池服务系统架构模型,讨论了用Pi演算对数据集成及数据访问的外部交互行为的描述方法。通过对数据访问交互行为的Pi演算描述,说明了访问结构的动态迁移。下一步的工作,可以依据所建立的Pi演算描述,用支持Pi演算规约规则的工具进一步验证交互行为中的一些特性。例如通过数据集成模式外部交互行为的描述,进一步使用Pi演算的动态演化规则,可以判断中心数据库能否在某段时间内保证对所有的独立数据源的访问,并且在某一个时刻只能与一个独立资源交互。通过分析推理可知,Pi演算作为一种过程代数,非常适合描述系统的并发行为及动态迁移行为。在实际应用中,当建立具有并发性以及动态拓扑结构的科技信息服务系统时,在系统投入使用前,用Pi演算对其行为进行描述并分析,有助于在系统开发初期及早发现其中存在的问题。

参考文献:

[1] 钱琼芬,李春林,张小庆,等.云数据中心虚拟资源管理研究综述[J].计算机应用研究,2012,29(7):2411-2415.

[2] 聂元丁.基于云计算的国家级气象资源池设计与建设[J].计算机技术与发展,2018,28(12):132-136.

[3] 苏朋程.云计算研究虚拟化技术的研究[J].计算机技术与发展,2017,27(4):29-33.

[4] LEAVITT N. Is cloud computing really ready for prime time? [J]. Computer,2009,42(1):15-20.

[5] 张建勋,古志民,郑超.云计算研究进展综述[J].计算机应用研究,2010,27(2):429-433.

[6] 孙大为,常桂然,陈东,等.云计算环境中绿色服务级目标的分析、量化、建模及评价[J].计算机学报,2013,36(7):1509-1525.

[7] 崔嘉.试析虚拟计算环境中资源池的资源聚合机制的研究[J].自动化技术与应用,2017,36(6):35-38.

[8] 文静,王明哲,郭法滨.面向云计算资源管理的交互仿真平台构建[J].计算机与数字工程,2013,41(5):753-756.

[9] 王健伟.一种全新的IT建设架构—统一资源池[J].数据通信,2011(5):10-12.

[10] 施其淮.探究云计算资源池虚拟化技术及实现方式[J].通讯世界,2016(16):15-16.

[11] LOPES R, BRASILEIRO F, MACIEL P D. Business-driven capacity planning of a cloud-based IT infrastructure for the execution of web applications[C]//IEEE international symposium on parallel & distributed processing, workshops and phd forum. Atlanta, Georgia: IEEE, 2010:1-8.

[12] MILNER R, PARROW J, WALKER D J. A calculus of mobile process. Part I[R]. Edinburgh: Edinburgh University, 1989.

[13] MILNER R, PARROW J, WALKER D J. A calculus of mobile process. Part II[R]. Edinburgh: Edinburgh University, 1989.

[14] 陈海鹏,邱际伦,邱峰,等.基于Pi演算的前沿调度算法并行性建模及实现[J].吉林大学学报:工学版,2019,49(1):242-247.

[15] 郭小群,郝克刚,侯红,等.复杂分支和同步模式的Pi演算描述[J].计算机科学,2010,37(2):139-140.

[16] 笪建,莫启,李彤,等.跨组织业务过程协同多视图模型[J].计算机应用研究,2015,32(9):2687-2692.

[17] 刘峰,王剑.适用于医院信息管理系统的Pi演算的业务流程建模[J].信息技术与信息化,2018(8):22-23.

[18] 刘亚亚,姜久雷.利用Pi演算实现对服务交互流程的设计[J].数字技术与应用,2016,(4):166.

[19] 付玉.基于PI演算的移动通信服务的建模与验证[D].长春:吉林大学,2008.