

# Web 服务组合性能评价的流逼近与扩散逼近研究

蹇志勇, 樊瑞娜

(燕山大学 经济管理学院, 河北 秦皇岛 066044)

**摘要:** Web 服务组合系统的规模尺度和复杂程度的不断提高以及调用请求的周期化和一般化, 使得 Web 服务组合系统的性能建模和分析更加困难。针对该问题, 建立了一组把服务组合的基本交互关系和 Web 服务器处理能力、负载等因素映射到开排队网络的映射规则。提出了一种基于逼近理论与排队模型相结合的分析方法, 克服了以往研究中调用请求到达过程和服务过程受限的情况, 得到了系统关键性能指标的流逼近与扩散逼近定理。最后, 分析了系统的性能及其变化规律。本文的结果和方法为研究更加一般的大型 Web 服务组合系统提供了一个新的途径。

**关键词:** Web 服务组合; 排队网络; 流极限; 扩散极限; 性能分析

**中图分类号:** TP302.7 **文献标识码:** A

## Research on Fluid Limit and Diffusion Limit for Performance

### Evaluation of Web Service Composition

Zhi-Yong Qian, Rui-Na Fan

(School of Economics & Management Sciences, Yanshan University, Qinhuangdao, Hebei, 066044)

**Abstract:** With the improvement of the Web service composition systems' size and complexity, the periodism and generalization of invoking requirements, the modeling and analyzing of systems become more difficult. To tackle this problem, a set of mapping rules was established to transform the interaction relations and factors including the processing capabilities and the load of web servers into open queueing network. This paper developed a new analytic technique based on approximation theories and queueing networks, overcame the restriction of arrival processes and service processes for invoking requirements in past studies and get the fluid and diffusion limit theorems of some key performance indexes. At last, this paper analyzed the performance and its change rules of systems. The results and methodology of this paper provide new highlight in the study of more general large-scale Web service composition systems.

**Keywords:** Web service composition; queueing networks; fluid limit; diffusion limit; performance analysis

---

**作者简介:** 蹇志勇 (1990-), 男, 河北邢台人, 硕士研究生, 研究方向为服务计算、排队网络、流逼近、扩散逼近, (qianzhiyong0926@stumail.ysu.edu.cn); 樊瑞娜 (1989-), 女, 河南安阳人, 博士研究生, 研究方向排队论、网络资源管理、大型服务系统。

## 0 引言

随着互联网技术的快速发展和企业业务流程复杂程度的不断提高,面向服务的计算<sup>[1][2]</sup>作为一种新型的分布式构造和计算模式应运而生。服务组合<sup>[2]</sup>作为服务计算研究中的一个十分重要分支,一方面是由于用户对服务质量<sup>[3]</sup>要求越来越高;另一方面人们也希望将若干个 Web 服务通过不同的交互过程有机的组合起来到达服务价值增值的目的。在满足用户对功能需求的前提下,性能是 Web 服务组合取得成功赢得用户的关键,因此大型 Web 服务组合系统的性能分析对研究、构建和优化实际 Web 服务组合系统具有非常重要的意义。然而随着服务系统的规模化尺度和复杂程度的不断提高,用户对 Web 服务调用的周期化和一般化,以及 Web 服务组合系统的性能受到 BPEL 流程的结构、用户的选择概率分布、服务器的负载和处理能力等因素的影响,使得服务组合系统的性能建模和分析越来越困难。因此,构建一套有效的服务组合性能评价方法,具有十分重要的理论意义与实际应用价值,同时也是国内外迫切需要解决的重要研究课题。

随机模型作为研究复杂网络性能的一个有效方法受到了广泛的认可,但是到目前为止,对 Web 服务组合性能进行随机模型分析的文献还比较少,以往的研究也主要是从 Petri 网、Markov 过程、进程代数 and 排队论几方面进行着手分析。其中,应用具有时间约束的 Petri 网<sup>[4-8]</sup>(随机 Petri 网<sup>[4][5]</sup>、广义随机 Petri 网<sup>[6][7]</sup>、模糊 Petri 网<sup>[8][9]</sup>、Markov 链<sup>[10-13]</sup>(其中包括将进程代数与 Markov 链相结合见肖芳雄等人<sup>[12][13]</sup>的研究)来研究服务组合性能,对完善服务组合的交互关系与规则、推进服务组合建模及理论研究做了很大的贡献。但这些研究都假定要调用的 Web 服务单元处于“随叫随到”状态,没能建模调用请求在服务器端排队等待的现象,而这种排队现象在计算机网络传输、CPU 执行过程中是普遍存在的;其次,它们建模了服务器处理能力对性能的影响,没有建模服务组合中各服务单元本身的负载对组合系统性能产生的影响。排队论可以很好的处理上面所面临的问题,对 Web 服务组合的真实性能进行更好的建模分析。到目前为止,主要是用排队论建模分析单个 Web 服务性能见 Xiong<sup>[14]</sup>, Gao 等人<sup>[15]</sup>和 Qu 等人<sup>[16]</sup>的研究。用于建模 Web 服务组合系统性能的文献还不是很多:顾军等人<sup>[17]</sup>将排队论与 Petri 网相结合,把服务组合五种交互关系映射到排队 Petri 网上,建模了队列库所的排队模

型,并利用 QPMEF 模拟工具分析了节点失效和恢复机制对服务组合性能的影响;Das 等人<sup>[18]</sup>和 Xie 等人<sup>[19]</sup>利用排队网络对多层 Web 服务系统的可靠性和在失效影响下的基于乘积网络的多层 Web 服务系统性能进行了分析;Zhou 等人<sup>[20]</sup>利用节点为 M/M/1 的排队网络模型分析了服务组合的响应时间,队长等指标,并通过蒙特卡罗模拟对航空订票系统进行实证分析验证了模型的高效性;汪浩等人<sup>[21]</sup>将 BPEL 流程结构和影响 Web 服务性能的内因和外因映射到排队网络上,并分析了 Web 服务响应时间和服务调用请求平均排队数量两个性能指标变化。然而,应用排队理论建模时,以往的研究在调用请求到达过程或处理过程的时间分布上受到指数分布的束缚,调用请求到达过程的非泊松化和处理时间的一般化是实际 Web 服务组合系统的最主要特征,同时也是理论研究的难点所在。

流逼近和扩散逼近在分析大型复杂网络问题上得到了广泛的应用。其相关理论可参考 Harrison<sup>[22][23]</sup>、Chen<sup>[24]</sup>和 Dai 等人<sup>[25][26]</sup>的研究。本文基于开排队网络理论对 Web 服务组合系统的关键性能指标进行了流逼近与扩散逼近分析,得到了调用请求数量和响应时间的变尺度过程在分布上收敛于反射布朗运动,克服了用户调用请求到达过程和服务过程受限的情况,建立了严格的分析框架。这对分析预测 Web 服务组合在互联网环境下的性能和优化大型 Web 服务组合系统,确定其瓶颈节点具有深刻的指导意义。

本文的学术贡献:第一,本文建立一套系统的服务组合交互关系和规则,拓宽了以往研究中用户调用请求到达时间间隔分布和 Web 服务器处理时间分布受限的情况,更具一般性;第二,建立了 Web 服务组合系统在稳态运行状态下的流逼近定理,刻画了关键性能指标的一阶线性近似。第三,在流逼近的基础上应用函数中心极限定理刻画了关键性能指标的扩散逼近定理,给出了关键参数的计算公式。基于此,本文首次将流逼近和扩散逼近理论应用到 Web 服务组合性能分析上,给出了一组建立在开排队网络基础上的性能分析指标体系及其相应的扩散逼近计算公式。本文所给出的分析方法和结果可以应用到研究更加一般的、复杂的、大型 Web 服务组合系统分析中。

本文的组织如下:第 1 节,描述了一个大型的 Web 服务组合系统,并给出了其相应的开排队网络

映射规则。第2节,给出了Web服务组合系统中各节点的联合排队过程、建立了节点的流平衡等式。第3节,给出了Web服务组合性能指标的流逼近定理。第4节,给出Web服务组合性能指标的扩散逼近定理。第5节,性能指标计算及数值算例。第6节,总结全文并做出评论。

## 1 模型描述

Web服务组合是将若干个Web服务单元通过不同的服务执行顺序和交互过程有机的组合起来,形成一个具有大粒度的、支持业务流程逻辑的全新的Web服务组合过程。其可以用集合关系进行表示如图1所示,

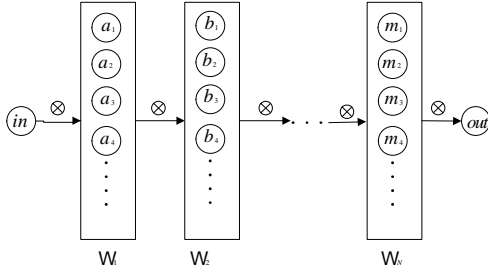


图1: Web服务功能集合图

令  $W_1 = \{a_1, a_2, \dots, a_n\}$  表示具有相同或相似功能  $a$  的Web服务单元的集合,同理,  $W_2, K, W_N$  具有相同的表示意义。 $in$  表示服务组合调用的起点;  $out$  表示调用完成的终点。令  $A_1 \hat{=} W_1$  表示集合  $W_1$  中的一个Web服务单元,同理有  $A_2 \hat{=} W_2, K, A_N \hat{=} W_N$ 。那么由不同功能的Web服务单元组合的全新的业务流程  $A = \{A_1, A_2, \dots, A_N\}$ , 功能集与功能集之间又存在着不同的交互关系。接下来我们将Web服务组合的运行机制、系统参数和交互关系表示如下:

- (1) Web服务单元: 设一个大型Web服务组合系统中有  $J$  个功能不同的Web服务单元组成,并且每个Web服务单元即可以处理正在对其进行性能分析的Web服务组合调用请求,也可以处理只对本单元的调用请求。
- (2) Web服务调用请求到达过程: Web服务组合调用请求和对第  $i$  个Web服务单元调用请求的到达时间间隔为一般分布。令  $u_0 = \{u_0(n), n \geq 1\}$  和  $u_i = \{u_i(n), n \geq 1\}$  分别表示一个独立同分布非负随机变量序列,  $u_0(i)$  表示第  $i$  个服务组合调用请求与第  $i-1$  个调用请求的到达时间间隔;  $u_i(l)$  表示第  $l$  个服务单元调用请求与第  $l-1$  个服务单元调用请求到达第  $i$  个Web服务单元的到达时间间隔。并且假设  $u_0$  的均值为  $\frac{1}{\lambda_0}$ , 均方差变异系数为  $c_{0,0}^2$ ;  $u_i$  的均值为  $\frac{1}{\lambda_i}$ , 均方差变异系数为  $c_{a,i}^2$ 。
- (3) Web服务单元的服务过程: 我们设同一个Web服务单元对服务组合调用请求的处理时间和对服务单元调用请求的处理时间是相同的,其处理时间服从一般分布。令  $v_i = \{v_i(n), n \geq 1\}$  表示一个独立同分布非负随机变量序列,  $v_i(l)$

表示Web服务单元  $i$  对第  $l$  个调用请求的处理时间。并且假设  $v_i$  的均值为  $m_i = \frac{1}{\mu_i}$ , 均方差变异系数为  $c_{s,i}^2$ 。

- (4) Web服务组合调用请求传输过程: 我们设Web服务组合调用请求在第  $i$  个服务单元处理完成后以概率  $0 < p_{ij} \leq 1$  进入第  $j$  个服务单元; 对第  $i$  个Web服务单元调用请求经本服务单元处理完成后直接离开系统, 因此有  $p_{i0} = 1$ 。令  $r^i = \{r^i(n), n \geq 1\}$  表示一个独立同分布的随机变量序列, 值集为  $\{e^j, K, e^j\}$  (其中  $e^j$  表示  $\hat{A}^j$  中第  $j$  个单位向量), 事件  $P_r\{r^i(l) = e^j\} = p_{ij}$  表示第  $l$  个服务组合调用请求经第  $i$  个Web服务单元处理完成后以概率  $p_{ij}$  进入到第  $j$  个Web服务单元继续接受处理。
- (5) Web服务调用请求离开原则: 我们设Web服务组合调用请求只有在接受完整个组合业务流程中相应的Web服务单元服务后才可离开Web组合系统; Web服务单元调用请求在经相应的Web服务单元处理后直接离开此组合系统。

Web服务组合系统复杂性行为主要是由各个功能集合之间的交互关系的复杂性引起的。开排队网络能够很好建模这种大型服务组合系统。本文将Web服务组合系统中服务单元的先后调用顺序映射到开排队网络中各个节点之间的拓扑次序, 系统地定义了6种基本交互关系, 并建立了其相应于开排队网络的映射规则, 如下:

- (1) 顺序关系映射规则: 只有当前驱服务单元  $S_i$  完成服务后, 后继服务单元  $S_j$  才能开始服务。其顺序关系映射为排队网络如图2所示:

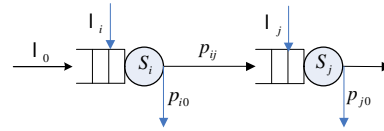


图2 顺序关系映射为排队网络

- (2) 条件关系映射规则: 当前驱服务单元  $S_i$  完成服务后, 会以相应的概率进入相应的后继服务单元  $S_j$  继续接受服务。其条件关系映射为排队网络如图3所示:

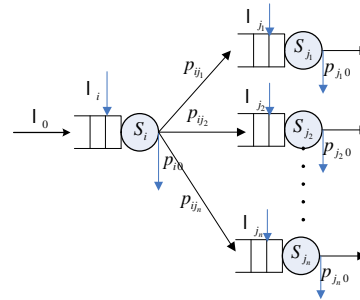


图3 条件关系映射为排队网络

- (3) 循环关系映射规则：服务单元  $S_i$  在循环条件为真时，重复执行。其循环关系映射为一个反馈排队网络如图4所示：

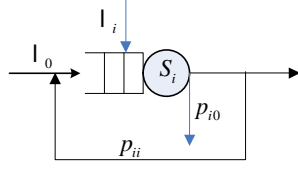


图4 循环关系映射为排队网络

- (4) 并发关系映射规则：服务单元  $S_i$  完成后并发调用其后继 Web 服务单元。其并发关系映射为一个 Fork-Join 排队网络中的 Fork 节点如图5所示：

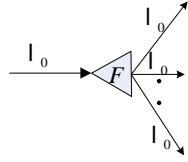


图5 并发关系映射为排队网络

- (5) 并行关系映射规则：Web 服务  $S_{j1}, K, S_{jm}$  相互独立并行执行，执行完成后其可以执行自己的后继服务，最终进入相应的等待队列节点  $q_{ji}$ ，等待其他并行分支的执行完成。其并行关系映射为一个并行的排队网络如图6所示：

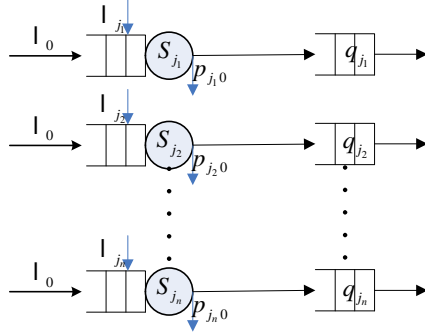


图6 并行关系映射为排队网络

- (6) 打包关系映射规则：Web 服务  $S_{j1}, K, S_{jm}$  执行完成后打包合并成一个新的任务。其打包关系映射为一个 Fork-Join 排队网络中的 Join 节点如图7所示：

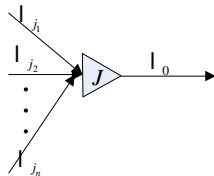


图7 打包关系映射为排队网络

根据上文所定义服务组合基本交互关系以及 Web 服务组合业务流程的多样化，可知服务组合形式有多种。我们给出一种典型的计划旅行过程的业务流程实例如图8所示及其映射的排队网络模型如图9所示：

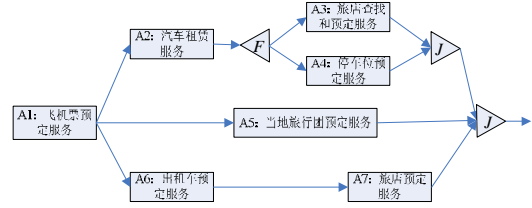


图8 旅客计划旅行过程的业务流程实例

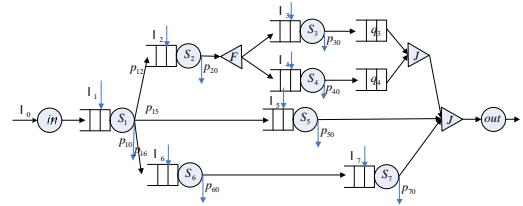


图9 计划旅行 Web 服务组合相应的开排队网络

由上可知，我们可以把 Web 服务组合系统建模成一个开排队网络，其系统参数描述如下：

- (1) 虚拟节点：虽然不同的 Web 服务单元具有不同的功能，我们在排队网络中抽象成相同的节点  $S_i (i=1, K, J)$ 。
- (2) 虚拟顾客：由于同一个 Web 服务单元对服务组合的调用请求和对服务单元调用请求的处理时间相同，因此我们将 Web 服务调用请求看成是同一类顾客。
- (3) 服务规则：排队网络中各个节点的服务规则为先到先服务(FCFS)。
- (4) 等待空间容量：排队网络中各个节点的等待空间容量无限大。
- (5) 到达率：输入节点的到达率为  $l_0$ ，网络中各个节点的外部到达率为  $l_i$ 。
- (6) 服务率：网络中各个节点的服务率为  $\eta$ 。
- (7) 路径矩阵：路径矩阵  $P$  的第  $i$  行第  $j$  列的组成部分为  $p_{ij}$ ，并且有： $\sum_{j=1}^J p_{ij} = 1$ 。由于服务单元调用请求在本 Web 服务单元处理完成后直接离开系统，因此  $p_{i0} = 1 (i=1, K, J)$ 。

## 2 排队网络中各个节点的联合排队过程

根据上文描述的 Web 服务组合运行参数和相应的开排队网络的相关参数，我们建立一个排队网络的顺向结构，其中各个节点平均顾客数量和平均响应时间来近似各个 Web 服务单元中 Web 服务调用请求的平均数量和 Web 服务组合的响应时间。首先，令  $\mathbf{e} = \{j: l_j > 0\}$  表示有 Web 服务单元调用请求进入的节点集合； $Q_j(0)$  是一个非负的整数随机变量代表服务台  $j (j=1, K, J)$  在  $t=0$  时刻拥有的 Web 服务单元调用请求数量；对于  $j \in \mathbf{e} \cup \{0\}$ ，令  $A_j = \{A_j(t) \geq 0\}$  表示第  $j$  个 Web 服务单元的到达更

新过程，涉及到的到达序列  $u_j$ ：  
 $A_j(t) = \sup\{n: U_j(n) \leq t\}$ ,  $U_j(n) = \sum_{l=1}^n u_j(l)$ ,  $n \geq 1$  和  $U_j(0) = 0$ ，在此我们说明，当  $j=0$  时  $A_0$  表示 Web 服务组合调用请求到达输入点的更新过程。对于  $j=1, \mathbf{K}, J$ ，令  $S_j = \{S_j(t) \geq 0\}$  表示第  $j$  个 Web 服务单元的服务更新过程，涉及到服务序列  $v_j$ ：  
 $S_j(t) = \sup\{n: V_j(n) \leq t\}$ ,  $V_j(n) = \sum_{l=1}^n v_j(l)$ ,  $n \geq 1$  和  $V_j(0) = 0$ ；令  $R^j = \{R^j(n) \geq 0\}$  表示第  $j$  个 Web 服务单元的路径过程，涉及到转移序列  $r^j$ ：  
 $R^j(n) = \sum_{l=1}^n r^j(l)$ ,  $n \geq 1$  和  $R^j(0) = 0$ ,  $R^j$  的第  $i$  个组成部分为  $R_i^j$ 。我们假设  $Q(0) \geq 0$  ( $Q_j(0)$ ),  $A_j$ ,  $S_j$  和  $R^j$  定义在相同的概率空间上，并且相互独立。

简而言之，一个由 Web 服务组合系统所抽象成的排队网络的初始数据主要有：到达过程  $A_j$ ，服务过程  $S_j$ ，路径选择过程  $R^j$  和初始的队长  $Q_j(0)$ 。从这些数据中，我们就可以构建出一些测量 Web 服务组合性能的重要指标：平均队长过程： $Q = \{Q(t), t \geq 0\}$ ，等待时间过程： $Z = \{Z(t), t \geq 0\}$ ，和忙期过程： $B = \{B(t), t \geq 0\}$ 。 $Q(t)$  的第  $j$  个组成部分  $Q_j(t)$  表示 Web 服务单元  $j$  在时刻  $t$  时 Web 服务单元调用请求数（包括正在接受服务的和等待接受服务的）； $Z(t)$  的第  $j$  个组成部分  $Z_j(t)$  表示在时刻  $t$  时完成当下在 Web 服务单元  $j$  的所有调用请求所需要的总时间； $B(t)$  的第  $j$  个组成部分  $B_j(t)$  表示 Web 服务单元  $j$  处于工作状态的累计时间。

从上面的描述中我们可知  $A_0(t)$  表示在  $[0, t]$  内到达的服务组合调用请求数； $A_j(t)$  表示的是在  $[0, t]$  内到达第  $j$  个 Web 服务单元的服务单元调用请求数； $S_j(B_j(t))$  表示在  $[0, t]$  内服务单元  $j$  总共处理完的调用请求数量； $R_i^j(S_j(B_j(t)))$  表示在  $[0, t]$  内从服务单元  $j$  转到服务单元  $i$  的服务组合调用请求数量。因此我们可以得到如下的流平衡关系式：对于起始 Web 服务单元  $j=1$ ,

$$Q_1(t) = Q_1(0) + A_0(t) + A_1(t) + \sum_{i=1}^J R_i^1(S_1(B_1(t))) - S_1(B_1(t)). \quad (1)$$

对于 Web 服务单元  $j=2, \mathbf{K}, J$ ,

$$Q_j(t) = Q_j(0) + A_j(t) + \sum_{i=1}^J R_i^j(S_j(B_j(t))) - S_j(B_j(t)). \quad (2)$$

对所有  $t \geq 0$  和  $j=1, \mathbf{K}, J$ ，我们有：

$$B_j(t) = \int_0^t \mathbf{1}\{Q_j(s) > 0\} ds. \quad (3)$$

进一步，我们构建等待时间过程  $Z$ ，因为  $V_j(n)$  表示 Web 服务单元  $j$  在服务完前  $n$  项调用请求所需要全部处理时间。因此可得：对于初始 Web 服务单元  $j=1$ ,

$$\begin{aligned} Z_j(t) &= V_j(Q_j(0) + A_0(t) + A_j(t)) \\ &+ \sum_{i=1}^J R_i^j(S_i(B_i(t))) - B_j(t) \\ &= V_j(Q_j(t) + S_j(B_j(t))) - B_j(t). \end{aligned} \quad (4)$$

对于 Web 服务单元  $j=2, \mathbf{K}, J$ ,

$$\begin{aligned} Z_j(t) &= V_j(Q_j(0) + A_j(t) + \sum_{i=1}^J R_i^j(S_i(B_i(t)))) - B_j(t) \\ &= V_j(Q_j(t) + S_j(B_j(t))) - B_j(t). \end{aligned} \quad (5)$$

另一个经常用到的是空闲时间过程  $I = \{I(t), t \geq 0\}$ ，其第  $j$  个组成部分可以如下：

$$I_j(t) = t - B_j(t) = \int_0^t \mathbf{1}\{Q_j(s) = 0\} ds. \quad (6)$$

对上式中的队长过程做“中心化处理”，把式(1)和(2)重写如下形式：

$$Q(t) = X(t) + (I - P\Phi Y)(t).$$

对于初始 Web 服务单元  $j=1$ ， $X(t)$  的第  $j$  个组成部分如下：

$$\begin{aligned} X_j(t) &= Q_j(0) + (1_0 + 1_j + \sum_{i=1}^J m_j p_{ij} - m_j)t \\ &+ [A_0(t) - 1_0 t] + \sum_{i=1}^J p_{ij} [S_i(B_i(t)) - m_j B_i(t)] \\ &+ [A_j(t) - 1_j t] - [S_j(B_j(t)) - m_j B_j(t)] \\ &+ \sum_{i=1}^J [R_i^j(S_i(B_i(t))) - p_{ij} S_i(B_i(t))]. \end{aligned} \quad (7)$$

对于 Web 服务单元  $j=2, \mathbf{K}, J$ ， $X(t)$  的第  $j$  个组成部分如下：

$$\begin{aligned} X_j(t) &= Q_j(0) + (1_j + \sum_{i=1}^J m_j p_{ij} - m_j)t + [A_j(t) - 1_j t] \\ &+ \sum_{i=1}^J p_{ij} [S_i(B_i(t)) - m_j B_i(t)] - [S_j(B_j(t)) - m_j B_j(t)] \\ &+ \sum_{i=1}^J [R_i^j(S_i(B_i(t))) - p_{ij} S_i(B_i(t))]. \end{aligned} \quad (8)$$

相应的  $Y(t)$  的第  $j=1, \mathbf{K}, J$  个组成部分如下：

$$Y_j(t) = m_j I_j(t) = m_j \int_0^t \mathbf{1}\{Q_j(s) = 0\} ds. \quad (9)$$

从上面的等式和相应的物理意义中我们可以发现，对于所有  $t \geq 0$ ， $j=1, \mathbf{K}, J$ ,

$$Q(t) \geq 0, \quad (10)$$

$$dY(t) \geq 0, \quad Y(0) = 0, \quad (11)$$

$$Q(t) dY(t) = 0. \quad (12)$$

**引理 1** (斜反射映射定理) 假设  $J' \times J$  的矩阵  $R$  是一个  $S$  矩阵，那么对于任何一个  $x \in D_0' = \{x \in D' : x(0) \geq 0\}$ ，在  $D^{2J}$  中存在一个唯一的  $(y, z)$  满足：

$$\begin{aligned} z &= x + R y \geq 0, \\ dy &\geq 0, \text{ 和 } y(0) = 0, \\ z_j dy_j &= 0, \quad j=1, \mathbf{K}, J. \end{aligned}$$

其中，过程  $z = F(x)$  和过程  $y = Y(x)$  被称为  $x$  的映射过程和  $x$  的调节过程， $(Y, F)$  依赖于矩阵  $R$ 。

引理 1 是文献[24]中定理 6.1 的转述，具体的证明部分可参考文献[24]的相应部分。由引理 1 我们可知关系式 (10-12) 唯一确定一对从  $X$  到  $Q := F(X)$  和  $Y := Y(X)$  的映射，并且，对于任意一

个固定  $t$ ，映射  $F$  和  $Y$  在  $D_0^J[0, t]$  是 Lipschitz 连续的。在本节的结束部分我们来证明矩阵  $R = I - P\Phi$  是一个  $S$  矩阵。首先我们给出  $S$  矩阵的定义，

**定义 1** 对于一个  $J' \times J$  的方阵  $R$ ，如果存在一个  $J$  维向量  $u \geq 0$  使得  $Ru > 0$ ，我们称方阵  $R$  是一个  $S$  矩阵。

**定理 1** 反射矩阵  $R = I - P\Phi$  是一个  $S$  矩阵。

**证明** 首先我们知道  $R$  是一个  $J' \times J$  的方阵，由式 (7-10) 我们可知

$$R_{ij} = \begin{cases} 1 - p_{ii} & i = j, \\ -p_{ij} & i \neq j, s(j) = i, \\ 0 & i \neq j, s(j) \neq i. \end{cases}$$

其中， $s(i) = j$  表示服务调用请求在第  $i$  个 Web 服务单元处理完成后进入第  $j$  个 Web 服务单元接受处理。由上节的反馈映射关系和路径矩阵规则可知  $0 \leq p_{ii} < 1$  和  $0 < p_{ij} \leq 1$ ，因此矩阵  $R$  是对角线元素由非负元素组成、非对角矩阵由非正元素组成的方阵，因此存在向量  $u \geq 0$  使得  $Ru > 0$ 。所以  $R$  是一个  $S$  矩阵。

### 3 Web 服务组合的流逼近

本节我们利用函数强大数定律给出 Web 服务组合系统的流逼近定理，从而可得到在稳态条件下系统关键性能指标的一阶近似。首先我们引入一个由 Web 服务组合系统所建模成的排队网的网络序列，用  $n=1, 2, \dots, K$  对这个排队网络序列中的各个排队网络进行标识。在第  $n$  个网络中，我们对其涉及到的过程和参数加一个上标  $n$ ，那么有  $Q^n(0)$ ， $A_0^n(t)$ ， $A^n(t)$ ， $S^n(t)$  及相应的参数  $l_0^n$ ， $l_j^n$ ， $m_j^n$ ，为简化计算，我们假设 Web 服务组合系统中的总服务单元个数  $J$  和路径序列不随着  $n$  发生变化。定义  $A_0^n(t) = A_0(l_0^n / l_0^n t)$ ， $A_j^n(t) = A_j(l_j^n / l_j^n t)$ ， $j=1, \dots, J$  和  $S_j^n(t) = S_j(m_j^n / m_j^n t)$ ，假设  $n \rightarrow \infty$  时，

$$l_0^n \rightarrow l_0 > 0; l_j^n \rightarrow l_j > 0; m_j^n \rightarrow m_j > 0. \quad (13)$$

$$c_0^n \rightarrow c_0; c_{a,i}^n \rightarrow c_{a,i}; c_{s,i}^n \rightarrow c_{s,i}. \quad (14)$$

$$\sqrt{n}(l_0^n + l_j^n + \sum_{i=1}^J p_{ij} m_i^n - m_j^n) \rightarrow q_j \quad (j=1). \quad (15)$$

$$\sqrt{n}(l_j^n + \sum_{i=1}^J p_{ij} m_i^n - m_j^n) \rightarrow q_j \quad (j=2, \dots, K, J). \quad (16)$$

其中上式 (13-16) 的极限结果为有限的常量，并模式 (13-16) 为高负荷条件。由函数强大数定律我们可以得到：当  $j=1, \dots, K, J$ ， $n \rightarrow \infty$  时

$$(\bar{A}_0^n(t), \bar{A}_j^n(t), \bar{S}_j^n(t), \bar{R}_i^j(t)) \rightarrow (l_0 t, l_j t, m_j t, p_{ij} t) \quad (17)$$

其中， $\bar{A}_0^n(t) = \frac{1}{n} A_0^n(nt)$ ， $\bar{A}_j^n(t) = \frac{1}{n} A_j^n(nt)$ ，

$\bar{S}_j^n(t) = \frac{1}{n} S_j^n(nt)$ ， $\bar{R}_i^j(t) = \frac{1}{n} R_i^j(\frac{nt}{m_j})$ 。在第  $n$  个网络中其相应的初始队长为  $Q^n(0)$  并且假设当  $n \rightarrow \infty$  时，

$$\bar{Q}^n(0) = \frac{1}{n} Q^n(0) \rightarrow \bar{Q}(0). \quad (18)$$

其中  $\bar{Q}(0) \geq 0$  是一个常数。相应地，在第  $n$  个 Web 服务组合网络序列中相应的性能指标记为  $Q^n$ ， $Z^n$  和  $B^n$  等。令

$$\bar{Q}^n(t) = \frac{1}{n} Q^n(nt), \bar{Z}^n(t) = \frac{1}{n} Z^n(nt), \bar{B}^n(t) = \frac{1}{n} B^n(nt)$$

接下来，我们给出 Web 服务组合系统中关键性能指标的流逼近定理。

**定理 2** 当 (17-18) 成立，那么当  $n \rightarrow \infty$  时，

$$\bar{B}^n(t) \rightarrow et. \quad (19)$$

**证明** 回想  $Q = F(X)$  队长过程的表达式， $X$  满足式 (7-8)，我们对队长的中心化过程  $Q(t) = X(t) + (I - P\Phi)Y(t)$  进行变尺度处理，在第  $n$  个网络中相应的有  $\bar{Q}^n(t) = \bar{X}^n(t) + (I - P\Phi)\bar{Y}^n(t)$ ，当  $j=1$  时，

$$\begin{aligned} \bar{X}_j^n(t) &= \hat{Q}_j^n(0) + \frac{1}{n} q_j^n t + [\bar{A}_0^n(t) - l_0^n t] \\ &\quad - [\bar{S}_j^n(\bar{B}_j^n(t)) - m_j^n \bar{B}_j^n(t)] + [\bar{A}_j^n(t) - l_j^n t] \\ &\quad + \sum_{i=1}^J [\bar{R}_j^n(\bar{S}_i^n(\bar{B}_i^n(t))) - p_{ij} \bar{S}_i^n(\bar{B}_i^n(t))] \\ &\quad + \sum_{i=1}^J [\bar{S}_i^n(\bar{B}_i^n(t)) - m_i^n \bar{B}_i^n(t)]. \\ q_j^n &= l_0^n + l_j^n + \sum_{i=1}^J p_{ij} m_i^n - m_j^n. \end{aligned}$$

当  $j=2, \dots, K, J$  时，

$$\begin{aligned} \bar{X}_j^n(t) &= \hat{Q}_j^n(0) + \frac{1}{n} q_j^n t + [\bar{A}_j^n(t) - l_j^n t] - [\bar{S}_j^n(\bar{B}_j^n(t)) \\ &\quad - m_j^n \bar{B}_j^n(t)] + \sum_{i=1}^J [\bar{S}_i^n(\bar{B}_i^n(t)) - m_i^n \bar{B}_i^n(t)] \\ &\quad + \sum_{i=1}^J [\bar{R}_j^n(\bar{S}_i^n(\bar{B}_i^n(t))) - p_{ij} \bar{S}_i^n(\bar{B}_i^n(t))]. \\ q_j^n &= l_j^n + \sum_{i=1}^J p_{ij} m_i^n - m_j^n. \end{aligned}$$

并且  $\bar{Q}^n(t)$ ， $\bar{X}^n(t)$  和  $\bar{Y}^n(t)$  满足属性 (10-12)，因此有  $\bar{Y}^n(t) = \sup_{0 \leq u \leq t} [-\bar{X}^n(t)]^+$  当  $t \geq 0$ ，因为  $\bar{Y}^n(t)$  非递减，如果对于任意的  $0 \leq s \leq t$ ， $\bar{Q}^n(t) - \bar{Q}^n(s) \leq 0$ ，那么 we 可知

$$\begin{aligned} 0 &\leq \bar{Q}^n(t) - \bar{Q}^n(s) = \bar{X}^n(t) - \bar{X}^n(s) + \bar{Y}^n(t) \\ &\quad - \bar{Y}^n(s) \leq \bar{X}^n(t) - \bar{X}^n(s). \end{aligned}$$

因此， $|\bar{Q}^n(t) - \bar{Q}^n(s)| \leq |\bar{X}^n(t) - \bar{X}^n(s)|$ 。如果  $\bar{Q}^n(t) - \bar{Q}^n(s) > 0$ ，我们考虑以下两种情况：

情况 1:  $\bar{Y}^n(t) > \bar{Y}^n(s)$ ，在这种情况下，

$$\bar{Y}^n(t) = \sup_{s \leq u \leq t} [-X(u)].$$

$$0 < \bar{Q}^n(t) - \bar{Q}^n(s) \leq \bar{Q}^n(t) = \bar{X}^n(t) + \sup_{s \leq u \leq t} [-X(u)]$$

$$= \sup_{s \leq u \leq t} [X(t) - X(u)].$$

情况 2:  $\bar{Y}^n(t) = \bar{Y}^n(s)$ ，在这种情况下，

$$0 < \bar{Q}^n(t) - \bar{Q}^n(s) = \bar{X}^n(t) - \bar{X}^n(s) \leq \sup_{s \leq u \leq t} [X(t) - X(u)].$$

因此当  $0 \leq s \leq t$  有  $|\bar{Q}^n(t) - \bar{Q}^n(s)| \leq |\bar{X}^n(t) - \bar{X}^n(s)|$ 。又因为  $0 \leq \bar{B}_j^n(t) \leq t$  对所有的  $t \geq 0$ ，由 (17) 和 (18) 式中的收敛可得：当  $n \rightarrow \infty$ ， $\bar{X}^n \rightarrow 0$ 。在任意一个时间段  $[t_1, t_2]$ ，由

$$0 \leq |\bar{Q}^n(t) - \bar{Q}^n(s)| \leq |\bar{X}^n(t) - \bar{X}^n(s)| = 0.$$

因此当  $t \in [t_1, t_2]$ ， $\bar{Q}^n(t) = 0$ ，又因为  $\bar{Q}^n(t) = \bar{X}^n(t) + (I - P\Phi)\bar{Y}^n(t) = 0$ ，所以有： $\bar{Y}^n(t) = m^n \bar{I}^n(t)$ ， $\bar{Y}^n(0) = 0$ ，所以我们可得

$$\lim_{n \rightarrow \infty} \bar{Y}^n(t) = 0.$$

由  $\bar{B}^n = et - M\bar{Y}^n$ ，所以  $\bar{B}^n = et$ ，综上定理得证。

### 4 Web 服务组合的扩散逼近

在条件 (13-16) 下，通过对 Web 服务组合关

键性能指标序列的变尺度处理，得到队长和等待时间的中心化过程弱收敛于反射布朗运动、从而我们可以用反射布朗运动平稳分布的均值逼近关键性能指标的稳态值。首先我们对上节给出的  $\bar{A}_0^n(t)$ ,  $\bar{A}_j^n(t)$ ,  $\bar{S}_j^n(t)$  和  $\bar{R}_j^{i,n}(t)$  做如下处理，令：

$$\begin{aligned}\hat{A}_0^n(t) &:= \sqrt{n}[\bar{A}_0^n(t) - l_0^n t] = \frac{A_0^n(nt) - nl_0^n t}{\sqrt{n}}, \\ \hat{A}_j^n(t) &:= \sqrt{n}[\bar{A}_j^n(t) - l_j^n t] = \frac{A_j^n(nt) - nl_j^n t}{\sqrt{n}}, \\ \hat{S}_j^n(t) &:= \sqrt{n}[\bar{S}_j^n(t) - m_j^n t] = \frac{S_j^n(nt) - nm_j^n t}{\sqrt{n}}, \\ \hat{R}_j^{i,n}(t) &:= \sqrt{n}[\bar{R}_j^n(t) - p_{ij} t] = \frac{R_j^i(nt) - np_{ij} t}{\sqrt{n}}.\end{aligned}$$

由上文更新过程定义和 Donsker 定理（更新过程部分和的极限定理），我们可得：

$$\begin{aligned}(\hat{A}_0^n(t), \hat{A}_j^n(t), \hat{S}_j^n(t), \hat{R}_j^{i,n}(t)) \\ \Rightarrow (\hat{A}_0(t), \hat{A}_j(t), \hat{S}_j(t), \hat{R}_j^i(t)).\end{aligned}\quad (20)$$

其中， $\hat{A}_0$  的漂移向量为  $\mathbf{0}$  向量，协方差为  $l_0 c_{0,1}^2$ ，对于  $\hat{A}_j, \hat{S}_j, \hat{R}_j^i$  我们可以表示成向量形式： $\hat{A}, \hat{S}, \hat{R}^i$  是  $J+2$  个相互独立的  $J$  维漂移布朗运动，漂移向量为  $\mathbf{0}$  向量，其相应的协方差矩阵为

$$\mathbf{G}^A = (\mathbf{G}_{jk}^A) \text{ 和 } \mathbf{G}_{jk}^A = l_{j,a,j} c_{a,j}^2 \mathbf{d}_{jk}.$$

$$\mathbf{G}^S = (\mathbf{G}_{jk}^S) \text{ 和 } \mathbf{G}_{jk}^S = m_j c_{s,j}^2 \mathbf{d}_{jk}.$$

$$\mathbf{G}^I = (\mathbf{G}_{jk}^I) \text{ 和 } \mathbf{G}_{jk}^I = p_{ij}(\mathbf{d}_{jk} - p_{ik}).$$

接下来我们对 Web 服务组合系统网络序列中的第  $n$  个网络中的关键性能指标， $Q^n(t)$ ,  $Z^n(t)$  和  $B^n(t)$  做如下变尺度处理，并给出其相应的收敛结果，

$$\begin{aligned}\hat{Q}^n(t) &= \frac{Q^n(nt)}{\sqrt{n}}, \quad \hat{Z}^n(t) = \frac{Z^n(nt)}{\sqrt{n}} \\ \hat{B}^n(t) &= \frac{ent - B^n(nt)}{\sqrt{n}}\end{aligned}$$

**定理 3** 假设条件式 (18-20) 成立，当  $n \rightarrow \infty$  时有：

$$(\hat{Q}_j^n(t), \hat{Z}_j^n(t), \hat{B}_j^n(t)) \Rightarrow (\hat{Q}_j(t), \hat{Z}_j(t), \hat{B}_j(t)),$$

其中，

$$\begin{aligned}\hat{Q}_j &= \hat{X}_j(t) + (I - P\Phi)\hat{Y}_j(t), \\ \hat{X}_j(t) &= \hat{Q}_j(0) + \mathbf{x}_j(t) + \mathbf{q}_j t, \\ \hat{Y}_j(t) &= \mathbf{Y}(\hat{X}_j(t)), \\ \hat{Z}_j(t) &= M\hat{Q}_j(t), \\ \hat{B}_j(t) &= M\hat{Y}_j(t).\end{aligned}$$

当  $j=1$  时：

$$\begin{aligned}\mathbf{x}_j(t) &= \hat{A}_0(t) + \hat{A}_j(t) - \hat{S}_j((l \dot{U} r_j)t) + \sum_{i=1}^J \hat{R}_j^i(m_j t) \\ &\quad + \sum_{i=1}^J p_{ij} \hat{S}_i(t),\end{aligned}$$

当  $j=2, \mathbf{K}, J$  时：

$$\mathbf{x}_j(t) = \hat{A}_j(t) - \hat{S}_j((l \dot{U} r_j)t) + \sum_{i=1}^J \hat{R}_j^i(m_j t) + \sum_{i=1}^J p_{ij} \hat{S}_i(t).$$

$\mathbf{Y}$  是一个相应于反射矩阵  $(I - P\Phi)$  的斜反射映射。

**证明** 首先由  $\bar{Q}^n(t) = \bar{X}^n(t) + (I - P\Phi)\bar{Y}^n(t)$ ，和

$\hat{Q}^n(t) = \frac{Q^n(nt)}{\sqrt{n}} = \sqrt{n}\bar{Q}^n(t)$ ，我们对网络序列的第  $n$  个网络中的队长指标做相应的中心化处理： $\hat{Q}^n(t) = \hat{X}^n(t) + (I - P\Phi)\hat{Y}^n(t)$ ，相应的状态空间  $S^n$  如下

$$S^n = \{x \in \mathbb{R}^J : x_j \geq 0\}$$

当  $j=1$  时，

$$\begin{aligned}\hat{X}_j^n(t) &= \hat{Q}_j^n(0) + \sqrt{n}[\bar{A}_0^n(t) - l_0^n t] + \sqrt{n}[\bar{A}_j^n(t) \\ &\quad - l_j^n t] - \sqrt{n}[\bar{S}_j^n(\bar{B}_j^n(t)) - m_j^n \bar{B}_j^n(t)] \\ &\quad + \sum_{i=1}^J \sqrt{n}[\bar{R}_j^n(\bar{S}_i^n(\bar{B}_i^n(t))) - p_{ij} \bar{S}_i^n(\bar{B}_i^n(t))] \\ &\quad + \sum_{i=1}^J [\bar{S}_i^n(\bar{B}_i^n(t)) - m_i^n \bar{B}_i^n(t)] + \sqrt{n} \mathbf{q}_j t.\end{aligned}$$

当  $j=2, \mathbf{K}, J$  时，

$$\begin{aligned}\hat{X}_j^n(t) &= \hat{Q}_j^n(0) + \sqrt{n}[\bar{A}_j^n(t) - l_j^n t] \\ &\quad - \sqrt{n}[\bar{S}_j^n(\bar{B}_j^n(t)) - m_j^n \bar{B}_j^n(t)] \\ &\quad + \sum_{i=1}^J \sqrt{n}[\bar{R}_j^n(\bar{S}_i^n(\bar{B}_i^n(t))) - p_{ij} \bar{S}_i^n(\bar{B}_i^n(t))] \\ &\quad + \sum_{i=1}^J \sqrt{n}[\bar{S}_i^n(\bar{B}_i^n(t)) - m_i^n \bar{B}_i^n(t)] + \sqrt{n} \mathbf{q}_j t.\end{aligned}$$

其中，当  $j=1$  时， $\mathbf{q}_j t = l_0^n t + l_j^n t + \sum_{i=1}^J p_{ij} m_i^n t - m_j^n t$ ；

当  $j=2, \mathbf{K}, J$  时， $\mathbf{q}_j t = l_j^n t + \sum_{i=1}^J p_{ij} m_i^n t - m_j^n t$ 。并且  $(\hat{Q}^n(t), \hat{Y}^n(t))$  满足属性 (10-12)，因此有  $\hat{Q}^n(t) = F(\hat{X}^n(t))$ ,  $\hat{Y}^n(t) = Y(\hat{X}^n(t))$ 。由第 4 节的流逼近定理 (19) 和连续映射定理，我们可知当  $j=1$  时，

$$\begin{aligned}\lim_{n \rightarrow \infty} \hat{X}_j^n(t) \Rightarrow \hat{X}_j(t) &= \hat{Q}_j(0) + \hat{A}_0(t) + \hat{A}_j(t) - \hat{S}_j(t) \\ &\quad + \sum_{i=1}^J \hat{R}_j^i(m_j t) + \sum_{i=1}^J p_{ij} \hat{S}_i(t) + \mathbf{q}_j t.\end{aligned}$$

过程  $\hat{X}(t)$  是一个初始点为  $Q(0)$ ，漂移量为  $\mathbf{q}$ ，协方差为  $\mathbf{G} = (\mathbf{G}_{kl})$  的布朗运动

$$\begin{aligned}\mathbf{G}_{kl} &= \sum_{j=1}^J m_j [p_{jk}(\mathbf{d}_{kl} - p_{jl}) + c_{s,j}^2(p_{jk} - \mathbf{d}_{jk})(p_{jl} - \mathbf{d}_{jl})] \\ &\quad + l_0 c_{0,l}^2 \mathbf{d}_{ll} + l_k c_{a,k}^2 \mathbf{d}_{kl}.\end{aligned}$$

当  $j=2, \mathbf{K}, J$  时，

$$\begin{aligned}\lim_{n \rightarrow \infty} \hat{X}_j^n(t) \Rightarrow \hat{X}_j(t) &= \hat{Q}_j(0) + \hat{A}_j(t) - \hat{S}_j(t) + \sum_{i=1}^J \hat{R}_j^i(m_j t) \\ &\quad + \sum_{i=1}^J p_{ij} \hat{S}_i(t) + \mathbf{q}_j t.\end{aligned}$$

过程  $\hat{X}(t)$  是一个初始点为  $Q(0)$ ，漂移量为  $\mathbf{q}$ ，协方差为  $\mathbf{G} = (\mathbf{G}_{kl})$  的布朗运动

$$\begin{aligned}\mathbf{G}_{kl} &= \sum_{j=1}^J m_j [p_{jk}(\mathbf{d}_{kl} - p_{jl}) + c_{s,j}^2(p_{jk} - \mathbf{d}_{jk})(p_{jl} - \mathbf{d}_{jl})] \\ &\quad + l_k c_{a,k}^2 \mathbf{d}_{kl}.\end{aligned}$$

接下来我们来证明  $\hat{Z} = M\hat{Q}$ ，我们把式 (4) 重写成相应的变尺度向量形式，并对下面的式子做相应的变形，对于  $j=1$

$$\begin{aligned}
\hat{Z}_j^n(t) = & \sqrt{n}V[\bar{Q}_j^n(0) + \bar{A}_0^n(t) + \bar{A}_j^n(t) + \sum_{i=1}^J \bar{R}_j^{i,n}(\bar{S}_i^n(\bar{B}_i^n(t)))] \\
& - \sqrt{n} \frac{1}{m_j} [\bar{Q}_j^n(0) + \bar{A}_0^n(t) + \bar{A}_j^n(t) + \sum_{i=1}^J \bar{R}_j^{i,n}(\bar{S}_i^n(\bar{B}_i^n(t)))] \\
& + \sqrt{n} \frac{1}{m_j} \bar{Q}_j^n(0) + \frac{1}{m_j} (\bar{A}_0^n(t) - l_0^n t) + \sqrt{n} \frac{1}{m_j} l_0^n t \\
& + \sqrt{n} \frac{1}{m_j} [\sum_{i=1}^J \bar{R}_j^{i,n}(\bar{S}_i^n(\bar{B}_i^n(t))) - p_{ij} \bar{S}_i^n(\bar{B}_i^n(t))] \\
& + \sqrt{n} \frac{1}{m_j} \sum_{i=1}^J p_{ij} [\bar{S}_i^n(\bar{B}_i^n(t)) - m_i^n \bar{B}_i^n(t)] \\
& + \sqrt{n} \frac{1}{m_j} \sum_{i=1}^J p_{ij} m_i^n \bar{B}_i^n(t) - \sqrt{n} (t - \frac{1}{m_j} \bar{Y}_j^n(t)).
\end{aligned}$$

进而，根据 *Random Time-Change* 定理和  $\hat{V}(t) = -M\hat{S}(mt)$  可得：

$$\begin{aligned}
\lim_{n \rightarrow \infty} \hat{Z}_j^n(t) & \circledast \hat{Z}_j(t) = \frac{1}{m_j} [\hat{Q}_j(0) + \hat{A}_0(t) + \hat{A}_j(t) - \hat{S}_j(t) \\
& + \sum_{i=1}^J \hat{R}_j(m_i t) + \sum_{i=1}^J p_{ij} \hat{S}_i(t) + q_j t] + \frac{1}{m_j} (m_j + \sum_{i=1}^J p_{ij} m_i) \hat{Y}_j(t) \\
& = \frac{1}{m_j} \hat{Q}_j(t).
\end{aligned}$$

对于  $j=2, \mathbf{K}, J$  时，我们可以做类似的处理这里不再重复证明，又由于  $\hat{Y}(t) = m\hat{V}(t) = m(t - \hat{B}(t))$ ，可知  $\hat{B}(t) = M\hat{Y}(t)$  至此定理证毕。

因为  $\hat{Q}_j(t)$  和  $\hat{Y}_j(t)$  同样满足属性式 (11-13) 因此我们可得  $\hat{Q}_j(t) = F(\hat{X}_j(t))$ ，并且  $F$  是 *Lipschitz* 连续的。由于  $\hat{X}_j(t)$  是一个布朗运动，因此  $\hat{Q}_j(t)$  是一个反射布朗运动并且相应的漂移向量和协方差如上所示。

### 5 Web 服务组合性能指标计算及数值算例

从第 4 节的逼近与第 5 节的扩散逼近中我们可知在 Web 服务组合的系统中，第  $i$  个 Web 服务单元调用请求的平均数量和平均响应时间可以用排队网络中第  $i$  个节点的平均队长和平均等待时间来近似。从上面的证明中，我们可以看到平均队长和平均等待时间同样满足 *Litter* 定理，并且平均队长过程可以建模成一个给定的反射布朗运动： $Q(t) \circledast \hat{Q}(t) = RBM(S, q, G, R)$ ，其中状态空间为  $S = \{x \in \mathbb{R}^J : x \geq 0\}$ ，漂移向量  $q$ ，协方差矩阵  $G$  和反射矩阵  $R$ ，其相应的表达式和计算过程如第 4-5 节所示。我们称概率测度  $p_0$  为反射布朗运动  $\hat{Q}(t)$  的平稳分布。文献[25]中证明当反射矩阵  $R$  是一个绝对  $S$  矩阵时，反射布朗运动有平稳分布并且唯一，我们定义  $p_0$  是平稳分布  $p_0$  的概率密度。对  $j=1, \mathbf{K}, J$  定义

$$q_i = \int_0^\infty x_i p_0(x) dx,$$

我们知道  $q_i$  是过程  $\hat{Q}_j$  的均值。因此我们可以得到下来性能指标：

- (1) Web 服务组合中第  $i$  个 Web 服务单元调用请求的平均数量为： $q_i$ ；
- (2) Web 服务组合中第  $i$  个 Web 服务单元的平均响应时间为： $\bar{m}_i = q_i / m_i$ ；

(3) Web 服务组合的平均响应时间为： $\bar{a} = \sum_{i=1}^J \bar{m}_i$ 。

对于任意一个给定的反射布朗运动 (*RBM*)，文献[25]给出了 *RBM* 在非负象限上求解平稳分布的 *BNA/FM* 算法，并且文献[24]证明了在 *RBM* 中，当  $R^{-1}q < 0$ ， $2G = RD^{-1}L + LD^{-1}R$  ( $2G_{jk} = -(p_{jk}G_{jj} + p_{kj}G_{kk}) (j \neq k)$ ) 时，其中  $D = \text{dig}(r_{jj})$ ， $L = \text{dig}(G_{jj})$ ，其平稳分布有乘积形式为：

$$f(x_1, \mathbf{K}, x_J) = \prod_{j=1}^J f_j(x_j),$$

其中， $f_j(x) = \begin{cases} h_j e^{-h_j x}, & x \geq 0, \\ 0, & x < 0, \end{cases}$  和  $h_j = \frac{2(m_j - l_j)}{G_{jj}}$  或

其向量形式  $\mathbf{h} = -2L^{-1}DR^{-1}q$ 。

简述而言，对于一个 Web 服务组合系统，在已知各个 Web 服务组合调用请求和 Web 服务单元的到达率  $l_0$ ， $l_j$  及相应的均方变异系数  $c_{0,1}^2$  和  $c_j^2$ ，Web 服务单元处理速率向量  $m_j$  及相应的均方变异系数  $c_{s,j}^2$ ，Web 服务单元处的顾客选择概率矩阵  $P$ ，那么 Web 服务单元中服务调用请求的平均数量就可以应用反射布朗运动  $RBM_{Q(0)}(q, G, R)$  的平稳分布的期望值来近似。本节将以图 9 给出的 Web 服务组合对应的排队网络为例，给出两个数值算例来讨论服务组合的性能指标及其变化规律，其中分析了服务组合网络的队长和等待时间是如何依赖于系统中一些关键参数的变化而变化的。

例 1：我们来以图 9 为例来计算 Web 服务组合排队网络中各个节点的平均队长和平均等待时间是如何依赖 Web 服务单元调用请求到达率的变化而变化的，我们取如下参数，路径矩阵  $P$ ，Web 服务组合中各服务单元的到达率、服务率及相应的方差变异系数如表 1 所示，

$$P = (p_{ij}) = \begin{pmatrix} 0 & 0.3 & 0 & 0 & 0.3 & 0.4 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

表 1 各节点到达率和服务率及方差变异系数

$l_0$	$l_j$	$m_1$	$m_2$	$m_3$	$m_4$	$m_5$	$m_6$	$m_7$
3	[1,10]	14	7	8	8	6	8	10
$c_0^2$	$c_j^2$	$c_{s,1}^2$	$c_{s,2}^2$	$c_{s,3}^2$	$c_{s,4}^2$	$c_{s,5}^2$	$c_{s,6}^2$	$c_{s,7}^2$
0.2	0.2	0.2	0.4	0.5	2	2	1	0.4

Web 服务组合中各个节点的队长和等待时间随着到达率的变化而变化趋势如图 10 和图 11 所示，如下：



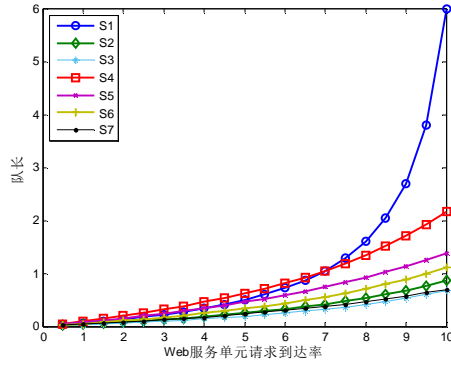


图 10 队长随服务单元请求到达率的变化规律

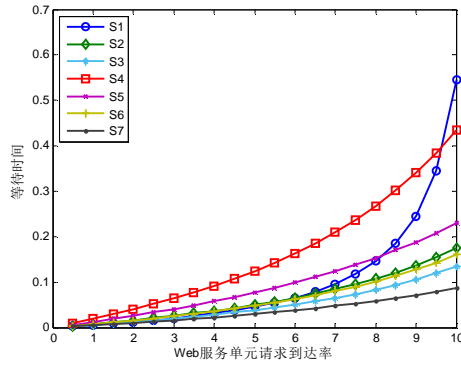


图 11 等待时间随服务单元请求到达率的变化规律

从图 10 上可以看出：当  $\lambda$  递增时，平均队长是递增的，并且随着  $\lambda$  的增加服务节点  $S_1$  和  $S_4$  的增幅速率要比其他节点的快，服务节点  $S_2$ ， $S_3$ ， $S_6$  和  $S_7$  的增速相对平稳。从图 11 中我们可以看出服务节点  $S_1$  和服务台节点  $S_4$  的等待时间增幅速率比较快，其余服务节点的增速也相对平稳，这也复合 Litter 公式的解释。

例 2：我们设服务组合排队网络中各个服务节点的服务率和均方变异系数及其路由矩阵依旧采用例 1 中的数据。我们设 Web 服务单元的调用请求外部到达率  $\lambda = 2$ ，Web 服务组合请求的外部到达率  $\lambda_0 \in [1, 10]$ ，观察此时各个节点的队长和等待时间是如何依赖 Web 服务组合调用请求到达率的变化而变化的，相应的如图 12 和图 13 所示：

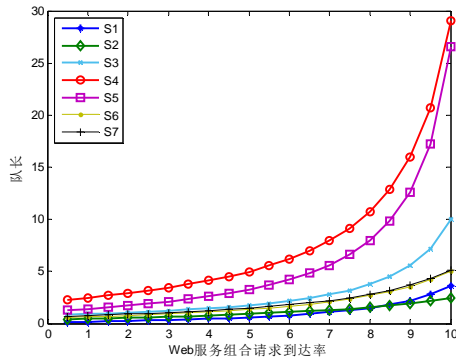


图 12 队长随组合请求到达率的变化规律

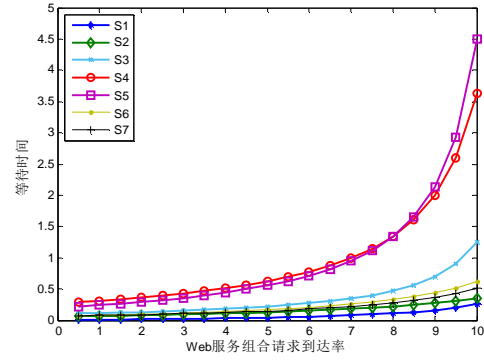


图 13 等待时间随组合请求到达率的变化规律

从图 12 和图 13 可以看出随着 Web 服务组合请求到达率的增加，当  $\lambda \leq 7$  时，各个节点的平均队长和等待时间增速相对平缓，当  $\lambda > 7$  时，服务节点  $S_4$  和  $S_1$  的平均队长和等待时间增速相对较快，在这种情况下其很可能成为影响整个服务组合系统性能的瓶颈节点。

## 6 结论与评论

随着信息服务的不断演化，以面向服务为基础的服务组合系统呈现出规模扩大化、结构复杂化、状态动态化的趋势，对这样的大型 Web 服务组合系统的性能进行分析对研究、构建、优化实际的 Web 服务组合系统具有十分重要的意义。随机过程尤其是马氏过程和排队网络是研究大型复杂网络的有效工具，本文针对 Web 服务组合问题，系统地建立了一套服务组合功能集合之间的基本交互关系及其映射为开排队网络的映射规则，首次将流逼近和扩散逼近理论应用到 Web 服务组合系统的性能分析上，充分考虑了 Web 服务单元的交互关系和负载对 Web 服务组合系统的性能指标的影响，克服了以往应用排队理论研究服务组合系统时，Web 服务调用请求到达时间间隔分布和 Web 服务器处理时间分布受限的情况。并且给出了 Web 服务组合调用请求的平均数量和响应时间的布朗近似及其相应的计算公式。最后通过数值算例分析了服务组合网络中任意一个服务节点的系统性能指标（调用请求的平均数量和平均响应时间）如何依赖于 Web 服务组合调用请求的到达率和 Web 服务单元的调用请求到达率的变化而变化的。基于此，本文在研究 Web 服务组合性能分析问题上给出一个新的研究途径，这对推进人们对大型的、复杂的 Web 服务组合系统的构建和研究具有非常重要的理论意义与实际价值。

本文提出的分析方法也能够解决更加一般的情况，当 Web 服务组合系统中存在部分 Web 服务组合的调用，以及当 Web 服务器的软硬件处理能力

和负载因环境发生变化、失效时的情况。在下一步的工作中,我们将研究当 Web 服务组合受到动态网络环境影响以及存在对 Web 服务质量有不同要求时的服务组合性能问题。

- (1) 多类顾客的 Web 服务组合性能分析问题: 将对整个 Web 服务组合调用请求作为一类顾客, 将对各个 Web 服务单元的调用请求作为一类顾客, 并且两类顾客的路径转移矩阵不同, 第二类顾客可以随时离开网络;
- (2) 分析周期性到达和具有优先权的 Web 服务组合性能评价问题;
- (3) 分析在 Web 服务器的软硬件处理能力和负载环境发生变化时的 Web 服务组合性能问题。

## 参考文献:

- [1] Bichier M, Lin K J. Service-oriented computing[J]. Computer, 2006, 39(3): 99-101.
- [2] Singh M P, Huhns M N. Service-oriented computing: semantics, processes, agents[M]. [S.l.]: John Wiley & Sons, 2006.
- [3] Menasce D A. QoS issues in Web services[J]. IEEE internet computing, 2002, 6(6):72-75.
- [4] Wang Y, Lin C, Ungsunan P D, et al. Modeling and survivability analysis of service composition using Stochastic Petri Nets[J]. The Journal of Supercomputing, 2011, 56(1):79-105.
- [5] Tan Z, Lin C, Yin H, et al. Approximate performance analysis of Web services flow using stochastic Petri net[C]//International Conference on Grid and Cooperative Computing, 2004: 193-200.
- [6] 朱俊, 郭长国, 吴泉源. 一种基于广义随机Petri网的Web服务组合性能预测模型[J]. 计算机科学, 2011, 38(8):125-129.
- [7] 杨怀洲, 李增智. 基于GSPN的Web服务组合与性能分析[J]. 计算机工程, 2009, 35(22):1-3.
- [8] 赵亮. Web服务组合性能建模与分析[D].武汉: 武汉大学, 2009.
- [9] Ding Z, Xiao L, Hu J. Performance analysis of service composition using ordinary differential equations[C]//In the 12th IEEE International Workshop on Future Trends of Distributed Computing Systems, 2008:30-36.
- [10] Ding Z, Shen H, Kandel A. Performance analysis of service composition based on fuzzy differential equations[J]. IEEE Trans on Fuzzy Systems, 2011, 19(1):164-178.
- [11] Sato N, Trivedi K S. Stochastic modeling of composite web services for closed-form analysis of their performance and reliability bottlenecks[C]//International Conference on Service-Oriented Computing, 2007:107-118.
- [12] 肖芳雄, 黄志球, 曹子宁, 等. Web服务组合功能与QoS的形式化统一建模和分析[J]. 软件学报, 2011, 22(11):2698-2715.
- [13] 肖芳雄, 李燕, 黄志球, 等. 基于时间概率代价进程代数的Web服务组合建模和分析[J]. 计算机学报, 2012, 35(5):918-936.
- [14] Xiong K. Web services performance modeling and analysis[C]//In the 2006 International Symposium on High Capacity Optical Networks and Enabling Technologies, 2006:1-6.
- [15] Cao J, Andersson M, Nyderg C, et al. Web server performance modeling using an M/G/1/K\*PS queue[C]//In the 10th International Conference on Telecommunications (ICT), 2003: 1501-1506.
- [16] Qu Z, Wang W, Li Z. Web Server Optimization Model Based on Performance Analysis[C]//In the 6th International Conference on Wireless Communications Networking and Mobile Computing, 2010: 1-4.
- [17] 顾军, 罗军舟, 曹玖新, 等. 基于排队Petri网的服务系统性能建模与分析方法[J]. 计算机学报, 2011, 32(12):2435-2455.
- [18] Das O, Woodside C M. Dependable LQNS: A Performability Modeling Tool for Layered Systems[C]//International Conference on Dependable Systems and Networks (ICDSN), 2003: 672-672.
- [19] Xie B, Chen Y, Sun X H, et al. Performance under failure of multi-tier Web services[C]//In the 15th International Conference on Parallel and Distributed Systems (ICPADS), 2009: 776-781.
- [20] Zhou G, Xia Y, Yu K, et al. A Queuing-network -based approach to performance evaluation of service compositions [C]//In the 12th International Conference on e-Business Engineering (ICEBE), 2015: 135-140.
- [21] 汪浩, 黄明和, 龙浩. 基于G/G/1-FCFS、M/G/1-PS和M/G/∞排队网络的Web服务组合性能分析[J]. 计算机学报, 2013, 36(1):22-38.
- [22] Harrison J M. Brownian models of queueing networks with heterogeneous customer populations[M]. [S.l.]: // Stochastic differential systems, stochastic control theory and applications. NY, Springer, 1988:147-186.
- [23] Harrison J M, Nguyen V. Brownian models of multiclass queueing networks: Current status and open problems[J]. Queueing Systems, 1993, 13(1): 5-40.
- [24] Chen H, Yao D D. Fundamentals of queueing networks: Performance, asymptotics, and optimization[M]. [S.l.]: Springer Science & Business Media, 2013.
- [25] Dai J G, Williams R J. Existence and uniqueness of semimartingale reflecting Brownian motions in convex polyhedrons[J]. Theory of Probability & Its

Applications, 1996, 40(1): 1-40.

- [26] Dai J G, He S. Many-server queues with customer abandonment: A survey of diffusion and fluid approximations[J]. Journal of Systems Science and Systems Engineering, 2012, 21(1):1-36.