

基于随机良构 workflow 的工作流进程性能分析

沙 静 庞善臣

(山东科技大学信息学院 青岛 266510)

摘 要 工作流进程中的 QoS 受很多非功能性因素的影响,例如性能、可靠性以及安全性等等。对 QoS 度量的管理直接影响到参与工作流具体应用的服务能否顺利完成。因此,当服务被工作流或 Web 进程创建或管理时,底层的工作流引擎必须估计、检测并控制用户的 QoS。基于已有的工作流模型分解算法,利用随机良构 workflow 网的数值分析,给出了分析工作流进程 QoS 的一种新方法,并结合实例加以验证。

关键词 QoS, 随机良构 workflow 网, 进程性能

中图分类号 TP302 **文献标识码** A

Performance Analysis of Workflow Process Based on Stochastic Well-formed Workflow

SHA Jing PANG Shan-chen

(College of Information Science and Engineering, Shandong University of Science and Technology, Qingdao 266510, China)

Abstract Quality-of-Service(QoS) in workflow processes encompasses various non-functional issues such as performance, dependability and security, etc. The management of QoS metrics directly impacts the success of services participating in workflow-based applications. Therefore, when services are created or managed using workflows or Web processes, the underlying workflow engine must be able to estimate, monitor, and control the QoS rendered to customers. An approach was given to give QoS of workflow processes based on the decomposing algorithm and numerical analysis of stochastic well-formed workflow(SWWF) model.

Keywords QoS, Stochastic well-formed workflow, Process performance

1 前言

在基于工作流的具体应用中,有关 QoS 度量的管理会直接影响到参与其中的任务能否成功。因此,当运用工作流或 Web 进程来创建或管理服务时,底层的工作流引擎必须能够估计、检测及控制提交给用户的 QoS。本文的关注点是工作流进程的 QoS 分析(研究的性能度量是系统响应时间)。

已有文献^[1]分析了指数分布的随机 Petri 网对工作流网的性能。为了避免结构上的错误,相关作者提出了良构 workflow^[2](由原子控制块、控制块的嵌套和简单的节点构成)的概念。本文所做工作的不同之处是:性能分析是基于所提出的随机良构 workflow (Stochastic Well-formed Workflow, SWWF),将变迁赋以任务的输入率和服务率,并提出 workflow 网基本结构新的性能计算公式和性能分析算法。

本文第 2 节提出 SWWF-net 的概念;第 3 节基于 SWWF-net 给出工作流进程的 QoS 分析;第 4 节运用具体实例验证方法的可行性;最后总结全文。

2 SWWF-net

由 van der Aalst 提出的 workflow 网^[3]是一种高级 Petri 网。一般来说,workflow 网是一个加了标识的图,包含两个特殊的库

所 i 和 o ,用以表示进程的开始和结束。理想情况下,每个变迁都在一条路径上,一个分叉变迁和一个合并变迁界定每条路径。分叉变迁有多于一个的输出库所,合并变迁有多于一个的输入库所。workflow 网不仅用于描绘和验证,还可用于工作流的证明。当我们计算它们的性能时,是假定它们没有结构错误的,已有方法用于证明 workflow 网的结构无错性^[4,5]。在给出本节所提出的 SWWF-net 之前,先看一下 workflow 网的定义。

workflow 网(Workflow net)^[6]是一种模拟 workflow 进程定义的 Petri 网,它满足两个需求:第一个需求是每个 workflow 网有一对特殊库所 i 和 o 。库所 i 没有输入变迁而库所 o 没有输出变迁。库所 i 和 o 分别用来标识一个 workflow 进程的开始和结束,如图 1 所示。在 workflow 管理系统中,库所 i 中的一个托肯表示一个等待处理的工作流实例;库所 o 中的一个托肯代表已被处理的实例。第二个需求是 workflow 网不存在任何悬而未决的任务变迁和条件库所。每个任务和条件都应该服务于实例的处理。因此对于 workflow 网中的每一个库所或变迁都存在一条从库所 i 到库所 o 的路径,使得该库所或变迁是这条路径中的一个节点。如果用一个辅助变迁 t 将库所 o 连接到库所 i ,则后一种需求意味着经辅助变迁 t 连接后的 workflow 网是强连通的。

到稿日期:2010-05-20 返修日期:2010-10-14 本文受国家“973”重点基础研究发展规划项目基金(2010CB328101),国家自然科学基金项目(60773034,60970001)资助。

沙 静(1976-),女,博士,讲师,主要研究方向为 Petri 网的理论及应用、模型验证,E-mail: xo5547@yahoo.com.cn;庞善臣(1974-),男,博士,副教授,主要研究方向为 Petri 网理论与应用、形式化方法、分布式并发系统分析与验证等。



图1 模拟一个工作流进程的工作流网

定义 1(WF-net) 一个 Petri 网 $N_1 = (P; T; F)$ 是一个 WF-net, 当且仅当:

- 1) $i \in P$, 为源库所, 且 $\cdot i = \phi$.
- 2) $o \in P$, 为终止库所, 且 $o \cdot = \phi$.
- 3) $\forall n \in P \cup T$, 必然在一条从 i 到 o 的路径上。

WF-net 不考虑时间, 但在 workflow 管理系统的某些情况下, 我们需要考虑时间因素。例如, 我们想知道整个 workflow 或者一些子网的结束时间, 以便判定 workflow 系统的安排是否符合我们对时间的要求, 并找到整个系统的瓶颈。这样在 workflow 网中引入时间概念就成为必需。

本文将 WF-net 概念做了延伸, 把 Poisson 分布的到达时间和指数分布的服务时间同每个任务联系在一起, 得到下面 SWWF-net 的定义。

定义 2(SWWF-net) $N_2 = (P, T, r, F, \mu, \lambda)$ 是一个 SWWF-net 当且仅当:

- 1) N_2 是一个 WF-net。
- 2) T 是变迁集合, 表示任务集。
- 3) P 是库所集合, 表示状态。
- 4) $r = \{r_i \mid r_i = \{pb_1, pb_2, \dots, pb_m\}, m \in N, i = 1, \dots, |P|\}$, $pb_j \in R$ 是路由率的集合, 对于 $(p_i, t_j) \in F (p_i \in P, t_j \in T)$, 路由率标识其上。
- 5) $\mu = \{\mu_i \mid i = 1, \dots, |T|, r_i \in r\}$ 是任务的到达率集合。
- 6) $\lambda = \{\lambda_i \mid i = 1, \dots, |T|, r_i \in R\}$ 是任务的服务率集合。

3 基于 SWWF-net 的 QoS 分析

SWWF-net 由于从结构上继承了 WF-net, 因此也包含 4 种基本模式^[3]: 顺序路由、并行路由、选择路由和可重复路由, 如图 2 所示。

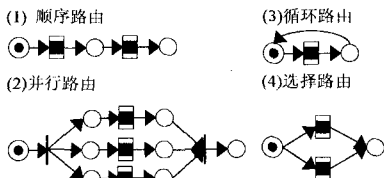


图2 WF-net 的基本路由模式

1) 顺序路由模式

在顺序路由模式中, 任务的执行是顺序进行的。也就是说, 前一个任务执行后接着执行后一个任务。在图 2(1) 中第一个任务执行后接着执行第二个任务。

2) 并行路由模式

在并行路由模式中几个任务可以同时执行。在图 2(2) 中, 3 个并行的任务可以同时执行。

3) 选择路由模式

在选择路由模式中, 从几个任务中有条件地选择执行其中一个任务, 如图 2(4) 所示。根据条件选择任务执行, 完成某一个任务后再进行下一步的操作。

4) 可重复路由模式

在可重复路由模式中, 一个或一组任务可以重复执行多次。在图 2(3) 中的任务执行一次后, 可以继续被重复执行多次。

实例分图^[7]代表工作流的任务集(变迁集)的一个子集, 代表工作流一个特殊实例的执行。更确切地说, 这个子集应包含相关的弧和库所。从 Web 服务的观点来看, 一个服务组合模式可能有多种执行方案, 那么一个子图代表某种服务组合的一种执行方式。为了获得一个工作流网的所有实例子图, 可用相应的算法^[8]对其进行分解。

对于一个 workflow 网来说, 它包含多个实例子图。在 workflow 情形下, 这些实例子图有不同的路由率。为了计算实例子图的路由率, 我们将 workflow 网的定义加以延伸, 为每条从库所发出的弧赋以路由率。当弧上的路由率为 1 时, 代表继续执行接下来的操作。我们用 RP_E 表示实例子图的路由率, 有

$$RP_E = \prod_{p \in P, p \cdot \geq 2} \left(\sum_{i=1}^m pb_i \right) \quad (1)$$

本文研究的是与时间有关的性能方面, 主要是响应时间, 因此每个任务的响应时间是整个分析过程的基础。下面给出根据每个任务的资源到达率和服务率来计算响应时间的方法。

3.1 任务响应时间

时间是 Web 服务 QoS 的一个普遍意义上的性能尺度, 更短的任务执行时间允许更多的服务调用, 也就从某种程度上提高了服务质量。时间的度量是通过任务响应时间实现的。任务响应时间对应于某个任务处理一个服务所需要的时间。这段时间可被划分成两个主要部分: 等待时间和服务时间。等待时间指的是某服务在所需资源还未到达的情况下附加的无价值时间, 其中包括排队延迟、建立延迟等。排队延迟指的是一个服务在被选择进行接下来的处理操作之前, 在任务表上的等待时间。建立延迟指的是服务等待任务被创建的时间。处理时间是任务处理服务所需的时间, 也就是服务的执行时间。响应时间中的每个组成部分都对应着一个重要的参量, 都不应该被忽视。

大量自然物理的和有机体的进程所展示的行为能够通过泊松过程加以模拟。例如, 某类放射性物质放射 α 粒子, α 粒子在哪些时刻出现这一问题就可用泊松过程来建立模型。指数分布通常用来为相邻事件的发生在时间上的间隔的分布建立模型。例如某时刻系统中的顾客, 他们相邻到达的间隔和服务时间都服从指数分布。考虑以上因素, 泊松过程和指数分布已被广泛应用于计算机工程方面的多个领域的分析^[9]。本文用泊松过程来模拟一个任务的输入, 用指数分布模拟任务案例的执行。

从排队网络^[10]的理论得知, 如果输入是一个泊松进程, 顾客的服务时间符合指数分布, 该排队模型称为马尔科夫排队网络。我们可直接将其中有关响应时间的公式加以应用, 并进一步推导整个模型的响应时间。

对于一个任务 t_i 来说, 假定其到达率为 μ_i , 服务率为 λ_i , 用 R_i 来表示 t_i 的响应时间。设 $\rho = \frac{\mu_i}{\lambda_i}$, 则有

$$WT_i = \frac{\rho}{\lambda_i(1-\rho)} = \frac{\mu_i}{\lambda_i(\lambda_i - \mu_i)} \quad (2)$$

$$R_i = WT_i + ST_i = \frac{\mu_i}{\lambda_i(\lambda_i - \mu_i)} + \frac{1}{\lambda_i} = \frac{1}{\lambda_i - \mu_i}$$

式中, WT_i 和 ST_i 分别表示等待时间和服务时间。

3.2 基本模式的 QoS 分析

为了计算整个 SWWF-net 模型的性能尺度, 即完成时间, 我们必须知道 SWWF-net 基本模式的响应时间。接下来

$$T_p = \max\{0.36028, 0.25\} = 0.36028(s)$$

剩下的任务加上这个并行块是顺序的,因此由式(3)可计算出第3个实例子图的响应时间为

$$T_3 = 0.25 + 0.2 + 0.36028 + 0.053 = 0.86328(s)$$

同样地,其它3个实例子图的响应时间为

$$T_1 = 0.795(s), T_2 = 0.82(s), T_4 = 1.153(s)$$

最后,整个模型的完成时间为

$$T = 0.12 \times 0.795 + 0.28 \times 0.82 + 0.18 \times 0.86328 + 0.42 \times 1.153 = 0.9646504 \approx 0.96(s)$$

模型的性能分析完毕。

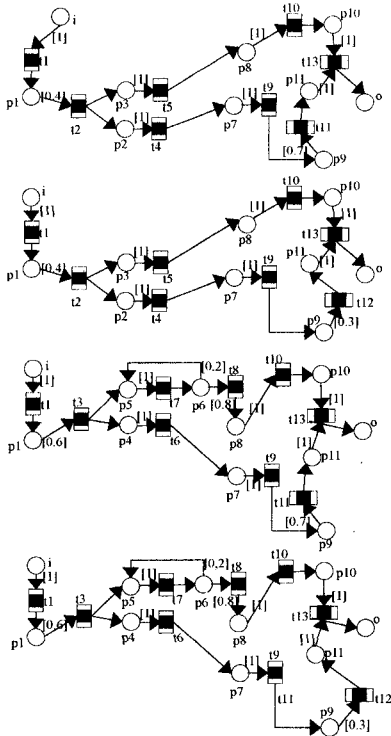


图4 实例子图

结束语 本文将 WF-net 概念做了延伸,把 Poisson 分布的到达时间和指数分布的服务时间同每个任务联系在一起,给出了随机良构工作流网的定义。结合排队网络的知识将资源到达情况及任务执行情况结合起来描述一个任务,推导出

了每个任务的响应时间、每种工作流模式响应时间的计算公式,结合模型的分解算法推导出了实例子图的响应时间,最终结合路由率定义计算出了整个 SWWF-net 模型的完成时间。本文最后结合实例给出模型分解算法和相应的分析,推导出了随机良构工作流网模型的平均响应时间。在今后的研究中将继续探讨及改进分析算法。

参考文献

- [1] Tan Zhangxi, Lin Chuang, Yin Hao, et al. Approximate Performance Analysis of Web Services Flow Using Stochastic Petri Net[C]// Proc. of Third International Conference Grid and Cooperative Computing(GCC 2004), 2004;193-200
- [2] Son J H, Kim J S, Kim M H. Extracting the workflow critical path from the extended well-formed workflow schema[J]. Journal of Computer and System Sciences, 2006, 70(1): 86-106
- [3] van der Aalst W, van Hee K. Workflow Management: Models, Methods, and Systems[M]. The MIT Press, 2002; 271-272
- [4] Aalst W. Workflow Verification: Finding Control-flow Errors Using Petri-net-based Techniques[M]. Springer-Verlag, 2000: 161-183
- [5] Sadiq W, Orlowska M. Applying Graph Reduction Techniques for Identifying Structural Conflicts in Process Models[C]// Proceedings of the 11th International Conference on Advanced Information Systems Engineering. 1999;195-209
- [6] Workflow Management Coalition Terminology and Glossary(WFMC-TC-1011)[R]. Workflow Management Coalition, Brussels, 1996
- [7] Sadiq W, Orlowska M E. Analyzing process models using graph reduction techniques[J]. Information Systems, 2000, 25(2): 117-134
- [8] Li Jianqiang, Fan Yushun, Zhou Mengchu. Performance Modeling and Analysis of Workflow[J]. IEEE Transactions on Systems, Man, and Cybernetics, 2005, 34(2): 229-242
- [9] Kleinrock L. Queueing Systems: Computer Applications[M]. New York; Wiley, 1974
- [10] Trivedi K S. Probability and Statistics with Reliability, Queueing, and Computer Science Applications[M]. Prentice-Hall, Inc., 1982

(上接第 212 页)

- [2] 鲍玉斌,王琢,孙焕良,等.一种基于分形维的快速属性选择算法[J].东北大学学报:自然科学,2003,24(6):527-530
- [3] 闫光辉,李战怀,党建武.基于多重分形的聚类层次优化算法[J].软件学报,2008,19(6):1283-1300
- [4] Traina C, Traina A, Wu L, et al. Fast feature selection using fractal dimension[C]//Proc. XV Brazilian Symposium on Databases. 2000
- [5] Barbará D, Chen P. Using the Fractal Dimension to Cluster Datasets [C]//Proceedings of the Sixth ACM SIGKDD. International Conference on Knowledge Discovery and Data Mining. 2000
- [6] de Sousa E P M, Traina A J M, Traina J C. SID: Calculating the Intrinsic Dimension of Data Streams [C]// Proceedings of the 2006 ACM Symposium on Applied Computing. 2006
- [7] Wong A, Wu L J, Gibbons P B, et al. Fast Estimation of Fractal Dimension and Correlation Integral on Stream Data[J]. Infao-

- ation Processing Letters, 2005, 93(2): 91-97
- [8] Alon N, Matias Y, Szegedy M. The space complexity of approximating the frequency moments[C]//1996 ACM Symposium on Theory of Computing. Philadelphia, 1996; 22-24
- [9] Thorup M, Zhang Yin. Tabulation Based 4-Universal Hashing with Applications to Second Moment Estimation[C]// Proceedings of the Fifteenth Annual ACM-SIAM Symposium on Discrete Algorithms. 2004, 18; 608-617
- [10] Anna O, Rasmus P. Uniform Hashing in Constant Time and Linear Space[C]//Proceedings of the Annual ACM Symposium on Theory of Computing. 2003; 622-628
- [11] Carter J L, Wegman M N. Universal Classes of Hash Functions [J]. Journal of Computer and System Sciences, 1979, 18(2): 143-154
- [12] Charikar M, Chen K, Farach-Colton M. Finding Frequent Items in Data Streams[J]. Theoretical Computer Science. 2004, 312(1): 3-15