Computers & Industrial Engineering 57 (2009) 874–879

Contents lists available at ScienceDirect

Computers & Industrial Engineering

journal homepage: www.elsevier.com/locate/caie

Dwelling time probability density distribution of instances in a workflow model

Liu Sheng [a,*], Fan Yushun [a], Lin Huiping [b]

[a] CIMS Center, Department of Automation, Tsinghua University, Beijing, China
[b] School of Software and Microelectronic, Peking University, Beijing, China

ARTICLE INFO

Article history:
Received 5 August 2006
Received in revised form 1 March 2009
Accepted 2 March 2009
Available online 9 March 2009

Keywords:
Workflow model
Performance
Dwelling time
Probability density

ABSTRACT

This paper presents a method to judge whether a business process is successful or not. A business process is deemed successful if a large enough proportion of instances dwell in a workflow (wait and be executed) for less than given period. By analyzing instances' dwelling time distribution in a workflow, the proportion of instances which dwell in the workflow for less than any given period will be achieved. The performance analysis of workflow model plays an important role in the research of workflow techniques and efficient implementation of workflow management. It includes the analysis of instances' dwelling time distribution in a workflow process. Multidimensional workflow net (MWF-net) includes multiple timing workflow nets (TWF-nets) and the organization and resource information. The processes of transaction instances form a queuing model in which the transaction instances act as customers and the resources act as servers. The key contribution of this paper is twofold. First, this paper presents a theoretical method to calculate the instances' dwelling time probability density in a workflow where the activities are structured and predictable. Second, by this method the analysis of instances' dwelling time distribution and satisfactory degree based on dwelling time can be achieved. The service time of an instance is specified by the firing delay of the corresponding transition (executing time of the corresponding activity). It is assumed that the service request (processing of a transaction instance) arrives with exponentially distributed inter-arrival times and the firing delay of a transition (executing time of the corresponding activity) follows exponential distribution. Then, the instances' dwelling time probability density analysis in each activity and each control structure of a workflow model is performed. According to the above results a method is proposed for computing the instances' dwelling time probability density in a workflow model. Finally an example is used to show that the proposed method can be effectively utilized in practice.

1. Introduction

Workflow technology is good for achieving a process oriented view of the organization and subsequently process automation. Workflow management is an effective means realizing full or partial automation of a business process (Fan, 2001). A business process is a set of one or more linked procedures or activities that collectively realize a business objective or policy goal, normally within the context of an organizational structure defining functional roles and relationships (Fan, 2001). Despite the abundance of workflow management systems developed for different types of workflow based on different paradigms (Adam, Atluri, & Huang, 1998; Sadiq & Orlowska, 2000; van der Aalst, 1998; van der Aalst & ter Hofstede, 2000), the lack of rigorous theoretic foundation and then effective model verification and analysis methods has blocked workflow techniques' research and application (Li, Fan, & Zhou, 2004; van der Aalst, 2003; van der Aalst, Rosemann, & Dumas, 2007).

The rationality and correctness analysis should be carried out from four aspects that are relevant for workflow modeling and workflow execution: process control logic, timing constraint logic, resource dependency logic, and information dependency logic (Li et al., 2004; van der Aalst et al., 2000). The correctness analysis of process control logic aims to avoid the deadlocks or structural conflicts in the execution of a workflow model caused by the errors in its process control. Some verification and conflict detection methods have been discussed in Workflow Management Coalition (1998), van der Aalst and ter Hofstede (2000), Han, Himmighofer, Schaaf, and Wikarski (1996), Hu (2001), van der Aalst (2003), Sadiq, Orlowska, and Sadiq (2005a, 2005b), Sadiq and Orlowska (1997), Hofstede and Orlowska (1999). The objective of resource dependency logic verification is to prove correctness of the static or dynamic resource allocation rules and consistence with the process control logic. The information dependency logic cares about the internal consistence of a workflow-related data and the correctness of temporary relation among different workflow applica-

* Corresponding author. Tel.: +86 10 62771152 1060; fax: +86 10 62770351.
E-mail address: liusheng@tsinghua.edu.cn (L. Sheng).

0360-8352/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.cie.2009.03.002

tion data. The timing constraint verification and analysis deal with the temporal aspects of a workflow model such as deadlines (Panagos & Rabinovich, 1997; Pozewauning, Eder, & Liebhart, 1997; van der Aalst et al., 2007), time scales (Marjanovic, 2000; Marjanovic & Orlowska, 1999a, 1999b; Qu, Lin, & Wang, 2002; Sadiq, Marjanovic, & Orlowska, 2000; van der Aalst et al., 2000; Zhuge, Cheung, & Pung, 2001) schedulability analysis (van der Aalst, 1996), and boundedness verification (Li, Fan, & Zhou, 2003) and time violation handling (Eder, Panagos, Pozewaunig, & Rabinovich, 1999; Eder, Panagos, & Rabinovich, 1999; Sadiq, Orlowska, Sadiq, & Schulz, 2005). Quality of Service in Flexible Workflows is discussed in Sadiq, Orlowska, Lin, and Sadiq (2006), Sadiq, Orlowska, Sadiq, and Lin (2005).

The above analysis can ensure only the functionally working workflow (correctness) but not its operational efficiency. The performance level (Eder, Panagos, Pozewaunig, et al., 1999; Eder, Panagos, Rabinovich, 1999; Ferscha, 1994a, 1994b; Kevin, Ng, Ghanmi, Lam, & Mitchell, 2002; Li et al., 2004; Lin, Qu, Ren, & Marinescu, 2002; Schomig & Rau, 1995; Son & Kim, 2001; van Hee & et al., 2000; Yamaguchi, Qi-Wei, & Tanaka, 2000),on the other hand, aims to evaluate the ability of the workflow to meet requirements concerning some key performance indicators such as, maximal parallelism, throughput, service levels, and sensitivity. The analysis of resource availability and utilization, and average turnaround time is performed at this level. Performance analysis of workflow has not get enough attention of researchers commensurate with its importance until now (Salimifard & Wright, 2001). The performance analysis of a workflow model (business process) is different from that of WfMS architecture (Gillmann, Weissenfels, Weikum, & Kraiss, 2000; Kim & Ellis, 2001).

The performance analysis can be conducted only after the rationality and correctness analysis has been carried out. So it is assumed that there are no temporal and logical errors in the considered workflow models at the performance analysis stage.

PN are the only formal techniques able to be used for structural modeling and a wide range of qualitative and quantitative analysis (Salimifard & Wright, 2001). PN-based workflow management systems are widely used because of formal semantics, local state-based system description, and abundant analysis techniques (van der Aalst & et al., 1998, chap. 10). So PNs are a naturally selected mathematical foundation for the formal performance analysis of workflow models. Many researchers use PN techniques to study workflow (Adam et al., 1998; Ferscha, 1994a, 1994b; Han et al., 1996; Hu, 2001; Li et al., 2003; Lin et al., 2002; Panagos & Rabinovich, 1997; Schomig & Rau, 1995; van der Aalst, 1998; van der Aalst & ter Hofstede, 2000; van Hee et al., 2000) since Zisman used PN to model workflow processes (Zisman, 1977).

A petri net (PN) is a graphical and mathematical modeling tool. It consists of places, transitions, and arcs that connect them. Input arcs connect places with transitions, while output arcs start at a transition and end at a place. There are other types of arcs, e.g. inhibitor arcs. Places can contain tokens; the current state of the modeled system (the marking) is given by the number (and type if the tokens are distinguishable) of tokens in each place. Transitions are active components. They model activities which can occur (the transition fires), thus changing the state of the system (the marking of the petri net). Transitions are only allowed to fire if they are enabled, which means that all the preconditions for the activity must be fulfilled (there are enough tokens available in the input places). When the transition fires, it removes tokens from its input places and adds some at all of its output places. PN which model workflow process definition are called WF-nets (van der Aalst, 1998; van der Aalst & van Hee, 1996). WF-nets are extended to MWF-nets with time, role, and resource information (Li et al., 2004). Methods are discussed to compute the workload that arrival transaction instances generate for the various resource pools and the lower bound of average turnaround time of transaction instances (Li et al., 2004). This paper adopts MWF-nets (Li et al., 2004) as a base mechanism to represent a performance analysis oriented workflow model.

2. Related works

A high-level stochastic PN (SPN) is used to model the routing constructs of a workflow, and then a method to compute throughput time of the process is presented (van Hee & et al., 2000). Based on four performance equivalent formulae, the performance of a workflow is approximately analyzed in Lin et al. (2002). These two techniques both aim at calculating instances' execution time and ignoring waiting time. The probability density of execution time is not taken into account, and cannot be applied to a workflow process of which the resources have stochastic service time. All the control structures are mapped into Generalized stochastic PN (GSPN) (Ferscha, 1994a, 1994b), and then a method based on a CTMC to obtain lower bounds of the execution performance is discussed. A so-called load equivalence aggregation model derived from GSPN has been developed in Schomig and Rau (1995), and then some performance-related measures of human resources in a workflow by obtained by simulating the model. By defining change time, a performance evaluation model for the dynamic workflow changes is brought forward in Yamaguchi et al. (2000). However, the technique can be used for only acyclic time WF-net in which the arrival intervals of transaction instances are constant. A queuing network is used to model the workflow (Kevin et al., 2002; Son & Kim, 2001). A method is yielded to identify the critical path of a workflow model and determine the minimum number of servers for the critical activity (Son & Kim, 2001). Some approximate approaches are employed in Kevin et al. (2002) for workforce configuration, and then the corresponding network is analyzed. But these techniques are not immediately applicable since they both assume that dedicated servers exist for an activity's execution. Existing modeling and analysis techniques used to resolve different aspects of workflow performance-related problems have mainly two shortcomings, which restrict their application in practice. One is that almost no information about instances' dwelling time (sum of waiting time and execution time) is considered. The other is that no accurate approaches are used to analyze probability density distribution of the instances' dwelling time in the whole workflow when the arrival interval and service time are stochastic. Although some literatures take instances' dwelling time into account in their workflow models, only some simple disciplines, such as constant arrival interval (Son & Kim, 2001; Yamaguchi et al., 2000) mean execution time (Lin et al., 2002; van Hee et al., 2000) are supposed. They are not necessarily true in actual workflow systems.

In this paper we discuss a method to judge whether a business process is successful or not. In order to get the proportion of instances which dwell in the workflow for less than any given period, instances' dwelling time distribution in a workflow must be analyzed. There are still no formal papers that focus on this problem so far. This paper will discuss this question.

Section 3 introduces some relevant Queuing Theory and queuing model in the workflow model. Section 4 provides an algorithm computing the instances' dwelling time probability density at a transaction (activity). Section 5 discusses the method to compute the instances' dwelling time probability density at four control structures of workflow models. Section 6 discusses the method to compute the instances' dwelling time probability density at a workflow model. Section 7 presents the method to shorten the instances' dwelling time by increase number of resources. Section 8 presents an example.