

Modeling and Analysis of Multilingual Information Resources Scheduling in Data Grid

Yaojun Han

College of International Business, Shanghai International Studies University, 200083, China

Xuemei Luo

College of International Finance and Commerce, Shanghai International Studies University, 200083, China

Abstract-The multilingual information resource scheduling problem becomes complex, as resources are distributed, heterogeneous, autonomous and multilingual. A multilingual information resources scheduling architecture for data grid is presented in this paper. Petri net is a powerful graphical and mathematics tool for describing the concurrent, asynchronous and dynamic events. This paper models and analyzes the multilingual information resources scheduling scheme using colored timed Petri nets (CTdPN). In CTdPN model, the color represents different language information resource, and the time of a transition is a function of tokens in input places of the transition. The reachable scheduling graph (RSG) of CTdPN is defined and the algorithm for constructing the RSG of CTdPN is given in this paper. Finally, we get some important results such as correctness and makespan on resource scheduling by analyzing reachability of Petri net.

Keywords-Data grid, multilingual, Resource scheduling, Petri net, modeling

I. INTRODUCTION

Multilingual, distributed, heterogeneous, dynamic and autonomous are the major characteristics of the information resource in data grid [1] environment. Because of the complexity of information resource scheduling in data grid environment, the need arises in information resource scheduling for powerful graphical and analytical tools.

Petri nets are promising tools for modeling and analysis information processing systems that are characterized as being concurrent, asynchronous, parallel and distributed [5]. So it has received increasing application in modeling and analyzing resource scheduling. W.M.P. van der Aalst modeled and analyzed scheduling problems using timed Petri nets in [3]. To support the modeling of scheduling problems, it provided a method to map tasks, resources and constraints onto a timed Petri net. But it is very complex and difficult for low-level Petri net to model a big and complex system. In order to settle this problem, one method is to use high-level Petri net such as colored Petri net instead of low-level Petri net. Another method is to introduce hierarchical Petri net. A hierarchical Petri net (HPN) model for the scheduling problems was given in [4]. But the resources that they discussed are homogeneous and single. So it is not suitable to the heterogeneous and multi-resources scheduling in data grid environment. This paper models and analyzes the multilingual information resource scheduling of data grid using colored timed Petri net (CTdPN). The colors represent different languages In CTdPN. The time

delay of transition is a function of colors in input place instead of time constant, which shows the dynamic property of information resource. Finally, we discuss the time and cost performance of the multilingual information resource scheduling by analyzing reachability of CTdPN.

The rest of this paper is organized as follows. In section 2, we give a multilingual information resource scheduling architecture for data grid. The concepts of Petri nets related to this paper were reviewed in section 3. The CTdPN model for multilingual information resource scheduling is presented. Some results about multilingual information resource scheduling are gotten by analyzing reachability of Petri nets in section 4. In section 5, we give an example. The conclusions of the paper are given in section 6.

II. MULTILINGUAL INFORMATION RESOURCE SCHEDULING ARCHITECTURE FOR DATA GRID

Figure 1. shows the multilingual information resource scheduling architecture for data grid. The architecture consists of four layers: user accessing layer, technology supporting layer, scheduling layer and resource layer.

In resource layer, there are two major resources: physical resources and information resources, where physical resources include computing resources, storage resources and network resources, and information resources include corpus, knowledge Base and original information Base. In technology supporting layer, the major technologies include task processing, language translation, ID authentication, model and analysis. The core layer is scheduling layer in multilingual information resource scheduling architecture for data grid. This paper focus on formal modelling and analysis of multilingual information resource scheduling.

III. PETRI NET MODEL FOR MULTILINGUAL INFORMATION RESOURCE SCHEDULING FOR DATA GRID

A. Concepts of Petri nets related to the paper

In this section, we simply review some concepts of Petri nets related to this paper. See references for the details of the definitions of Petri nets.

Definition 1 [5] A Petri Net is a bipartite directed graph represented by a four-tuple $PN = (P, T, F, M_0)$, where,

- (1) $N = (P, T, F)$ is a net,
- (2) $M: P \rightarrow Z$ is marking function, where M_0 is an initial marking, $Z = \{0, 1, 2, \dots\}$ is non-negative integer set,

Definition 2 [5] Let $PN = (P, T; F, M_0)$ be a Petri net. For $x \in P \cup T$,

$$x^* = \{y \in P \cup T \mid (y, x) \in F\} \text{ and}$$

$$\bullet x = \{y \in P \cup T \mid (x, y) \in F\}$$

are called the pre-set and post-set of x respectively.

Definition 3 [5] Let $PN = (P, T; F, M_0)$ be a Petri net.

(1) A transition $t \in T$ is enabled in M iff $\forall p \in \bullet t: M(p) \geq 1$,

(2) A transition t enabled in M can fire and yield a new marking $M'(p) = M(p) - 1$ for any $p \in \text{pre-set of } t$ and $M'(p) = M(p) + 1$ for any $p \in \text{post-set of } t$.

Definition 4 [6] A colored Petri net (CPN) is a seven-tuple $CPN = (P, T; F, M_0, I, I_+, C)$, where,

(1) $(P, T; F, M_0)$ is a Petri net,

(2) C is a function: $P \cup T \rightarrow \rho(D)$, where $\rho(D)$ is a power set of color set D , such that: for any $p \in P, C(p)$ is a set of all possible colored tokens in place p ; for any $t \in T, C(t)$ is a set of all possible colors occurred in transition t ,

(3) I and I_+ are negative function and positive function of $P \times T$ respectively, such that: for any $(p, t) \in P \times T, I(p, t) \in [C(t)_{MS} \rightarrow C(p)_{MS}]_L$ and $I(p, t) = 0$ iff $(p, t) \notin I$; $I_+(p, t) \in [C(t)_{MS} \rightarrow C(p)_{MS}]_L$, and $I_+(p, t) = 0$ iff $(p, t) \notin I_+$, $M_0: P \rightarrow D_{MS}$ is an initial marking that satisfies $\forall p \in P: M_0(p) \in C(p)_{MS}$.

For $t_j \in T$, let $C(t_j)$ be a color set associated to $t_j, c_k \in C(t_j)$, M is current marking of CPN, then t_j is enabled with respect to c_k in a marking M (hereinafter represented by t_j/c_k iff $M(p_i) \geq I(p_i, t_j/c_k)$ for any input place p_i of c_k . A transition t_j enabled with respect to c_k in a marking M can fire and yield a new marking M' .

$$M'(p_i) = \begin{cases} M(p_i) - I_-(p_i, t_j/c_k) & \text{if } p_i \in \bullet t_j - t_j \\ M(p_i) & \text{others} \end{cases}$$

Definition 5 [7] A timed Petri net (TDPN) is a five-tuple $TDPN = (P, T; F, M_0, D)$, where,

(1) $(P, T; F, M_0)$ is a Petri net,

(2) $D: T \rightarrow R$ is a firing time delay, where R is a real set.

B. Colored timed Petri net model for multilingual information resource scheduling in data grid

Definition 6 Suppose that there are n sites in a data grid. The i th site includes information resources of m_i kinds of languages. The colored timed Petri net model (CTdPN) for multilingual information resource scheduling in the data grid is a eight-tuple $CTdPPN = (P, T; F, M_0, I, I_+, C, D)$, where,

$P = \{p_0, p_b, p_f, p_e\} \cup \{p_{j1}, p_{j2} \mid j=1, 2, \dots, n\}$ is a finite set of places, where, p_0 represents that user is ready to submit a task, p_b represents the task submitted by user, p_f represents the expected results on site j , p_e represents the results expected by user, p_{j1} represent the tasks scheduled to site j for executing, p_{j2} represent the results returned from site j .

$T = \{t_b, t_e\} \cup \{t_{j1}, t_{j2}, t_{j3} \mid j=1, 2, \dots, n\}$ is a finite set of transitions, where, t_b is used to submit tasks to schedule, t_e is used to return the results to user, t_{j1} is used to schedule a task to site j according to some algorithm, t_{j2} is used to execute the task on site j , t_{j3} is used to finish the task on site j .

$$F = \{(p_0, t_b), (t_b, p_b), (t_e, p_e)\} \cup$$

$$\{(p_b, t_{j1}), (t_{j1}, p_{j1}), (p_{j1}, t_{j2}), (t_{j2}, p_{j2}), (p_{j2}, t_{j3}), (t_{j3}, p_f), (p_f, t_e) \mid j=1, 2, \dots, n\}$$

is a finite set of arcs,

$C = \{LG\} \cup \{(c_j, e_j) \mid j=1, 2, \dots, n\}$ is a set of colors, where, LG is a set of languages in the task submitted by user, c_j and e_j are communication time from user site to site j and execution time of the task on site j respectively.

$$M_0(p_0) = 1, \forall p \in P \neq p_0, M_0(p) = 0.$$

$$D(t_{j1}) = [c_j], D(t_{j2}) = [e_j], D(t_{j3}) = 0, j=1, 2, \dots, n$$

$$D(t_b) = D(t_e) = 0.$$

In the CTdPN, the firing time delay associated with transition t is a function of color token of input places of transition t .

The graphical representation of CTdPN is shown as Figure 2.

IV. ANALYSIS OF PERFORMANCE OF MULTILINGUAL INFORMATION RESOURCE SCHEDULING IN DATA GRID

A major strength of Petri nets is their support for analysis of many properties and problems associated with concurrent systems. After modeling the with Petri nets, we will analyze the correctness and performance of the multilingual information resource scheduling system in order to ensure the correctness and effectiveness of the system. A fundamental and most widely applied method for analyzing behavioral properties of Petri net models is coverability tree [5]. In this paper, we use the reachable scheduling graph of the model instead of coverability tree to analyze the correctness and effectiveness of the system.

Definition 7 Let CTdPN be a colored timed Petri net model for multilingual information resource scheduling. The reachable scheduling graph (RSG) of CTdPN is defined as a directed graph with labeled directed edges and nodes.

$$RSG(CTdPN) = (V, E, FT). \text{ Where,}$$

$$V = \{R(M_0)\},$$

$$E = \{(M_i, M_j) \mid M_i, M_j \in R(M_0), \exists t_k \in T: M_i[t_k > M_j],$$

$FT(M_i, M_j) = t_k/ct_k$, where ct_k is the current time of transition t_k firing.

The reachable scheduling graph of CTdPN is constructed by the following algorithm.

Algorithm 1 Construction of reachable scheduling graph of CTdPN

(1) Let $V = \{M_0\}, E = \{\emptyset\}$, tag M_0 "new".

(2) If there exists no "new" node in V , then the algorithm ends, otherwise go to (3).

(3) Select a "new" marking M and do the following:

(3.1) If there no exist enabled transition t at M , then tag M with "end node".

(3.2) While there exist t at M , do the following for each enabled transitions t at M :

(3.21) Obtain M' that results from firing t at M .

(3.22) If $M' \notin V$, then $V = V + \{M'\}$ and tag M' with "new".

(3.23) $E = E + \{M, M'\}$, tag $\{M, M'\}$ with t_i/ct_i , where ct_i is the current time of transition t_i firing.

(3.3) Remove "new" from M and go to (1).

The correctness of the algorithm 1 can be easily proven according to the definitions of CTdPN and RSG.

Proposition 1 Let CTdPN be a colored timed Petri net model for a multilingual information resource scheduling. The scheduling is correct iff for any end node $M \in \text{RSG}(M_0)$, $M(p_e) \neq 0$ and any $p \neq p_e$, $M(p) = 0$.

The proff of proposition 1 is obvious.

Proposition 2 Let $\text{RSG}(\text{CTdPN}) = (V, E, FT)$ be the reachable scheduling graph of CTdPN and the scheduling be correct. The makespan for the scheduling is equal to $ct_e - ct_0$, where, ct_e is the current time of transition t_e firing, ct_0 is the current time of transition t_0 firing.

The proff of proposition 2 is obvious.

V. EXAMPLE

Suppose that there are two machines A and B in a data grid. Both machines store information of four kinds of languages L1, L2, L3 and L4. Suppose that a user submits a task including four kinds of languages L1, L2, L3 and L4 at time θ . The execution time and transmission time of the task are 5 and 3 in machine A. The execution time and transmission time of the task are 6 and 7 in machine B.

First, we construct the CTdPN for multilingual information resource scheduling of the task. The graphical representation of the CTdPN for the example is similar to Figure 2.

Second, We construct RSG for the example, as show in Figure 3.

From Figure 3, we know that the scheduling of this example is correct because for end node M_9 , $M_9(p_e) \neq 0$ and any $p \neq p_e$, $M_9(p) = 0$. Also we know that the makespan for the scheduling is 8 if the task is executed in machine A, and the makespan for the scheduling is 13 if the task is executed in machine B. If we use the XMin-min algorithm [2] for scheduling the task, the makespan for the scheduling is 8.

VI. CONCLUSION

In this paper, we propose a multilingual information resource scheduling architecture for data grid. We use a

colored timed Petri net (CTdPN) to model multilingual information resource scheduling. In CTdPN model, the time interval of a transition is a function of tokens in input places of the transition, which is convenient for modeling and analyzing dynamic performance of the data grid. The reachable scheduling graph (RSG) of CTdPN is defined and the algorithm for constructing the RSG of CTdPN is given in this paper. Finally, we get some important results such as throughput, load balance and makespan on resource scheduling by analyzing reachability of Petri net. The example shows that the model given in this paper is simple and effect. We get some important results such as correctness and makespan on resource scheduling by analyzing reachability of Petri net.

ACKNOWLEDGEMENT

This paper is supported partially by The Philosophy and Social Sciences Planning Project in Shanghai (2010BTQ001).

REFERENCES

- [1] I. Foster and C. Kesselman, ed., The Grid: blueprint for new computing infrastructure. Morgan Kaufmann Publishers. San Francisco, CA, 1999
- [2] Yaojun Han, Changjun Jiang, You Fu, Xuemei Luo. Resource Scheduling Algorithms for Grid Computing and Its Modeling and Analysis Using Petri Net. LNCS 3033, Springer-Verlag Berlin Heidelberg, 2004, PP73-80.
- [3] W.M.P. van der Aalst. Petri net based scheduling. Computing Science Reports. No. 95, Eindhoven University of Technology, 1995.
- [4] Yu DA, Zhang Bo, Chen Chen. An approach to HPN model of the scheduling problem. Computer Research & Development. Vol. 33(5), 1996, PP321-328.
- [5] T. Murata. Petri nets: properties, analysis and application. Proceedings of IEEE, Vol.77, Apr.1989, PP541-584.
- [6] Jensen, K. Coloured Petri nets: basic concepts, analysis methods and practical use. Berlin, Heideberg, New York: Springer-Verlag, Vol. 1, 2. ed., 1996.
- [7] W.M. Zuberek. Timed Petri nets: definitions, properties and applications. Microelectronics and Reliability, vol.31, no.4, 1991, PP627-644.

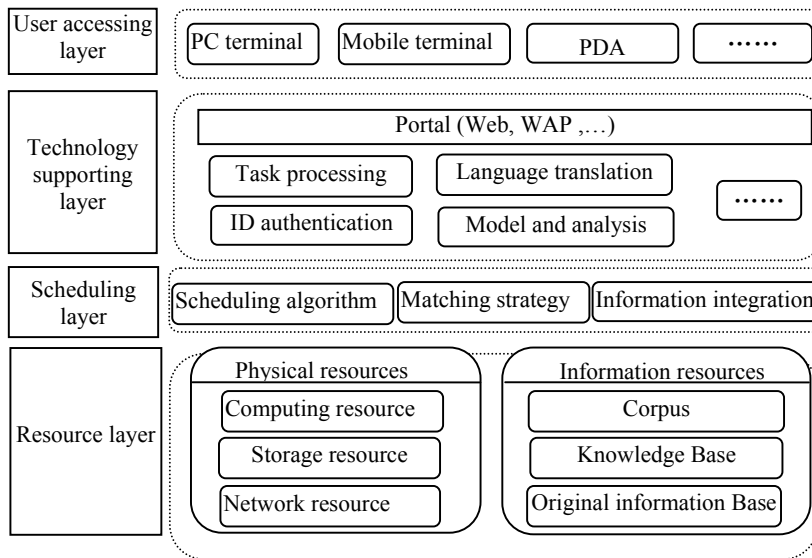


Figure. 1 Multilingual information resource scheduling architecture for data grid

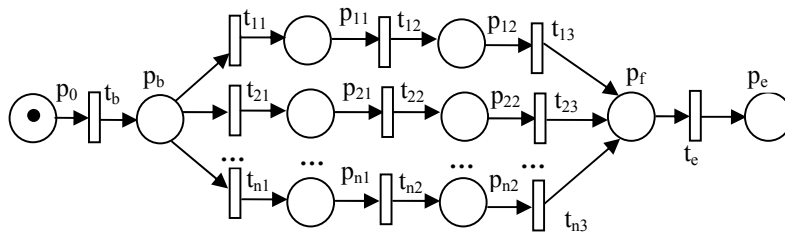


Figure 2. CTdPN model

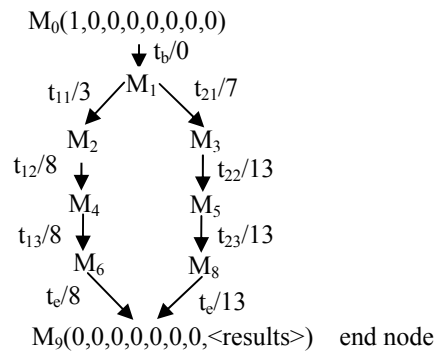


Figure 3. RSG of CTdPN model for the example