Modeling the Scheduling Problem of Identical Parallel Machines with Load Balancing by Time Petri Nets

Sekhri Larbi
Industrial Computing and Networking Laboratory, Computer Science Department, University of Oran, BP 1524 Oran, Algeria
E-mail: larbi.sekhri@univ-oran.dz

Slimane Mohamed
Industrial Computing and Networking Laboratory, Computer Science Department, University of Oran, BP 1524 Oran, Algeria
E-mail: Slimanemhd@hotmail.com

Abstract—The optimal resources allocation to tasks was the primary objective of the research dealing with scheduling problems. These problems are characterized by their complexity, known as NP-hard in most cases. Currently with the evolution of technology, classical methods are inadequate because they degrade system performance (inflexibility, inefficient resources using policy, etc.). In the context of parallel and distributed systems, several computing units process multitasking applications in concurrent way. Main goal of such process is to schedule tasks and map them on the appropriate machines to achieve the optimal overall system performance (Minimize the Make-span and balance the load among the machines). In this paper we present a Time Petri Net (TPN) based approach to solve the scheduling problem by mapping each entity (tasks, resources and constraints) to correspondent one in the TPN. In this case, the scheduling problem can be reduced to finding an optimal sequence of transitions leading from an initial marking to a final one. Our approach improves the classical mapping algorithms by introducing a control over resources allocation and by taking into consideration the resource balancing aspect leading to an acceptable state of the system. The approach is applied to a specific class of problems where the machines are parallel and identical. This class is analyzed by using the TiNA (Time Net Analyzer) tool software developed in the LAAS laboratory (Toulouse, France).

Index Terms— Scheduling Problem, Load Balancing, Time Petri Net, Identical Parallel Machines

I. INTRODUCTION

Recent works in scheduling tend to classify problems and to find techniques to reduce their complexity by developing optimal and efficient algorithms. Among all classifications existing in the literature, Conway’s and Graham’s classifications are considered as the most used to encode such problems [1] and [2]. In [2], a clear separation is made between the three parts of the problem: task, resource and the objective to optimize. In these works, a task is an important element and it doesn’t carry about resources just using them without conflicts [3]. Nowadays with the technology advances, these traditional methods have proved to be inadequate because degrading system’s performances (inflexibility, lack of strategy in the adopted solutions for use resources, etc.)

Petri nets are a formal tool for modeling and analyzing systems that exhibit behaviors such as competition, conflicts and the causal dependencies between events. They are appropriate to model and validate discrete event systems (DES). They are used for modeling concurrent systems such as Occam programs [4], the design of an automated production system [5] and recently in the protocol modeling for wireless sensor network [6].

Petri nets based approaches address the issue of flexibility of tasks and resources. The modeling power and analysis power of Petri nets theory can explore easily different aspects of the scheduling problem (behavioral and structural properties) and integrate factors ignored in previous works [7].

In scheduling related literature, Petri nets theory is applied in several fields. Ramamoorthy used graphical techniques to analyze the cyclical flow in automated production systems in order to represent the scheduling problem by cyclic Petri Nets [8]. Chrétienne provides a timed Petri nets based approach for cyclic generic scheduling problem [9]. Van Der Aalst exploits Petri nets to minimize the cycle time for planning repetitive problems [10]. Korbaa proposes a colored Petri nets based approach to model cyclic scheduling and to reduce the complexity analysis [11]. In [12], a Petri net-based model for scheduling problems with a branch and bound algorithm is proposed using dynamic branching strategy and resource-based lower bound. Other works on scheduling flexible manufacturing systems using Petri-nets and genetic algorithms are proposed in [13]; and colored Petri nets [14].

In these works, performance evaluation of the system is defined only in term of the task execution ordering without conflicts and resources sharing. This fact leads to a bad system performance.

In this paper, we extend classical methods by taking into account the load balancing aspect in the system, and
ensuring an acceptable operating level. The method improves the conventional Petri nets based approaches by ensuring an economic resources using and achieving an acceptable state of the system. The modeled system can be analyzed to verify some properties like: finding a feasible scheduling, detecting conflict, eliminating redundancy, finding the lower/upper limit for the total execution time.

This work focuses on the formal modeling of identical parallel machines scheduling problem by using time Petri net and the formal validation of its properties by using TiNA tool. The preliminary results reveal that the obtained model load is bounded and adjusted among the resources.

The paper is structured as follows. Section 2 introduces some definitions on scheduling problem with an illustrative example. In section 3, we introduce briefly the time Petri nets model. In section 4, we detail our method by explaining different mapping steps translating a scheduling problem into a time Petri net. The method is applied to an illustrative example and analyzed by TiNA tool. Finally we conclude with a conclusion and future prospects.

II. SCHEDULING PROBLEM (SP)

Scheduling can be defined as the allocation of scarce resources over time to perform a collection of tasks. It is concerned with finding the optimal one.

Scheduling techniques are used to answer questions that arise in production planning, project planning, computer control, manpower planning, cloud tasks scheduling, distributed simulations, etc. [15] and [16]. Many techniques have been developed for a variety of problems and classified by categories (single machine, parallel machine, flow job, job shop, etc.). To fix the terminology, we begin by defining the general scheduling problem.

Many problems fit into this definition. In essence, scheduling boils down to the allocation of resources to tasks over time. Some authors refer to resources as ‘machines’ or ‘processors’ and tasks are called ‘operations’ or ‘steps of a job’. Resources are used to process tasks. However, it is possible that the execution of a task requires more than one resource, i.e. a task is processed by a resource set. Moreover, there may be multiple resource sets that are able of processing a specific task. The processing time of a task is the time required to execute the task given a specific resource set. By adding precedence constraints it is possible to formulate requirements about the order in which the tasks have to be processed. We will assume that resources are always available, but we shall not necessarily assume the same for tasks. Each task may have a release time, i.e. the time at which the task becomes available for processing. A schedule is an allocation of resources to tasks over time and for each task it is specified when it is processed and which resources process it.

A. Definitions

Formally, an SP is defined as a triplet \( (G_T(T, C_\cdot), R, C_r) \) where \( G_T(T, C_\cdot) \) is the tasks graph with:
- \( T \): finite set of tasks \( T = \{T_i\}_{i=1}^n \) with :
  - \( s_i \): starting time of the task execution \( T_i \)
  - \( p_i \): processing time of task \( T_i \)
  - \( c_i \): completion time of task \( T_i \)
  - \( r_i \): release time of the task \( T_i \)
  - \( d_i \): due time
- \( C_\cdot \): precedence constraint (partial order relation on \( T \times T \)).
- \( T \) and \( C_\cdot \) make \( G_T \) task-Graph.

Feasible scheduling: A scheduling is feasible if it satisfies the precedence constraints and resource availability constraint.

Objective function: The objective function is a metric that describes the quality of the solution based on specific objectives.

\[ \phi : \Pi \rightarrow \mathbb{R} \]  
(1)

\[ \theta \mapsto q = \phi(\theta) \]  
(2)

Optimal scheduling: The optimal schedule is a feasible solution that optimizes the objective function on the state space.

\[ \theta_\ast = \min_{\theta \in \Pi} \phi(\theta) \]  
(3)

\( R \): finite set of resources \( R = \{R_i\}_{i=m} \) with the maximum available amount \( m \).

\( C_r \): resources constraint. It defines for each task, resource requirements with \( a_i \) the necessary amount of resources \( R_i \) for the execution of \( T_i \).

Scheduling: The scheduling of an instance of an SP is an application that associates to each task \( T_i \) and its resource requirements, start time of its execution during \( p_i \).

\[ \theta : T \times R \rightarrow \mathbb{R} \]  
(4)

\[ (T_i \times R_j) \mapsto t_{ij} \]  
(5)

State space (I): The state space is the set of all possible scheduling on \( T \times R \).

In the sequel we adopt the Graham’s notation [4] \(<a|\beta|\gamma>\) for SP where:
- \( a \) represents the machine environment,
- \( \beta \) represents the characteristics of tasks,
- \( \gamma \) is the objective function.

B. Illustrative Example.

The example illustrated by the Fig. 1 will be used in the sequel as a reference for our mapping method.
According to Graham notation, Fig. 1 represents an instance of the SP, where the environment machine contains two identical parallel machines with precedence constraints and optimizing the total execution time as objective. The different elements of the SP are:

\[ T = \{ T_0, T_1, T_2, T_3, T_4, T_5, T_6, T_7, T_8 \} \]
\[ P = \{ 2, 2, 2, 3, 2 \} \]
\[ < = \{(T_0, T_1), (T_0, T_2), (T_0, T_3), (T_1, T_4), (T_5, T_3), (T_2, T_5), (T_3, T_5)\} \]
\[ R = \{ R_0, R_1 \} \] two instances of this kind of resources are available (processors).

\[ C_k \] each task uses a unit of this resource for its execution.

Fig. 1 gives a possible solution to this instance, its configuration is:

\[ \text{Scheduling} = \{(T_0, R_1, 0), (T_1, R_1, 4), (T_2, R_1, 2), (T_3, R_2, 3), (T_4, R_1, 6), (T_5, R_2, 6), (T_6, R_2, 10)\} \]

C. Working hypotheses

In this study, a number of assumptions on the SP model are to consider:

- Simultaneous use of a resource by tasks is prohibited.
- Once launched, a transaction must complete without interruption.
- Lack of communication between resources and tasks.
- No duplication of tasks.
- Processing times are fixed and known in advance.
- Model used is parallel identical machines SP where resources are independent and having the same features.

This model is classified by Graham as <P/μ/γ>.

III. TIME PETRI NETS

Time Petri nets (or TPN for short) [17] and [18] are a convenient model for real time systems and communication protocols. TPN extend Petri nets by associating two values (min, max) of time (temporal interval) to each transition. The value min (min≥0), is the minimal time that must elapse, starting from the time at which transition \( t \) is enabled until this transition can fire and max (0≤max≤∞), denotes the maximum time during which transition \( t \) can be enabled without being fired. Time min and max, for transition \( t \), are relative to the moment at which \( t \) is enabled. If transition \( t \) has been enabled at time \( a \), then \( t \) cannot fire before \( a+\text{min} \) and must fire before or at time \( a+\text{max} \), unless it is disabled before its firing of another transition.

We assume that the reader is familiar with the definitions and basic concepts of Petri nets [19].

IV. MODELING SP BY TIME PETRI NETS

This section presents main steps of modeling SP by Petri nets. Entities of SP (task, resource, constraint) are associated to elements of Petri net model (place, transition, edge). Two main known translating methods are referenced in literature. These methods enable to map a SP into a Petri net model. The first method is due to Chrétienne and the second one is due to Van der Aalst. In both methods, translating process is made up of six steps:

- Modeling tasks,
- Modeling precedence constraints,
- Modeling resources,
- Modeling constraints of available resources,
- Initial marking of time Petri net
- Synchronization mechanism.

Unfortunately in these methods, there is no way to control the policy of the resources used by the system. This fact is considered as the main drawback of such methods. For this reason, in this paper, we propose an extension of these traditional approaches by taking into account this fact.

A. Scheduling Problem Modeling with Load Balancing

In previous methods there is no way to know how resources were used; also the interest is oriented only to tasks availability not for their loading. Our approach tries to combine the best of both previous methods by incorporating load balancing of resources. The main idea is to model separately each constraint, then separate the Petri net corresponding to the tasks graph from those corresponding to resources. Both Petri nets are connected in specific points through which communication is done. This approach allows us to adjust the load on resources resulting from the execution of tasks and therefore achieve a steady state system with a feasible scheduling. The method will be applied to the case study formed by identical parallel machines. Our mapping method proceeds in three main steps: mapping tasks graph into time Petri net task, mapping resources into time Petri net resources and finally connect these two generated nets [20] and [21].

\[ SP = (G_T (T, C_s), R, C_k) \] be an identical parallel machine scheduling problem.

- Mapping tasks graph

\[ G_T (T, C_s) \] is transformed into tasks time Petri net

\[ I.J. ~Intelligent ~Systems ~and ~Applications, ~2015, ~01, ~42-48 \]
**Step1.** Modeling tasks Ti (see Fig. 3 (a)).
- $T_{Si}$: start transition to launch $T_i$.
- $P_i$: associated place for $T_i$.
- $T_{Ci}$: completion transition to end $T_i$.
- Create an arc from $T_{Si}$ to $P_i$ and another from $T_{Ci}$ to $P_i$.

**Step2.** Modeling the precedence constraint $T_i < T_j$ by adding (see Fig. 3 (b));
- $P_{ij}$ place associated with this constraint.
- Create an arc from $P_{ij}$ to $T_{Sj}$ and another from $T_{Ci}$ to $P_{ij}$.

**Step3.** Defining the initial marking
- All places are empty (initially no task is in execution).

**Step4.** Modeling the synchronization mechanism
- The time is associated with completion transitions $T_{Ci}$ and equal to $p$.

![](image1.png)

**Fig. 3. Tasks Time Petri net**

In the resulting net, two specific places are added:
- $P_C$: Start place to start the process. $P_S$ is initialized with a token that passes through the rest of the net.
- $P_C$: Completion place to end the process. The process stops when the token reaches the place $P_C$.

By applying the mapping process to the $SP$ illustrated by the Fig. 1, we obtain the tasks time Petri net represented by the Fig. 2. This part of solution illustrates a task graph side as a time Petri net node starting by task start transition. The place having a token represents the task processing and the task end transition represents the task delay. When the task finishes then delivers its outputs to the next step. The overall process starts with token on its first place $PS$ and ends when the last transition $PC$ expires it’s waiting time as execution of the last task in the task graph.

- **Mapping resources**
  Resources are translated into Resources Petri Net.
  **Step1.** Modeling Input / Output Resources:
  - $R_{ip}, R_{op}, P_{SW}$: Input / Output / Switch place.

  **Step2.** Modeling resource $R_j$ is achieved by introducing:
  - $P_{Rj}$ place
  - $Q_{Rj}, Q_{Oj}$: place points of entry / exit non-reentrancy be associated with $R_j$
  - $T_{Rj}, T_{Oj}$: transitions input / output of $R_j$
  - $(P_{Rj}, T_{Rj})$, $(T_{Oj}, P_{Oj})$: arcs are connected to the outside
  - $(T_{Oj-1}, Q_{Oj})$, $(T_{Rj-1}, Q_{Rj})$: arcs connecting the preceding and following resources

  **Step3.** Defining the initial marking
  - $P_{Rj}$ contains $m$ tokens.

By applying these steps to $SP$ of Fig. 1, we obtain the resource time Petri net illustrated by the Fig. 4.

- **Connection of Tasks time Petri Net and Resources time Petri Net**
  Before the connection we model resource constraints and allocation of resources to tasks.
  - $(P_{OP}, T_{SP})$: Resource application.
  - $(T_{Si}, P_{SW})$: Notification of use of the resource.
  - $(T_{Ci}, P_{IR})$: Release of the resource.

Connection of tasks time Petri net (Fig. 2) and resources Petri net (Fig. 4) forms a time Petri net illustrated by the Fig. 5.

When the task is ready to run it consumes available resources modeled by the place $P_{DR}$. Then, this task generates a token on the place $P_{SW}$ meaning that the next resource is ready to be consumed. The load distribution is ensured by the fact that each resource can be fairly used. When the place $P_{SW}$ is marked and goes out to the next resource that will put the token in the place $P_{sw}$ (see Fig. 5). This process is repeated while exist enough resources to satisfy requests made by tasks. In other words, we expect the release of used resources. When a task completes then it releases the possessed resource. By recovering this resource, the place $P_{IR}$ gives the control to the next resource.

This solution respects two principles; the first one using the resources in round robin fashion and the second concerns the availability of the resource for task execution. In this situation, the execution is blocked until a resources is released (become available) then it will be allocated to one of the waiting tasks. After this we will pass turn token to the next resources as marked for allocating next time.

**B. Properties Analysis**

The analysis of time Petri net illustrated by the Fig. 5 is performed using the TiNA tool software version 8.0 (www.laas.fr/bernard/tina.htm). The analysis has shown that the network has good properties. Fig. 6 summarizes some analytical results given by Net Control Draw of TiNA tool. It is possible to analyze all firing sequences of a time Petri net corresponding to a scheduling problem. Each firing transitions sequence starting from initial marking to final marking corresponds to a feasible
scheduling. Transition sequences belonging to time Petri net given by Fig. 5 define the using of resources by the tasks. A transition sequence from the initial place \( P_S \) to final place \( P_C \) models ordered execution of tasks and ordered use of resources.

In the time Petri net of Fig. 5, a transition sequence generated by using TiNA tool is given as scheduling solution such a 5-tuple \( (\text{Task}, \text{begin}, \text{Used-Ressource}, \text{Execution-time}, \text{end}) \):

\[
\begin{align*}
\text{TRACE} &= \text{TOR1}@0.0 \quad \text{TS0}@0.0 \quad \text{TOR2}@0.0 \\
&\quad \text{TC0}@2.0 \quad \text{TS2}@0.0 \quad \text{TIR1}@0.0 \quad \text{TOR1}@0.0 \quad \text{TS1}@0.0 \\
&\quad \text{TC2}@2.0 \quad \text{TIR2}@0.0 \quad \text{TC1}@0.0 \quad \text{TOR2}@0.0 \quad \text{TS4}@0.0 \\
&\quad \text{TIR1}@0.0 \quad \text{TOR1}@0.0 \quad \text{TS3}@0.0 \quad \text{TC4}@2.0 \quad \text{TIR2}@0.0 \\
&\quad \text{TOR2}@0.0 \quad \text{TC3}@1.0 \quad \text{TS5}@0.0 \quad \text{TIR1}@0.0 \quad \text{TOR1}@0.0 \\
&\quad \text{TC5}@3.0 \quad \text{TS6}@0.0 \quad \text{TIR2}@0.0 \quad \text{TOR2}@0.0 \quad \text{TC6}@2.0 \\
&\quad \text{TIR1}@0.0
\end{align*}
\]
The reachability graph shows all accessible states. In the context of our case study by time Petri nets, the reachability graph represents a spanning tree of all possible paths for task execution. Based on the properties of such graph, we can determine not accessible states, all terminal states (leaves of the tree), etc. If one of these leaves matches a final task in the task graph then we find a critical path that shows the longest sequence of tasks execution in this graph.

After launching the simulation, we will save its trace. By analyzing it, we will depict the most important features of this modeling method. Fig. 7 depicts that time Petri net of Fig. 5 is bounded and the reachability graph (state classes’ graph) contains 174 markings (state classes) and 318 transitions. The trace generated by TiNA simulator shows that the obtained solution has 12 units length, this means that the optimal solution has a Makespan $C_{max} \leq 12$ units for this instance of problem.

So, based on this result, we can fix the parameters of other algorithms to get this optimal solution.

V. CONCLUSION AND FUTURE PERSPECTIVES

In this paper we have proposed a new approach to mapping a scheduling problem into a time Petri net. This work extends previous works by taking into consideration the load balancing factor. Our mapping method is applied to specific class of SP (identical parallel machines). The method is based in a clear separation between main parts of a SP: tasks graph and resource constraints. Both parts are translated independently into a time Petri net. The resources time Petri net can be apportioned equally (resource use policy) serving requests from the first time Petri net (tasks Petri net). The transformation of a SP into a time Petri net allows us to analyze properties as: detecting resources conflicts, search of feasible scheduling, determining bound inf/sup of global run time, load balancing, etc. This validation can be done by software TiNA tool. Actually, as future prospects, we plan to extend our method to other classes of SP as no identical heterogeneous parallel machines.

ACKNOWLEDGMENT

This work was supported in part by the Industrial Computing and Networking Laboratory of Oran University.

REFERENCES


**Authors’ Profiles**

**Sekhri Larbi** is an Associate Professor at the Computer Science Department of Oran University. His current research area of interests include formal modeling in distributed and mobile systems, wireless ad-hoc and sensor networks, systems modeling using Petri nets, diagnosability and monitoring of automated production systems. He is member of the Industrial Computing and Networking Laboratory at Oran University. He has been a visiting professor at Cedric-CNAM research laboratory, in Paris, France, and Ecole Centrale de Lille (LAGIS) where he worked in Diagnosis of Industrial Systems; and LIUPA Laboratory at the University of Pau, France and distinguished lecturer at University of Ottawa, Canada.

**Slimane Mohamed** is a Ph. D. candidate in Department of Computer Science; University of Oran, Algeria. His current research area includes Flexible Manufacturing Systems Modeling, Scheduling Problems, distributed and parallel programming, simulation and information security. He received his Ms.Bc degree in scheduling and mapping distributed applications from Le Havre University, France. He is an assistant professor in Computer Science Departement at University of Chlef, Algeia.

**How to cite this paper:** Sekhri Larbi. Slimane Mohamed,"Modeling the Scheduling Problem of Identical Parallel Machines with Load Balancing by Time Petri Nets", *International Journal of Intelligent Systems and Applications (IJISA)*, vol.7, no.1, pp.42-28, 2015. DOI: 10.5815/ijisa.2015.01.04