Solutions of JDF-Workflow Scheduling Problem using Genetic Algorithm

Hao YIN[1], Guangxue CHEN[1],Mirko SEGELEK[2]; 1. State Key Laboratory of Pulp and Paper Engineering, South China University of Technology, 510640, Guangzhou, China; 2. Hochschule der Medien, Nobelstraße 10, 70569, Stuttgart, Germany*

Abstract

In print media industries networked production requires advanced print and media specific job definitions and the communication carrying job relevant information alongside the complete production chain. The relatively young XML based communication standard JDF/JMF allows to specify workflow models and data exchange.

Based on an object-oriented approach, a scheduling and modeling framework is proposed, which implements modules for the resources and controls applied in this particular industry. For this framework the available JGAP library is extended to enable modeling of networked print production using the same JDF-/JMF-communication algorithms as applied in the real production. This leads toward a realization of JDF Workflow for operation planning and control of networked print production in digital printing factory environment. This paper theoretically describes and defines a method of expression for execution sequence of JDF Process Nodes and Process Group Nodes using Hierarchical Dependence Graph, meanwhile generates the Connectivity Matrix for practical use. In order to prepare the Device data for scheduling software, a small database has been set up, with functions to manage Device resources.

The resolves of limited scheduling problems in the field of traditional and digital printing workflows using genetic algorithm may present its basic principle and algorithms for calculation of Connectivity Matrix of JDF nodes, this shall lead to adding more limitations to broader applications in solving print shop scheduling problems.

Introduction

Application of digitalization in printing shops can be distinguished into 3 levels.

On the 1st level, there is further automation of production unit. It carried on the development of analog electronic equipment, both persue precision and safety during the unites' preset, operation and monitoring. After the densely integration of monitoring and control in front of technical staff in form of control panels, productivity can be improved. Highlights of its application are digital control system of a single unit and integration of equipment console.

On the 2nd level, there is networking, in order to solve "information island" problem in a networked production, in other words, the difficulty of effective information sharing and exchange among the whole production process. Acquisition, transmission, storage and processing of necessary data for the management are in the top priority.

On the 3rd level, the system shall intelligently deliver solutions to solve practical problem to to technical and management personnel, involving in decision makings along factory design, daily production, operation management and evaluation[3].

In daily production, under the pressure of striving for profit, printing shop want to reduce equipment vacancy rate, maximize the use of existing resources and improve productivity[1]. Although manufacturers of printing machinery have launched many complex and expensive management information system (MIS) software, but given the unique nature of the printing production process, such as its flexibility and heteromerous, there are no satisfactory automatic solution yet. the scheduling of printing workflow heavily rely on the experience of on-site manager to be done.

A workflow is a set of working steps, which are activated by events individually.Workflow software is integrated system to support workflow management. They implement management of operation in print shop. In this study, better scheduling plan will encounter actual constraints of process order, equipments, in production process. Right and a reasonable description of these restrictions of scheduling scheme is precondition to test the feasibility of scheduling solutions[8].

Before the proposing of the algorithm for finding optimal solutions of scheduling problem, in order to better describe scheduling constraints of print shop, extraction of process priorities and resource constraints information is achieved via JDF files. Through the perspective of directed acyclic graph, Connectivity Matrix that describes the relationship between the process are established. Meanwhile, in order to better grasp the key process nodes, a simple device resource scheduling database are established, with the method of allocation device resources to process node, description of devices restrictions are set for further research of equipment-driven scheduling algorithm.

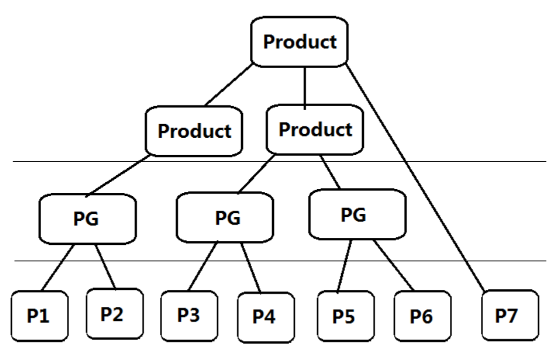

***Figure 1.** JDF Document Tree Structure*

Data structure analysis

From JDF to Workflow

Job definition language defines what the product looks like and how to product it[1] .There are four different kinds of JDF nodes in a JDF document[5]:

1.Product node and product intent node, referred as Product.
2.ProcessGroup node, referred as PG node.
3.Combined process node.
4.Process node, referred as P node.

Not all nodes are representative of some process, only some of them can represent concrete steps in production. Typically, the process node is a child node on the tree structure JDF. In this sense, JDF data structure is only partially based on the process and resource model[7].

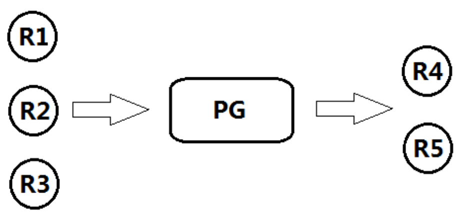

Figure 2*. Process and Resource Model is a Resource Driven Process Model, The PG node here can also be replaced by a P node*

For upcoming production, each part of the product or product component are described by product nodes, For example, in Figure 1, a book is formed by a cover and contents, so they are three corresponding JDF product nodes: book, cover, contents.

Process node defines production steps of print shops or a single work step, such as interpreting, format conversion, screening, cutting, folding, etc..

For various reasons processes can together forms groups or combinations. These type of process nodes are named as ProcessGroup node and CombinedProcess node. For example, a RIP process can be represent by a ProcessGroup node or a CombinedProcess node including interpreting, conversion, screening these three processes. The differences between ProcessGroup node and CombinedProcess node are that processes in ProcessGroup node share resources and scheduling data and processes in CombinedProcess node share same device[2].

JDF workflow scheduling constraints are numerous and various in kind, the following are a few examples.

1)The last process node of one job must be completed prior to its final delivery time.

2)Device status at a given time. DeviceID is that each device driver corresponding unique ID, at any given time, each Device resource can serve as one role, corresponding to one resource ID.

3)Equipment performance capability and the processed Amount.

The sequence of process nodes in JDF workflow.

……

This is obviously not the complete list of constraints, nor the situation of unexpected circumstances, such as interruption of a process node, has been considered. JDF provides a cross-enterprise, cross-system, across-software common language for describing a print job, which provides the basis for integrated automated workflow, integration of equipment, personnel and management control. But JDF is not a detailed workflow description language [11]. Although job informations are well described in JDF files, including detailed processes, but how these processes are performed, as well as relationships between them and the relationships between them and various resources referenced, can not be reflected by JDF document. Since printing is a capital intensive manufacturing industry, a huge proportion of the total cost is the cost of equipment and supplies, this study will focus on equipment scheduling problem.

Directed acyclic graph describing print workflow

Graph for print workflow has following characteristics.

1. No self loop. Self loop means two ends of any edge can be the same vertical.

2. Vertical types are divided into two categories: the process verticals and resources verticals.

3. Resource verticals sometimes only in the output side of an edge, and sometimes only in the input side of an edge, sometimes both.

4. Generally, process verticals can be on both input and output side of any edge, but there are exceptions. For example, JobExport process vertical can have no output edge, therefore no output resource.

5 Nesting verticals. This is a feature of JDF workflow graph, ProcessGroup nodes or Process nodes are nested within product node, Process nodes are nested within ProcessGroup nodes. Figure 2-8-2 shows multiple Process nodes are nested in PG1 and PG2 nodes.

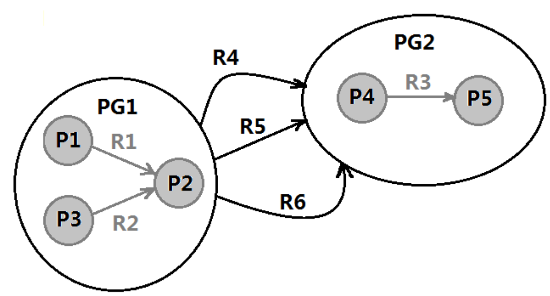

Figure 3. *A simple of directed acyclic graph describing a workflow at Process Group level*

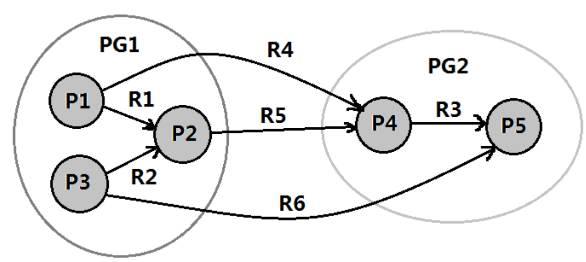

Figure 4. *A simple of directed acyclic graph describing a work-flow at Process level*

Connectivity Matrix is commonly used in comprehensive relationship evaluation. It mainly expressed the relationship between nodes in matrix form other than graph, it is ready for relationship data analysis and processing and traversal of relationships graph .

JDF workflow model has two types of nodes, resource node and Non-resource node. Non-resource node like product node, processes, and process group node, they are "consumer" and "producer" of all the resources. Hence, expression of JDF workflow model diagram have two different perspectives, one is from resource nodes, another is based on non-resource nodes[11].

	PG1	PG2	P1	P2	P3	P4	P5
R1	0	0	-1	1	0	0	0
R2	0	0	0	1	-1	0	0
R3	0	0	0	0	0	-1	1
R4	-1	1	-1	1	0	0	0
R5	-1	1	0	-1	0	1	0
R6	-1	1	0	0	-1	0	1

Figure 5. *Connectivity Matrix base on resource node perspectives*

	PG1	PG2	P1	P2	P3	P4	P5
PG1	0	1	0	0	0	1	1
PG2	-1	0	-1	-1	-1	0	0
P1	0	1	0	1	0	1	0
P2	0	1	-1	0	-1	1	0
P3	0	1	0	1	0	0	1
P4	-1	0	-1	-1	0	0	1
P5	-1	0	0	0	-1	-1	0

Figure 5. *Connectivity Matrix base on resource node perspectives*

In the above Connectivity Matrix from non-resource node perspective, each row or column refer to one JDF node, numbers for each coordinates are of three types:

1. Number 0 indicates that the execution order of row or column nodes is not specific.

2. Number -1 indicates that row node should be executed after to the column node.

3. Number 1 indicates that row node should be executed prior to the column node.

Since these two perspectives are based on the same directed acyclic graph, and they have the JDF tree architecture, in JDF workflow management, we can use a set of different levels of abstraction of directed acyclic graph to describe JDF job information at different stages of its life cycle as long as its tree structure system [6]. The life cycle of JDF includes product intent stage, process group stage and the process execution stage.

Suppose before production, the output resources are not yet available, the device resources are scarce, but other input resources are ample and delivered on time, then the scheduling problems can be divided into two facets, one is allocation of device resource for proper input resources, the another is to ensure that when the device was needed, the assigned device resources are available for use, that is to say, making the status attributes of this device resources has value of “available”.

Based on the characteristics of the scheduling problems, in order to explore solutions to solve scheduling problems, Connectivity Matrix for directed acyclic graph must be generated, which contains process nodes or ProcessGroup nodes that have independent scheduling information. The independence means during scheduling process, equipment, personnel, start time assigned by MIS for each node are totally independent and different from each other.

Along print workflow, the determination of execution time for a process node haw certain constrains, such as:

1. The necessary resources must be in place before start. In JDF document, intended resource node ArtDeliveryIntent node can give detailed description of a resource, will in what time frame, with what manner be sended to print shop, for use.

2. The node took precedence over must be started before succeeding process. If it is running a pipe resources in between, such as paper, then the succeeding process node can be started after completion of some output resources from preceding process. If the output resource from preceding process is indivisible, such as one plates, then the succeeding process must wait until that the preceding process completes its execution. The finish time of the succeeding process is always later than the finish time of the succeeding process.

3. Each piece of device, with intensifier DeviceID, at the same time, only can be used as a resource, with intensifier resourceID.

Total production time of a process is the sum time when its status are “InProgress”, “Setup” or “Cleanup”. The concrete time span depends on the type of the device. To be simple, duration time of a process can be categorized into the following 4 models:

1. The length of process duration time is a constant value. Such as plate making.

2. The length is linear, preset and post maintenance time together are the minimum value, while actual production time depends on the production capacity of device (DevCap Element) and quantity of prints(Amount).

3. The duration time is small enough to negligible. Any phase take time, but if time is so short, that it can not be scheduled, it can be neglected. Calculation of some parameter resources are examples.

4. When pipeline resource is involved, it may occur that speed of the preceding node is slower than the succeeding node, so that progress of the succeeding node have to wait after a period of time. In that case, the end time of succeeding node have to extend some certain time necessary after the previous node is finished [34].

Software design

Open source Java Genetic Algorithms Package, the JGAP, is chosen to be a software core. As evaluator, DeviceFitnessFunction class inherits FitnessFunction class in from JGAP[9]. DeviceFitnessFunction object use genetic algorithm to filter randomly generated schedule solutions.

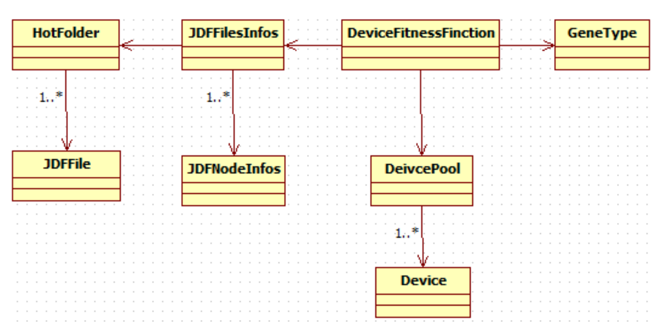

Figure 6. *Design Model of the Software*

Design and coding of gene

Each node from JDF workflow with devices resources is corresponding to two genes units [27]:

The first gene unit indicate option from a collection of eligible devices, each instance of this gene corresponding to one choice. The code is

```
sampleGenes[index] = new IntegerGene(conf, 0, uperBound - 1);
```

, in which the device options are as many as uperBound, the index is the sequence of a gene at this time.

The second gene unit is the start time of this process. The start time will take Time Table coordinates. The code is

```
sampleGenes[index] = new IntegerGene(conf, 0, deadline - 1)
```

The deadline above is the coordinate of delivery time positioning on the Time Table.

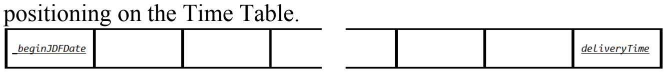

Figure 7. *Time Table Model*

A simple Time Table are set for calculation of time.This Time Table consists of a series of time intervals, each interval in units of 10 minutes, each time interval has a coordinate value. This means that all events fall into time units of a one time block of ten minutes. The reason for this design is that deviation of the actual production time from its scheduling plan is inevitable, time unit set too small will have no practical significance.

There is a common start time for a Time Table, it will be initialized when Fitness function is established. Timetable class has methods converting JDFDate time into the schedule coordinates.

Maximum coordinate value in the Time Table is the delivery time for this job.

The constraints determine the fitness value. Aforementioned theoretical analysis shows the biggest constraint is the order of execution for process node. If those constraints are violated then such a scheduling plan is unfeasible, however, in generating of solution pool in some genetic algorithm steps, the existence of such solutions are still valid meaningful, that is to become preconditions for evolution of the next generation to effective solution. If each generation rules out invalid solutions, it will affect the efficiency of generating of a better solution pool. The fitness of those unfeasible plan are set to be a minimal value, i, in this case.

After determination of its feasibility, the fitness value of a solution can be set as the Time Table length from the finish time of the last process node to its delivery time. This fitness valuation of a solution depends on the completion time of a job, sooner the job is done, better the scheduler is, it is a simplified model.

In accordance with the design principles[24]the advantages and disadvantages of this design are analyzed.

Design principle of GA genes emphasize all genes are independent from each other. Because one device can be used for many process nodes. It is preferably that genes are designed so that controller can assign device according to the genetic codes of process node, so that each gene has nothing to do with each other, there is no such thing when one gene is determined, the other is facing situation with limited choice. Meanwhile, selection of device and start time of process do not limit the choice of each other.

The second design principles is that gene design should keep solutions having trends, which aggregate optimal solutions together, it will help to find a bast solution faster. This design does not follow this principle, but in the fitness value assessment, the impact on efficiency because scattering of optimal solutions can be eliminate.

```
......
Number = 1
type = ProcessGroup
ID = n081128_113710140_000354
DeviceID = ImageSetter:Suprasetter@PMDEMO2_1
DescriptiveName = phototypesetter_1
Start Time = "2008-12-16T16:00:00+01:00"
-------------------------
Number = 2
Type = ProcessGroup
ID = n_081128_113736359_000378
DeviceID = 4810_2
DescriptiveName = SpeedMasterSM74-8-P_2
Start Time = "2008-12-16T16:20:00+01:00"
......
```

Figure 8. *One Scheduling Result (Sample)*

Conclusion

In perspective of digital printing workflow, this paper analyzes process structure of JDF document, the device constrains for solutions of scheduling problem in printing shop are throughly discussed, a design of solution filter using Genetic Algorism are proposed with Java language, CIP4 JDF development kits and extended JGAP framework.

*Corresponding author. E-mail: chengx@scut.edu.cn

References

[1] Wolfgang Kühn, Martin Grell. JDF: Prozessintegration, Technologie, Produktdarstellung[M], Springer, 2004. 20-22

[2] Thomas Hoffmann-Walbeck, Sebastian Riegel. Der JDF-Workflow: Lehrbuch zur Automatisierung in der grafischen Industrie[M]. Beruf + Schule, 2009. 76-77

[3] Oliver Margraf. JDF-basierte Produktionsplanungs- und Steuerungssysteme in der Druckindustrie, Hochschule der Medien[D], Dipl.-Arb, 2004: 25-32

[4] CIP4, JDF Specification Release 1.4a[EB/OL]. http://www.cip4.org/, 2009-11-17

[5] CIP4. XML Schema for JDF Version 1.4a[EB/OL]. http://www.cip4.org/documents/jdf_specifications/schema.php, 2009-11-17

[6] Thomas Hoffmann-Walbeck, Druckformherstellung Offset Job DefinitionFormat[EB/OL]. http://www.hdm-stuttgart.de/ctp/vorle-sung/druckform/Jobtickets-JDF_2_Seiten.pdf. 2011-02-20

[7] Buckwalter C. A JDF-enabled Workflow Simulation Tool[EB/OL].. http://www.cip4.org/document _archive / documents /JDF_Simulation_Tool_-_Buckwalter.pdf. 2010-09-12

[8] Workflow Management Coalition. [EB/OL], http://www.wfmc.org/, 2012-02-02

[9] Klaus Meffert, Neil Rotstan. JGAP[EB/OL], http://jgap.sourceforge.net/, 2012-12-20

[10] W3school. XML技术[EB/OL]. http:/www.W3school.com, 2011-04-02

[11] T. Sun, Hierarchical Dependence Graphs for Dynamic JDF Workflows[J] Systems, Man and Cybernetics, 2006. SMC '06. IEEE International Conference on Volume 4: 2747-2752

Author Biography

Yin hao(1986-), male, Doctoral candidate from South China University of Technology, Guangzhou, China. His work focuses on research direction of digital printing workflow. He Studied in Hochschule der Medien in Germany for Master Degree. He Graduated from Xi'an University of Technology in school of Printing and Packaging Engineering.