QoS-based Dynamic Scheduling for Manufacturing Grid Workflow

Yiping YUAN, Tao YU, Feng Xiong, Minlun FANG
CIMS & Robot Center of Shanghai University, Shanghai, china, 200072
ypyuan@mail.shu.edu.cn

Abstract

Manufacturing Grid (MG) workflow can be defined as the composition of manufacturing activities which execute on heterogeneous and distributed manufacturing resources in virtual organization to accomplish a specific goal. Uncertainties within MG environment pose new challenges for MG workflow. Dynamic scheduling is one of the most critical components in MG workflow. Through analyzing the QoS properties in MG, a scheduling architecture based on QoS for MG workflow is presented. Workflow engine can dynamically schedule the resources according the activity requirements and resource QoS capabilities. If multiple resources that meet the QoS requirements are available to an activity at a time, workflow scheduling can select the 'best match' resource according to scheduling algorithms based on AHP (Analytic Hierarchy Process).

Keywords: Manufacturing Grid, MG workflow, Dynamic Scheduling, QoS, AHP.

1. Introduction

With the rapid technological innovations of the networked manufacturing, much more is learned about inherent limitations of the network technology. Grid is regarded as the next generation Internet as well as Manufacturing Grid (MG) is then presented as an advanced solution for the bottleneck of networked manufacturing. Based on Grid technology, MG is proposed to realize resource sharing and collaborative working among manufacturing resources such as CAD, CAPP, CAE, and various kinds of machine tools that are diverse, heterogeneous and owned by different enterprises [1].

Because the resource is distributed, and heterogeneous, product development for these large-scale collaborative is a difficult task in MG environment. To construct the sophisticated dynamic system, it becomes ever more important and urgent to provide workflow to control and manage the collaborative development process.

This paper is supported by "SEC E-Institute: Shanghai High Institutions Grid" project.

Some projects and systems such as GridPHyN [2], McrunJob [3] and GridAnt [4], also adopt the workflow or planning with the characteristics of workflow to automate the complex Grid applications. However, manufacturing resources have their special particularities: Controlled by software, such as CAD/CAPP/ CAE/CAM; Not always online; Kinds of resources, which are distributed; Different QoS (Quality of Service) capabilities, such as functions and attributes; Task plans are ordinal. Utilization is exclusive. So utilization of MG workflow is distinguished from other data Grid or computing Grid applications. The MG workflow has to be dynamic enough to cope with rapidly change user's requirements and resources with QoS capabilities attribute. To assign dynamically appropriate manufacturing resources for workflow activities, a scheduling approach based QoS is proposed in the paper. It allows the decision of resources allocation to be made dynamically at the time of execution of activity in the workflow according the activity requirements and resource QoS. At the same time, when multiple resources that meet the QoS requirements are available to an activity at a time, workflow scheduling can determine which resource takes care of which activity.

2. The Characteristics and Architecture of MG workflow

2.1. The Characteristics of MG workflow

MG workflow is similar with traditional workflow, which covers fundamental components: workflow model, workflow running mechanism, user and application. There are some distinctions between each other.

Virtual Organization (VO): In the MG environment, organizational structures known as virtual organization are dynamic federation of individuals and organizations aimed at solving a problem of common interests via collaboration. Resources are shared among users in the VO.

Application based on grid services: OGSA defines mechanisms for creating, managing and exchanging information among grid services. A grid service is a Web service that conforms to a set of conventions (interfaces and behaviors) that define how a client interacts with a Gird service. In the MG environment, various kinds of manufacturing resources, software or machine tools, are encapsulated into Manufacturing Service Nodes according to the OGSI specification, with the methods of WSDL description, service mapping, service realization and service deploying [5].

Complexity of the business process: MG focuses on the whole product life-cycle of design, manufacturing and service. The whole processing of MG workflow could be quite complex. A manufacturing job is usually composed of several activities in a pre-defined sequence. The activities of grid workflows are expected to be executed on heterogeneous resources which are geographically distributed. Many different resources may be involved in one workflow execution.

Selection resources based QoS: Manufacturing resource is heterogeneous and distributed. Different manufacturing resource has different function and attributes. Therefore, the processing of resource discovery and selection could be quite complicated according to different QoS levels provided and demanded.

2.2. MG Workflow Architecture

As mentioned above, MG workflow has new features consistent with the MG environment and OGSA. A layered MG workflow architecture is shown in Fig.1.

The user Portal layer: It is an integrated GUI that mainly covers grid workflow model and simulation tool, MG task list, visualization tool and other GUI tools for user to manage workflow. It provides friendly interfaces and facilitates for user to construct and manage the MG workflow.

Workflow core function layer: it includes workflow engine, workflow Admin &monitor and QoS guarantee mechanism. Workflow engine provides operational functions to support execution of business processes based on the process definitions. Workflow Monitor/Management provides administration and monitoring functions for end users to administrate and monitor information about processes. QoS guarantee mechanism provides QoS negotiating and end-to-end QoS guarantee mechanism for MG workflow.

Grid service runtime layer: It offers basic services provided by GT3, such as security, resource management, data management, information management, and lifecycle management, etc.

MG resource layer: It consists of all the globally distributed manufacturing resources that are accessible from anywhere on the Internet. Resources are encapsulated into the grid service in MG, and are transparent to the workflow definition process.

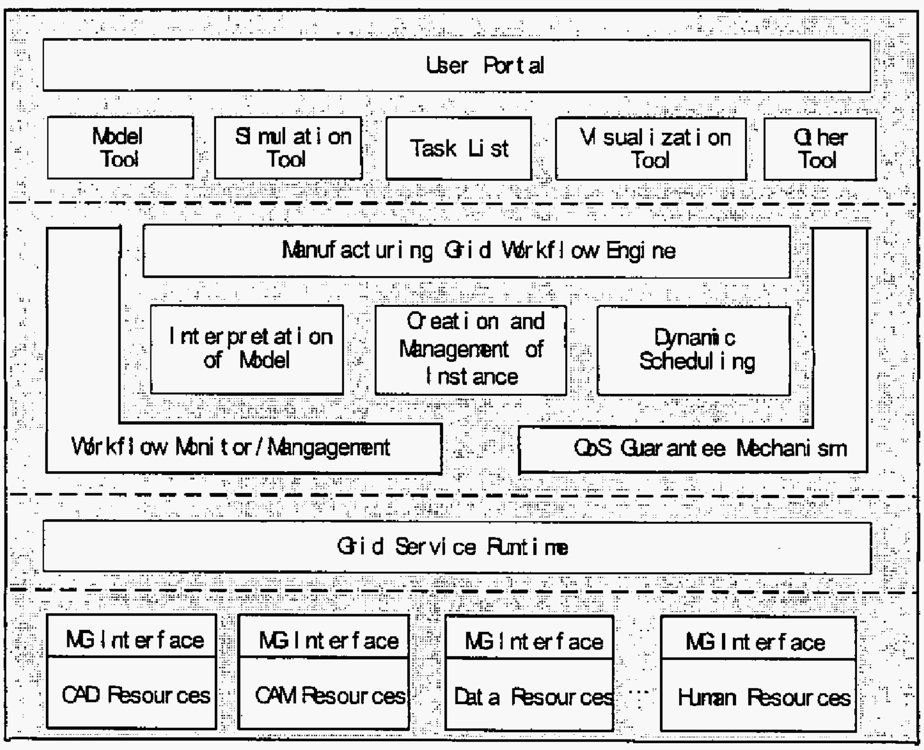

Fig.1. MG workflow architecture

3. QoS Properties

Quality of Service (QoS) is a synthetically guideline, which is used for measuring the satisfaction of a service [6]. It describes some characteristics of a certain service. QoS has been a major concern in the areas of networking, real-time applications and middleware [7]. Comparing QoS criteria (Bandwidth, Throughput, etc.) to general computing Grid, QoS are classified into five criteria in manufacturing area, namely TQCSR (Time, Quality, Cost, Service, and Reliability). Table 1 shows all these essential criteria.

Note that the OGSA specification for service interface definition documents, which defines the elements and their grammars in WSDL, does not include any element for QoS provisions. Therefore, we imagine the incorporation into the main definitions element a new subelement called MG-QoS with various QoS attributes, as showed in the following WSDL service interface definition document:

```
<? Xml version="1.0" encoding="UTF-8"
<wsdl:definitions …>
…
   <MG-QoS Properties>
   <service_name>LOM </service_name>
    <service_provider> CIMS & Robot Center of
     Shanghai University </service_provider>
    <service_Data_Format>STL</
      service_Data_Format>
   <service_type> RP <service_Type>
   <MG-QoS>
      <Reputation_Level>A</Reputation_Level>
      <QoS_T>3 days</QoS_T>
      <QoS_Q>high</QoS_Q>
      <QoS_C>$300</QoS_C>
      …
   </MG-QoS>
    </MG-QoS Properties>
…
</wsdl:definitions>
```

In this way, service providers will be able to define their services using WSDL documents along with their QoS capabilities, and subsequently services would be registered with Monitoring and Discovery Service (MDS) [9]. In GT3, the MDS provides the following features: generating, publishing, storing, searching, querying, and displaying information. Having considered the technologies used by OGSA, namely WSDL, Simple Object Access Protocol (SOAP) and MDS, it can be said that Index Service is the primary registry and service discovery engine. When a service is registered to local index service, aggregator mechanism in the Index Service will collect the corresponding service data, including QoS properties, which is provided in WSDL. Via the service data querying mechanism provided by MDS, we can discover the desired service based on QoS properties.

4. Dynamic Scheduling based QoS

4.1. Scheduling Strategy

MG workflow consists of a collection of coordinated activities designed to carry out a complex process. MG workflow scheduling is to assign appropriate manufacturing resources (services) for workflow activities of which execution obeys the imposed constraints. Scheduling of workflow has two types: static scheduling and dynamic scheduling.

Static scheduling: at the workflow build time, activity is bound statically to a certain resource. At run time, engine invokes the corresponding resource according to the process description. Static scheduling

Table 1: QoS Criteria

QoS Criteria	Description	Measure
Time	The time taken to deliver services between service requestors and providers.	$t = t_o - t_i$ t_i is the timestamp when the service is invoked and t_o is timestamp when the service is delivered.
Quality	The quality of the product service offered including design quality, material quality, processing quality, function quality etc.	The criterion is an abstract and fuzzy concept which is difficult to quantitative calculation. Benchmarking method [8] is adopted to measure the criteria.
Cost	The cost of the service execution.	Assigned by the service provides
Service	Customer service including the before service and after service.	Benchmarking method.
Reliability	This parameter is related to the number of success of a service.	$R(success) =$ Num of successful executions / N, N: Total number of invocations.

requires the specification of the exact location of resources assigned to every activity. The workflow execution failure could be caused by the status change of the specified resources during the workflow.

Dynamic scheduling: at workflow build time, the activity is bound to description of resource QoS rather than a certain resource. The exact location of resource that meets activity requirements is unknown. At run time, engine dynamically discoveries and invokes the resources according to activity requirements. When multiple resources that meet the QoS requirements are available to an activity at a time, workflow scheduling is also needed to determine which resource takes care of which activity.

Because resources are adding and removed all the time in MG environment, static scheduling can't continue and lead to fail usually. So the MG workflow engine utilizes the dynamic scheduling. Dynamic scheduling makes the workflow fault-tolerant. Failure of one or more resources would not completely jeopardize the execution of underlying workflow.

4.2. Scheduling Architecture of MG workflow

Based on previous considerations, the dynamic scheduling architecture based QoS is shown Fig.2. It is divided into two parts: the left part is the simplified traditional workflow architecture and the right part is the extension for supporting the proposed scheduling based QoS. It includes workflow definition, resource discovery, resource selection, QoS negotiation, resource contract & reservation and resource invocation. The following is the description of the architecture:

Phase 1: workflow definition A workflow is defined using workflow modeling tools. A workflow model interpreter interprets the workflow definition to extract the activity specifications. Elements of an activity specification include DataStructure, ActivityName, Input/output Data, which are extracted from workflow definition, and QoSReq (QoS Requirements), which is an element added to describe the requirements of activity implementation. QoSReq is major factors that should be considered when search resource. The description of QoSReq is similar to the QoS description of a service in section 3.

Phase 2: Resource discovery Resource discovery module assures that selected logical resources can satisfy user's minimum QoS requirements. Activity specification is used as service requirements to match with corresponding service descriptions by interacting with MDS in the step. If not finding resource that can satisfy the minimum demands, the Error Process module would inform the user and terminate the applying;

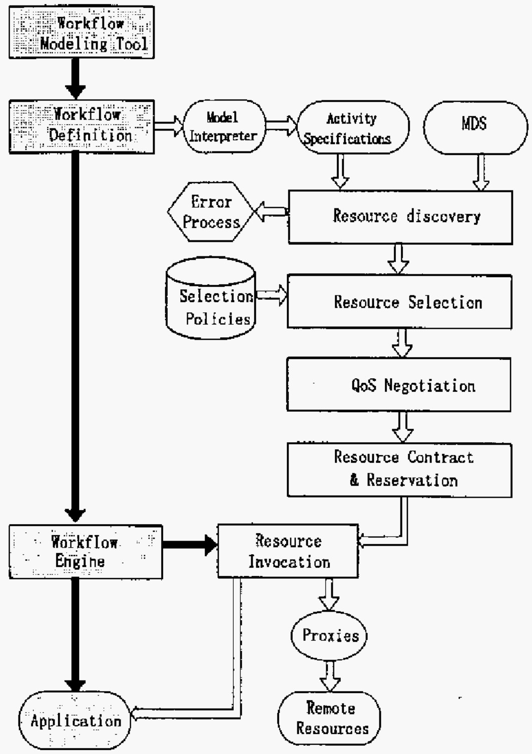

Fig.2. Scheduling architecture of MG workflow

Phase 3: Resource Selection Based on the discovered resource list, Resource Selection module attempts to find the 'best match' resource according to activity specification and certain select police. The resource is expected to meet the time or cost constraints imposed by the user. In order to fulfill the restrictions, it has to gather dynamic information about resource accessibility, machining precision, machining capability, and resource status, etc.

Phase 4: QoS Negotiating Through the Negotiation with resource owners, system judges whether resource QoS can satisfy the user's demands. When presenting resources cannot fully satisfy the user's demand, QoS Negotiation modules should interact with relevant modules and inquire whether the user can reduce QoS demand.

Phase 5: Resource contract & reservation When suitable resources are found, the contracting module establishes resource contracts with the resource providers to guarantee the resource quality they expect during the service session. Such assurance can be contained in an agreement protocol, such as Service Level Agreements (SLAs). Then the Resource reservation module is responsible for setting resources reservation flag and waiting to be invocated.

Phase 6: resource invocation When an activity is instantiated by the workflow engine, the invocation flow from the workflow engine is redirected to the resource invocation module. The resource invocation module binds the activity instances with the corresponding resource and activity instance information. To invoke a remote resource, the invocation message with the relevant data are sent or received by a proxy. For example, a HTTP/SOAP proxy is used to exchange invocation messages and data with a remote HTTP/SOAP server that provides resource. If the resources are a local application, it will be invoked directly.

4.3. Scheduling algorithms

In resource selection phase, the choice of the 'best match' resources is a multi-objective decision-making problem in manufacturing. In this step, scheduling algorithms have to adapt to the different optimization criteria that an activity specification for each particular resource. In order to resolve scheduling problem, many heuristics have been proposed, such as Multi-Objective Decision Making (MODM), Data Envelopment Analysis (DEA), Analytic Hierarchy Process (AHP), Fuzzy Comprehensive Evaluation (FCE), Mathematical Statistics (MS), and Genetic Algorithm (GA), and so on. Among these, AHP is the optimal solution for its characters in dealing with qualitative and quantitative problems in real applications [10]. The AHP, introduced by Saaty [11-12], is particularly useful for evaluating complex multi-attribute alternatives involving subjective criteria.

There are three steps for considering decision by AHP: constructing hierarchies; comparative judgment; and synthesis of priorities and the measurement of consistency, described as follows.

1) Establishment of a structural hierarchy

We establish the AHP model with QoS including three levels, as shown in Fig.4.

Level I: Initially, the objective or the overall goal of the decision is presented at the top level of hierarchy. Specifically, the overall goal of this application is to select the best or most suitable manufacturing resource.

Level II: the second level represents the evaluation criteria (see Table 1) affecting decisions.

Level III: finally, at the lowest level of the hierarchy, the alternatives of the manufacturing resources are identified, which are decision options.

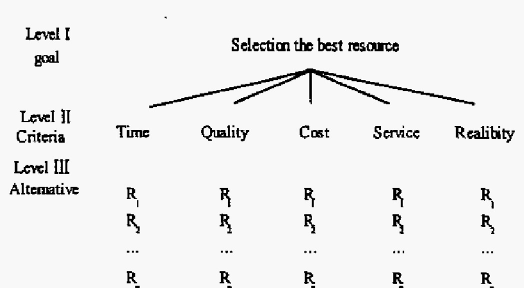

Fig 3.QoS-based scheduling model

2) Establishment of comparative judgments

Construct a set of pair-wise comparison matrices (size n×n) for each of the lower levels with on matrix for each element in the level immediately above by using the relative scale measurement shown in Table 2. The pair-wise comparisons are done in terms of which element dominates the other. The number of matrices depends on the number elements at each level. Five matrices need to be establishment (example: Table 3).

Table 2: Pair-wise comparison scale for AHP preferences

Preference weights /level of importance	Definition
1	Equally preferred
3	Moderately preferred
5	Strongly preferred
7	Very strongly preferred
9	Extremely preferred
2、4、6、8	Intermediates values
Reciprocals	Reciprocals for inverse comparison

Table 3: Pair-wise comparison matrix for quality

Quality	R_1	R_2	$\cdots$	R_n
R_1	1	1/3	$\cdots$	2
R_2	3	1	$\cdots$	4
$\vdots$	$\vdots$	$\vdots$		$\vdots$
R_n	1/2	1/4	$\cdots$	1

3) Synthesis of priorities and the measurement of consistency

After all matrices are developed and all pair-wise comparisons are obtained, hierarchical synthesis is now used to weight the eigenvectors by the weights of the criteria and the sum is taken over all weighted eigenvector entries corresponding to those in the next lower level of the hierarchy. Eigenvectors or the relative weights (the degree of relative importance amongst the criteria), global weights and the maximum eigenvalue (λ_{max}) for each matrix are then calculated using Expert choice software (Expert Choice, 2000).

The consistency is determined by using the eigenvalue, λ_{max}, to calculate the consistency index, CI as follows:

$$CI = \frac{\lambda_{max} - n}{n-1}$$

Where n is the matrix size. Judgment consistency can be checked by taking the consistency ratio (CR) of CI with the appropriate value in Table 4. The CR is acceptable, if it does not exceed 0.10. If it is more, the judgment matrix is inconsistent. To obtain a consistent matrix, judgments should be reviewed and improved.

Table 4 Average random index (RI) based on matrix size

Size of matrix	1	2	3	4	5	6	7	8	9	10
Random consistency	0	0	0.52	0.89	1.11	1.25	1.35	1.40	1.45	1.49

Through using a pair-wise comparison procedure to set priorities for all five criteria (T, Q, C, S, and R) in terms of importance of each in contributing to the overall goal, We combine the criterion priorities and the priorities of each decision alternative priorities and the priorities of each decision alternative relative to each criterion in order to develop an overall priority ranking of the decision alternative which is termed as the priority matrix. The calculations for finding the overall priority of all resources are selected. The resources are ranked according to their overall priorities, highest priority resource is chosen to schedule.

5. Conclusion

As the complexity of the business process and resources, it becomes more important and urgent to provide MG workflow to construct and manage the manufacturing activity in the MG environment. Because manufacturing resources have their special particularities and resources are adding and removed all the time in MG environment, workflow engine should dynamically discovery and invoke the manufacturing resources according to activity requirements. So, we presented a dynamic scheduling architecture based on QoS in this paper. And, when multiple resources that meet the QoS requirements are available to an activity at a time, workflow scheduling can select the 'best match' resource according to scheduling algorithms based on AHP. The goals of our scheduler are to achieve better Quality of Service (QoS) performance during the application, such as higher product quality and service, and lower failure rate, time-to-market, and cost, etc.

Now we are developing MG workflow middleware including the scheduling functions for 'Shanghai High Institutions Grid" project. In future work, we plan to formalize the MG workflow description language about the parameters of QoS, which is based on XML.

6. References

[1] Liu Lilan,Yu Tao, Liu Shi Zhanbei, etc: Self-organization Manufacturing Grid and Its Activity Scheduling Algorithm. Computer Integrated Manufacturing Systems(2003)

[2] E.Deelaman, .Blyth,Y.Gil, Carl Kesselman: Workflow Management in GriPhyN. Chapter in the Grid Resource Management: State of the Art and Future Trends. Kluwer Publishing (2003)

[3] G.E. Graham,D. Evans and I. Bertram: McRunjob: A High Energy Physics Workflow Management in Grids with application onto Position Resolved Diffraction. Midwest Software Engineering Conference. Chicago. DePaul University, 6 (2003)

[4] G. von Laszewski, N. Zaluzec, M. Hatgan, et al: GridAnt: Client Side Workflow Management in Grids with application onto Position Resolved Diffraction. Midwest Software Engineering Conference. Chicago. DePaul University, 6 (2003)

[5]Liu Lilan, Yu Tao, Shi Zhanbei. Research on Rapid Manufacturing Grid and Its Service Nodes [J]. Machine Design and Research, 2003, 19 (5): 57-59. (in Chinese)

[6] Rashid Al-Ali, Kaizar Amin, Grregor von Laszewski. An OGSA-based Quality of Service Framework, The Send International Workshop on Grid and Cooperative Computing, (2003) 1075-1087

[7] A. Sheth, J. Cardoso, J. Miller, K. Koch, QoS for Service-oriented Middleware (2002) "Web Services and Grid Computing," Proceedings of the Conference on Systemics, Cybernetics and Informatics, Orlando, FL, July 2002.

[8] Arthanari T S, Dodge Y. M athemat ical Programming in Statistics. New York: JohnWiley & Sons,1981: 35～52

[9] Information Services in the Globus Toolkit 3.0 Release [EB/OL]. http://www.globus.org/mds.

[10] X. N. Chu, S. K. Tso, W. J. Zhang, etc. Partnership Synthesis for Virtual Enterprises. Int. J. Adv Manuf Technol. 2002, 19: 384-391

[11] Saaty TL. The analytic hierarchy process. New York: McGraw-Hill, 1980.

[12] Saaty TL. Decision making for leaders. Belmont, California: Life Time Leaning Publications, 1985.