A Digital Archives-Based Active Service Framework for Scientific Workflow System

Rui Wang

**School of Management Science and Engineering, Shandong University of Finance and Economics
Jinan, 250014, China**

Abstract

This paper describes the architecture of active service model of scientific workflow system. Digital archives technologies are available for all the resources needed to construct services specific group. This paper reviews the introduction and motivation for active service approach, describes the architecture of scientific workflow system, discusses the technologies used in active service , which uses digital archives technology to enable multi-service organizations achieve scientific workflow goals.

Keywords: *Active service; Digital archives; Scientific workflow system*

1. Introduction

Workflow management systems (WfMS) provide the automation of science processes where a collection of tasks is organized between participants according to a set of defined rules to accomplish some science goals [1-2]. Often, traditional WfMS can only coordinate workflows and their enacting agents within a single organization. However, advanced WfMS can now interact with various types of distributed agents over the Internet [3-4]. A science process can be viewed as a process to solve a scientific problem with goals such as knowledge discovery, knowledge innovation, and so on. A large-scale complex science process needs to handle very complicated logics in scientific problems.

Although many research projects are exploring the problems about computation- and data-rich scientific collaboration., there have been very few attempts to investigate the architectural issues in scientific workflow system. Not only the number of organizations in the workflow cannot be easily predicted, but the available services may change over time, too. Moreover, it is desirable that services are provided in a user adapted and context-adapted way [5-6]. science processes are more data-centric and knowledge-intensive and, active services will provide customized services to better satisfy the needs of every individual participants. Active services also enable science processeshave highly creative, innovative and dynamic.

We present architecture for a scientific workflow system management offering active services, which uses digital archives technology to enable multi-service organizations achieve scientific workflow goals.

2. Scientific Workflow System

A scientific workflow system contains a specification of a science goal. For a given condition regime, the goal set G will comprise the set $\{g_0, g_1, \cdots g_n\}$, where each g_i will have a relationship to one or more other g_k. A part of certain workflow goal is implemented by one migrating agent regarded as a service consumer. Attributes include Goal_id, a set condition on goal achievement, and a validity time and cost.

Alliance develops an architecture that facilitates fault tolerant cooperative control of heterogeneous organizations towards a common workflow goal. By negotiation between organization partners, we mean all the conversation acts made between a requester partner need a workflow service and one or several selected provider partner(s) able to provide the requested service. Alliance is robust, reliable and flexible in dynamic and unpredictable environments. Attributes include Alliance_id, a set of organization, lifecycle.

Science organizations offer services to the workflow agents who visit them. Each organization has a service agent as a leader and may provide one or more services. Every service is a representation of a goal type that can be handled by this organization. Attributes include organization name, one anchorage (the place of migrating agents moving to), one or more workstations.

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 2, No 1, March 2013
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org

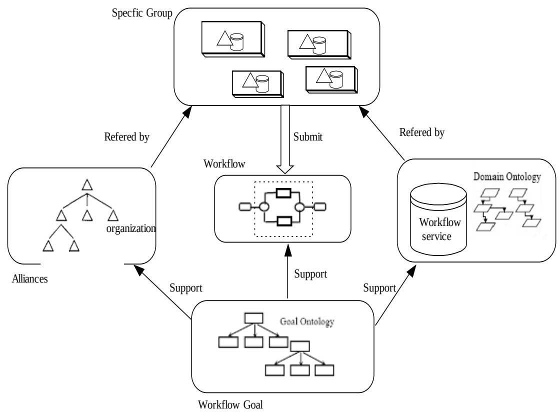

Figure 1. A goal-oriented scientific workflow system

A specific group is a set of service agents that also determines a virtual interaction space: a service agent may communicate with another service agent only if they belong to the same group. The specific group gives prominence to the core competitive and cooperative ability of each member organization, and optimizing the value chains of group, then mapping a collection of comparable or related workflow services to a single logical service.

A workflow service is a computational entity which is able to achieve a service requestor goal, having different criteria (due time, price, visibility of the service evolution and way of executing the service). Our active service model is based on OWL-S, which is a set of ontology definitions for describing service behaviors.

The active service agent is just another site, but provides only one service: s_info, which returns a list of service descriptions that match (exactly) the service name. On the one hand, the active service agent silently records histories of migrating agents' interaction and builds profiles by observing their behavior. On the other hand, the active service agent aggregates and manages multiple service organizations as a single virtual service.

3. Active Service Model

Active service mechanism provides a distributed environment where multiple application organizations can coexist in virtual or physical resources, such that migrating agents are unaware of the complexities inherent to workflow-computing. The active service model makes extensive use of virtualization technology for the creation of dynamic specific groups of virtual services that can be aggregated on-demand.

Active service is based on matching abstracted workflow goal descriptions with semantic annotations of workflow services. This process can only happen on an ontological level. In a distributed environment, different users and web services may use different terminologies, which lead to the need for mediation in order to allow heterogeneous parties to communicate. Given the previous assumption, we can optimistically assume that a mapping has already been established between the used terminologies. For this, the concrete migrating agent need has to be generalized to more abstract goal descriptions, and concrete services and their descriptions have to be abstracted to the classes of services that the active service agent can provide.

3.1 Acquiring Goals

The first task of the active service agent is to identify data consumers' needs and wants accurately, then to provide information of workflow organizations that will satisfy them. For forecasting to be successful, it is not sufficient to merely discover what service customers require, but to find out why it is required. Only by gaining a deep and comprehensive understanding of clients' state and workflow context can active service's goals be achieved. Therefore, active service agent attempt to visit the workflow engine which is the core of workflow management system to acquire migrating agents desires.

3.2 Building Specific Group

A service specific group maps a collection of comparable or related services to a single logical service. It is managed by the active service agent consisting of a service entry component, a rule-based policy component, a service catalog, and a service-rendering component. Building service specific group's objective is not to define new application programming interfaces (APIs) or new standards, but to construct from the existing organizations a new, higher-level structure that can hide complexities from service consumers, simplify deployment for service supplies.

Figure 2 shows the topology of a logical active service hierarchy. It is a simple topology that consists of a main specific group node and few secondary organization nodes. Service workstations of various types are provided through the secondary organization nodes. The service organizations that are similar are aggregated on the main group node and presented as virtual ports. Each specific group has a set of policy rules that governs its operations and a service catalog (or registry) that stores the information collected about its services. In the second layer, workflow organizations are instantiated as services which can be connected as needed to create virtual service group. This layer decouples the process of using and composing services from that of managing the execution of the underlying workflow applications.

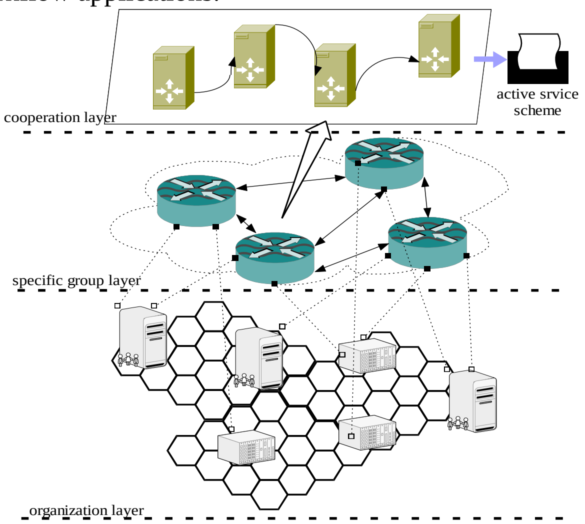

Figure 2. A logical active service hierarchy.

Let $SWI=(O,E)$ be an information system of the science workflow system, where O is a non-empty finite set of organizations and E is a non-empty finite set of effect such that $e:O\rightarrow S_e$ for every $e\in E$, S_e is the service *set* for effect *e*. With any $EF\subseteq E$, there is an associated similar relation:

$$SIM(EF)=\{(x,y)\in Or^2 \mid \forall e\in EF, s(x,e)=s(y,e) \text{ or } s(x,e)=\sigma \text{ or } s(y,e)=\sigma\}$$

If $(x,y)\in SIM(EF)$, x and y are similar by effect from *EF*. The similar class of the EF-similarity relation is denoted SG(Specific Group). $O/_{SIM(EF)}$ defines a service group granule, then this is not partition but overlay of O.

3.3 Organization Using Digital Archives

Three new digital archives and a selection mechanism are devised for organizations of specific groups. Competing operator and cooperating operator realize competition and cooperation among organizations. Self-learning operator increase the energy of organizations by knowledge.

On the basis of effect significance, matching degree of organization is calculated according to formula:
Given that $o\in$ organization and $r\in$ rule, $|condition_r|$ denotes the number of <*condition*> in the part of *r*, and $|condition_r^o|$ denotes the number of condition satisfied by organization o.

The matching degree value:
$MD_r^o=|condition_r^o|/|condition_r|$

According to the definition, the range of the match value is [0,1], while 0 is the worst case and 1 the best case. The rule with the maximum match value is used to predict the optimal organizations in this specific group.

- Competing digital archives

Performing competing operator on organizations in one specific group is to find the max matching degree of service organization. If org_i satisfy $MD(org_i)>=MD(org^{\max})$, org_i is recorded as winner, and $MD(org^{\max})=MD(org_i)$.

- Cooperating digital archives

Given that two parent organizations, $org_1=\{p_1,p_2,\cdots,p_n\}$ and $org_2=\{q_1,q_2,\cdots,q_n\}$. If they do not achieve goal respectively, then one child org^v is determined by cooperating strategy to expand influence and enforce energy:

- Self-leaning digital archives

All organizations in one service group could not satisfy the goal of conditions. Organizations improve their matching degree through self-learning from other alliance's group to.

4.Active Service Process in Scientific Workflow

This section presents an active service process that uses digital archives to solve clients lacking of the perfect knowledge of all the details. In addition, this is a more flexible way providing workflow with the isolation property.

When a new scientific workflow goal is confirmed, a scientific workflow engine must initiate utilities to activate appropriate applications for the execution of desired goal and decompose goal then derive migrating instances in order to achieve sub-goals, as shown in steps 1 and 2 of Figure 2. Service agents register them with an active service agent and the scientific workflow engine submits the migrating instances' identity and desire to the active service agent as shown in steps 3 and 4. Then, the active service agent find possible service agent according to the demand. The active service agent searches for a recommendable service organization alliance and returns the results to the migrating instance as shown in steps 5 and 6. The migrating instance selects proper service organizations recommended by the active service agent and obtains bids all the designated service agents for moving to as shown in steps 7-8. The service agent grants the request as shown in step 9 of the figure. Finally, as in step 10, the migrating instance transfers its code (specification) and its execution state to the anchorage server, negotiates a service to be executed, and moves on.

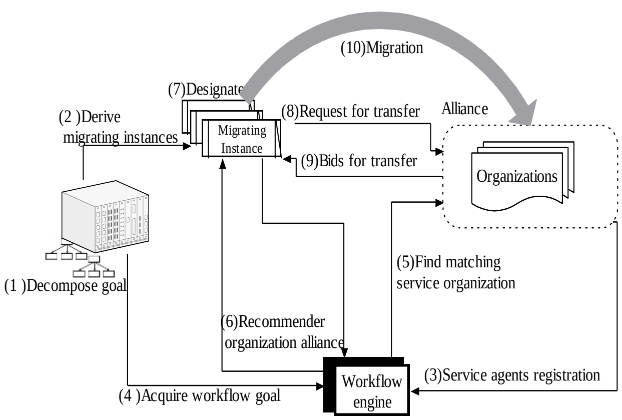

Figure 3. A service recommender process

5. Performance Evaluation

To evaluate service recommendation based on participation matrix contract net (SRPMCN), we use F_1 which is a well-known performance measure for recommender system [7-8]. We start by dividing our data sets into two parts: the training set and the test set. SRPMCN works on the training set first, and then generates a set of recommended services, called service recommendation set. F_1 integrates recall and precision and is given by:

$$F_1 = \frac{2rp}{r+p} \quad (1)$$

where r is recall, counting the ratio of the number of services correctly assigned to the test set t to the total number of services belonging to the test set t; and p is precision, counting the percentage that services assigned to the test set t actually belong to the group service recommendation set g. For recall and precision, there exists contradiction between them. Emphasizing one side alone will lead to unacceptable low value in the other side. So F_1 metric is widely used as an optimization criterion for binary decisions. For a good service recommendation, the F_1 value should be high.

This section presents experimental results performed by different parameter set of SRPMCN. And the performance of SRPMCN is compared to that of the benchmark system. The following parameters are selected, which may affect the quality of recommendations generated by SRPMCN: the size of service population and the global minimizing strategies. We observed the sensitivity of each of those parameters and, used the optimal values for the rest of the experiments.

5.1 Impact of Population Size

The population size has a significant impact on the recommendation quality. To determine the sensitivity of population size, we performed an experiment where we varied the number of population members' from10 to 30 and computed the corresponding F1 metric. Our experimental results are shown in Fig. 4. Looking into the results, we can conclude that the size of the population members does affect the quality of service recommendation. Generally, the recommendation quality increases as the number of population members increases. However, after a certain peak, the improvement gains diminish and the quality becomes worse. In our experiment, the peak was reached at the area between 10 and 15.

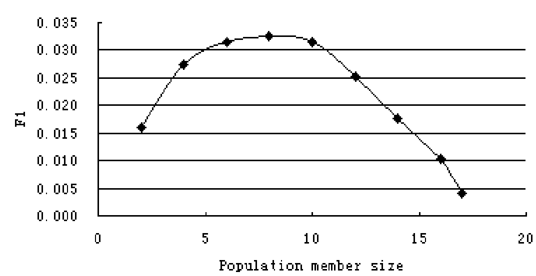

Figure 4. Impact of Population Member Size on Recommendation Quality

5.2 Impact of Global Minimizing Strategies

In order to evaluate the impact of global minimizing strategies on the recommendation quality, we performed an experiment with the annealed version of the Nelder&Mead strategy (ANM)[9] which is appealing because of its adaptive scheme for generating random parameter deviations. We tried to optimize each of recommendation by experimenting to find the SRPMCN provided fastest and smoothest convergence. Fig. 4 shows the comparative results obtained based on these two methods. Looking into the results, we can see that the recommendation quality using SRPMCN is better than that of ANM. This indicates that PM usually converged faster, especially in the more difficult cases to lead to the better quality of recommendations.

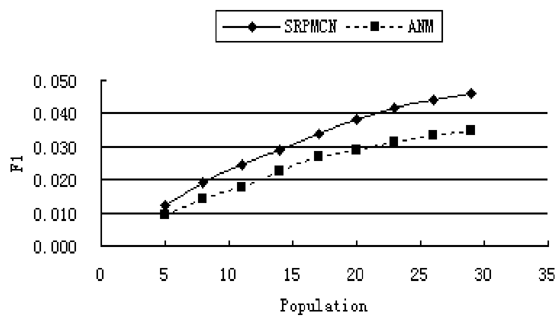

Figure 5. Impact of Global Minimizing Strategies on the Recommendation Quality.

6. Conclusion

This paper motivates the need for an architectural framework of active services in a heterogeneous workflow environment. In this paper, we have proposed a goal-oriented scientific workflow system based on the mobile agent technology. We introduced the concept of service specific group which mapping a collection of comparable or related workflow services to a single logical service. We discuss a way to do active service by allowing runtime selection, integration and coordination of distributed resources and also accommodate dynamic science requirements.

Acknowledgment

A Project of Shandong Province Higher Educational Science and Technology Program : J10LG29.

References

[1] D. Georgakopoulos, M. Hornick, and A. Sheth, "An overview of workflow management: From process modeling to workflow automation infrastructure," Distributed and Parallel Databases, vol. 3, pp. 119-153, 1995.

[2] M. A. de, A. W. van, and A. J. Weijters, "Workflow Mining: Current Status and Future Directions," in Lecture Notes in Computer Science, vol. 2888, Lecture Notes in Computer Science, 2003, pp. 389-406.

[3] W. L. Seng and Z. Arkady, Towards Distributed Workflow Enactment with Itineraries and Mobile Agent Management, 2011.

[4] J. Cheng, G. Zeng, "An agent-oriented approach to process partition and planning in migrating workflow systems " Engineering Applications of Artificial Intelligence, vol 25, Issue 4, 2012, pp. 837-845.

[5]H. Maus, "Workflow Context as a Means for Intelligent Information Support," in Lecture Notes in Computer Science, vol. 2116, Lecture Notes in Computer Science, 2001, pp. 261-267.

[6]X. H. Le, T. Doll, "An enhancement of the Role-Based Access Control model to facilitate information access management in context of team collaboration and workflow," Original Research Article Journal of Biomedical Informatics, Volume 45, Issue 6, 2012, Pages 1084-1107.

[7]J. S. Breese, D. Heckerman, and C. Kadie, "Empirical analysis of predictive algorithms for collaborative filtering," Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence, vol. 461, 1998.

[8]B. Sarwar, G. Karypis, J. Konstan, and J. Reidl, "Item-based collaborative filtering recommendation algorithms," Proceedings of the 10th international conference on World Wide Web, pp. 285-295, 2001.

[9] C. Peter, E. Terry, "Parameter sensitivity study of the Nelder–Mead Simplex Method,"Advances in Engineering Software, Volume 42, Issue 7, 2011, Pages 529-533

WangRui received the Master degree in 2005 and the Doctor degree in computer application in 2009 both from the Shandong University, She has been employed at Shandong University of Finance and Economics since 2000, and currently, she is a professor at Shandong University of Finance and Economics; She has been supported by the Shandong province higher educational science and technology program. Her articles published are over 20. His interests are in Intelligent Information Processing and Scientific Workflow.