Multi-User Tool for Scientific Work Flow Composition

Sheksha Vali[#1], Srikanth Sreerama[*2]

[#1]*M.Tech, Computer Science Engineering, MLRIT, Hyderabad, Andhra Pradesh, India*
[#2] *Associate Professor, Department of CSE, MLRIT, Hyderabad, Andhra Pradesh, India*

Abstract- Scientists with diverse expertise and located in different geographical locations may work together in order to achieve desired results. This collaboration is required in modern society as skill set required are spread multiple scientists. In order to utilize the diversified skills associated with human experts, it is essential to have a mechanism to share the skills and work in scientific inventions in collaborative fashion. In this context it is important to have a scientific workflow for automated communication and data management among the scientists involved in the collaboration. Many existing work benches are providing single user flows. They lack in providing multi-user environment where many scientists can work collaboratively. Recently Zhang et al. proposed a novel tool that facilitates collaborative scientific workflow among multiple scientists. The tool allows collaborative scientific workflow among scientists. In this paper we implement such tool practically as a prototype. The experimental results revealed that the tool is effective in providing workflow for scientific collaborations.

Index Terms – Service oriented architecture, collaborative computing, and scientific workflow

I. INTRODUCTION

Modern science is growing in fast pace. It has produced large volumes of data. Such diversified data needs to be used by multiple scientists with different expertise in order to obtain desired results. Processing huge amount of scientific datasets is not beyond the scope of one individual scientist. Many scientists with different skills are to collaborate in order to achieve tasks assigned [1]. Scientists from multiple domains need to work together to make any task successful. Huge amount of data is being produced by various satellites and applications continuously. Working with such data in collaborative fashion is the need of the hour. Dataflow-oriented processing is required for scientific data analysis. Such process is known as scientific workflow. The scientific workflow is not similar to business work flow as they are data flow oriented and streamline a set of scientific tasks in order to produce desired results [2], [3]. Scientists make use of such scientific workflows in order to integrate local and remote data sources to process in collaborative fashion [4], [5], [6], and [1].

There are two important techniques such as scientific collaboration and scientific workflow. The convergence of these two techniques will naturally lead to collaborative scientific workflow [7]. There are existing scientific workflow tools but they do not support collaborative composition. They include VIEW [8], [9], Swift [10], Pegasus [3], VisTrails [11], Traina [12], Taverna [13], and Kelper [14]. All these tools provide scientific workflow with a limitation. The limitation is that they support only a single user environment.

In this paper we build a tool for supporting scientific workflow in multi-user environment. The remainder of this paper is structured as follows. Section II reviews relevant literature. Section III provides description of the proposed approach. Section IV provides tool implementation details. Section V presents experimental results while section VI concludes the paper.

II. RELATED WORK

In this section we review literature with respect to collaborative workflow composition, business workflow coordination, and scientific workflow management tools. However, the main focus of this paper is workflow composition. Service oriented middleware architecture was proposed by Balasooriya et al. [15], [16] for composing workflows using a two layer framework. Work flow management in grid environment is proposed in [17], [18]. For each task in scientific workflow post and pre conditions are studied by Lu and Sadiq [19]. The general purpose collaborative design problem was studied in [20] and [21]. They explore collaborative business processes with human interactions.

With respect to collaborative workflow composition, it is understood that only system to system integration or program to program integration is not sufficient in truly business environment. The business community recognized the need for human experts to involve as part of business process automation. Thus human centric workflows came into existence. Such workflows provide realistic workflows that are useful in the real world enterprise applications that run in distributed environment. One of the examples for workflow model is PBEL4People [22] which allow seamless integration among multiple services and do not enable human interaction. However, these collaborative processes are not adequate for scientific collaborations. This is because the scientific workflows and business workflows are fundamentally different.

The requirements of scientific workflows are different. For instance BPEL4People can't provide suitable solution to scientific workflows because it does not model to include human interactions. The reason behind this is that the BPEL process is made up of multiple web services that provide program to program communication. Web services run in heterogeneous environment but they are programs which can be run from programs written in any programming language. They are basically meant for integrating multiple homogenous and heterogeneous systems. However, they can provide interoperability among different programs and platforms. BPEL is made up of multiple web services that run in a pre-defined workflow. In this paper we focus on building a tool that provides scientific workflows in collaborative fashion.

III. PROPOSED WORKFLOW COMPOSITION MODEL

The proposed scientific workflow model assumes complex workflow among multiple scientists in collaborative fashion. The collaborative workflow is presented as ontology as it can represent knowledge in the form of concepts and relationship among the concepts. Sample ontology that models collaborative workflow composition is presented in figure 1.

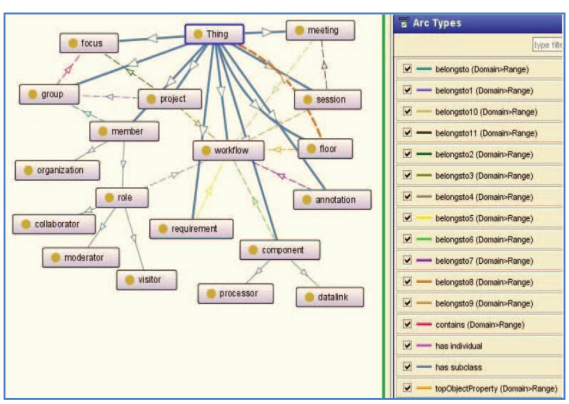

Fig. 1 – Sample ontology representing composition of collaborative workflow (excerpt from [23])

As seen in figure 1, it is evident that the workflow contains many concepts and relationships among them. This kind of knowledge representation is known as ontology. In figure 1 there are two parts. The left part shows ontology while right part shows legend that illustrates the relationship among concepts. Scientific workflows contain organized tasks, channels and comments. Meeting represents long term collaboration. Short term collaboration for synchronization represents a session.

Collaborative Composition Model

Ontology is our basic foundation for making collaborative composition model. We also made it a file based workflow that can be serialized. The ontology is saved in the form of an XML file which can be altered programmatically. A sample workflow which is in the form of an XML file is presented in figure 2.

```
<schema>
  <complexType name="Workflow">
    <sequence>
      <element name="dataflow" type="tav:Dataflow"
maxOccurs="unbounded" minOccurs="1"/>
...
  </complexType>

  <complexType name="Dataflow">
    <sequence>
      <element name="name" type="string"/>
      <element name="inputPorts"
type="tav:AnnotatedGranularDepthPorts"/>
      <element name="outputPorts" type="tav:Ports"/>
      <element name="processors" type="tav:Processors"/>
      <element name="conditions" type="tav:Conditions"/>
      <element name="datalinks" type="tav:Datalinks"/>
      <element name="annotations" type="tav:Annotations"
maxOccurs="1" minOccurs="0"/>
    </sequence>
  </complexType>
</schema>
```

Fig. 2 Sample workflow (excerpt from [23])

As can be seen in figure 2, the XML content is a schema which is related to a workflow. The advantage of using an XML file is that it can be altered or customized programmatically. The collaborative composition model used by us provides many possible scenarios. The schematic view of the workflow is visualized in figure 3.

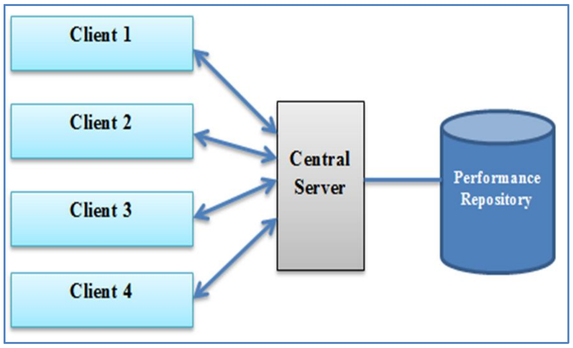

Fig. 3 – Schematic overview of workflows

As can be seen in fig. 3, it is evident that there is much collaboration and many groups of people are interactive with others through a central server. The conversations are being stored in a repository. As part of advanced collaboration model two algorithms are used. The first algorithm is known as floor granting algorithm, while the second algorithm is known as floor releasing algorithm.

```
Algorithm 1: Floor Granting Algorithm
Input: A collaborator releases a floor
Requirements: Release a floor.
   1: Check the waiting list.
   2: if waiting list ≠ empty then
   3:    get the top requestor of the waiting list
   4:    floor owner ← requestor
   5:    notify all members
   6:    remove the top requestor from waiting list
   7: else if waiting list is empty then
   8:    floor flag ← unoccupied
   9:    notify all members
   10:endif

Algorithm 2: Floor Releasing Algorithm
Input: A requestor requests a floor
Requirements: Decide whether a floor should be granted.
   1: Check floor flag.
   2: if floor ≠ taken, then
   3:    floor flag ← occupied
   4:    floor owner ← requestor
   5:    notify all members
   6:    return true
   7: else if floor = taken then
   8:    add requestor to waiting list
   9:    return false
   10: endif
```

Fig. 4 – Floor granting and releasing algorithms (excerpt from [23])

As can be seen in fig. 4, in case of floor granting algorithm, when the floor is not occupied, the floor will be granted to the requester. On releasing the floor the requester who is on the waiting will get control over the floor. In this model collaboration process and workflow is controlled. More technical details can be found in [23].

IV. EXPERIMENTAL RESULTS

We did experiments with our tool that demonstrates proof of concept. The experiments are made in terms number of collaborators, number of updates per minute, and tasks involved. The experimental results are presented in the form of a series of graphs.

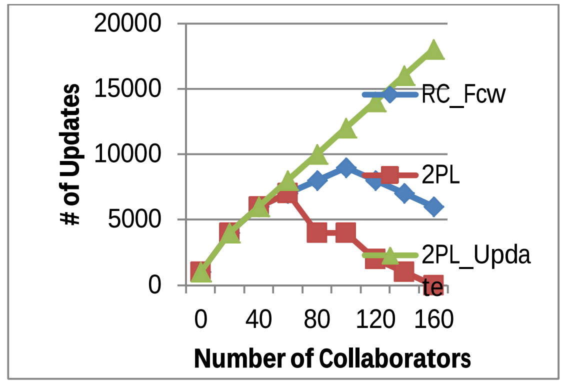

Fig. 5. Comparison of successful update rates.

As shown in the above figure represents the horizontal axis represents number of collaborators while vertical axis represents number of updates

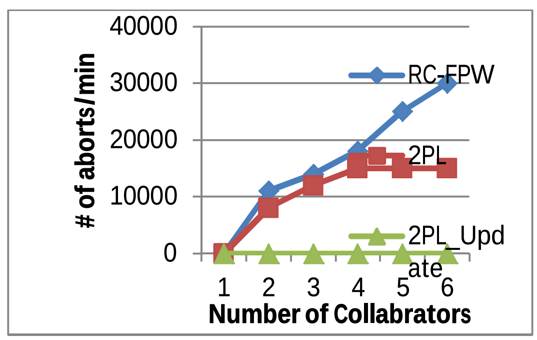

Fig. 6. Failed task update rates under numbers of collaborators.

As shown in the above figure represents the horizontal axis represents number of collaborators while vertical axis represents number of aborts/min.

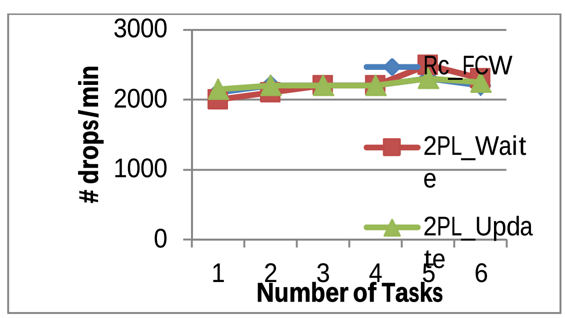

Fig 7 Comparison of successful update rates

As shown in the above figure represents the horizontal axis represents number of tasks while vertical axis represents number of drops/min.

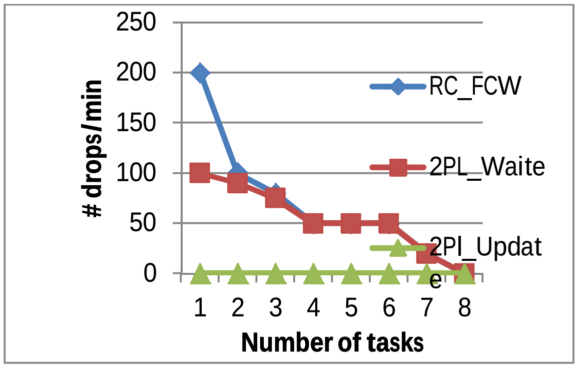

Fig 8 Comparison of successful update rates

As shown in the above figure represents the horizontal axis represents number of tasks while vertical axis represents number of drops/min.

V. CONCLUSIONS

In this paper, we study collaborative work among multiple scientists with diverse expertise located in different geographical locations. To make their work effective a tool is required that can facilitate automatic workflow communication and data management among the scientists. The existing workbenches are providing single user environment. They lack support for multiple scientists to have diversified workflow scenarios. In this paper we implement a tool that supports multiple scientists to have automated workflow and communication seamlessly in service oriented architecture. Empirical results revealed that the proposed tool is very effective in providing environment where multiple scientists can continue their work collaborative fashion.

REFERENCES

[1] G.M.Olson , A.Zimmarman and N.Bos, eds., Scientific collaboration o n the internet, 2008, MIT Press, Cambridge, MA, USA.

[2] B.Ludascher, “Scientific workflows: Cyberinfrastructure for e-Science”, in Proceedings of Pacific Neighborhood Consortium (PNC), 2007, Brekeley, CA, USA, Oct.19, PP.

[3] Y.Gil, E.Deelman, J.Blythe, C.Kesselman and H.Tangmunarunkit, “Artificial Intelligence and Grids:Workflow Planning and Beyond”, IEEE Intelligent Systems , Jan.-Feb., 2004, 19(1):PP.26-33.

[4] N.R.Council “Facilitating Interdisciplinary Research “. 2004, National Academis Press, Washington, DCA.

[5] S.Wuchty, B.Jones and B.Uzzi, “The increasing Dominance of Teams in Production of Knowledge”, Science 2007, 316: pp. 1036-1039.

[6] E.Deelman and Y.Gil , “NSF workshop on the Challenges of Scientific workflows”, (ed), May1-2, 2006.

[7] S.Lu and J.Zhang “Collabortaive Scientic workflows supporting Collaborative Science”, International Journal of Business Integration and Management(IJBPIM), 2011, 5(2):pp. 185-199.

[8] A.Chebotko, C.Lin , X.Fei, Z.Lai, S.Lu, J.Hua, and F.Fotohui, “VIEW: A Visual Scientific Workflow Management”, in Proceedings of the IEEE 2008 Internatioanl Workshop on Scientific workflows, 2007, Salt Lake City, UT, USA,Jul., pp.207-208.

[9] , C.Lin , , A.ChebotkoX.Fei, Z.Lai, S.Lu, J.Hua, and F.Fotohui, “VIEW: A Visual Scientific Workflow Management”, in Proceedings of the IEEE 2008 Internatioanl Conference on Services Computing(scc), 2008, Honolulu, HI, USA, Jul9-11., pp.335-342.

[10] Y.Zhao, M.Hategan, B.Clifford, I.Foster, G.Laszewski, V.Nefedova, I. Raicu, T.Step-Praun and M.Wilde, Swift:Fast, Reliable, Loosely Coupled Parallel Computation “, in proceedings of IEEE International workshop on Scientific workflows, 2007, Salt Lake City, UT, USA, Jul.9-13, pp.199-206.

[11] J.Freire, C.T. Silva, C.P.Challan, E.Santos and C.E.Scheidegger, “Managing Rapidly-Evolving Scientific workflow”, Lecture notes in Computer Science, May, 2006, 4145/2006: pp.10-18.

[12] D.Churches, G.Gombas, A.Harrison, J.Maassen, c.Robinson, M.Shields, I.Taylor and I.Wang, “Programming scientific and Distributed workflow with traina services”, concurrency and Computation: Practice and Experience , 2006, 18(10): pp.1021-1037.

[13] T.oinn, m.Greenwood, M.addis, M.N. Alpdemir, J.Ferris, K.Glover, C.Goble, A.Goderis, D.Hull , D.Marvin, P.li, P.lord, M.R.Pocock, M.senger, R.Stevens, A.wipat and C.Wroe, “Taverna Lessons in Creating a workflow environmentfor the Life Sciences”, concurrency and computation: Practice and Experience, 2006, 18(10): pp1067-1100.

[14] B.Ludascher, I.Altintas, C.Berkley, D.Higgins, E.Jaeger, M.Jones, E.A.Lee, J.Tao and Y.Zhao, “Scientific workflow management and the Kepler System” concurrency and computation: Practice and Experience, 2006, 18(10): pp1039-1065.

[15] J.Balasooriya, J.Joshi, S.K.Prasad and N.B.Savathe, “ A Middleware Architecture for Enhancing web services Infrastructure for Distributed coordination of workflows”, in Proceedings of IEEE Internatioanl Conference on Services Computing(scc), 2008, Jul., pp.370-377.

[16] J.Balasooriya, J.Joshi, S.K.Prasad and S.B.Navathe, " A Two Layered Software Architecture for distributed workflow Coordination web services ",", in Proceedings of IEEE Internatioanl Conference on web services (ICWS), 2006, Sep., pp.933-934.

[17] T.miller p. Mc burney,J.MCginnis,and k.sathis,"first_class protocol for agent_based coordination
Of scientific instrument "in nproceeding of 16th IEEE international work shops on enabling technologies:Infrastructure for colaberative enterprises {wetice},2007jun,pp:41-46.

[18] z.nemeth,c.perez,and T.priol, "distributed work flow coordination;melucules and reactions,;in proceeding of20th IEEE international parallel&distributed procesn in symposium,2006,apr.,pp.260-267.

[19] R. Lu and S.W. Sadiq, "A Survey of Comparative Business Process Modeling Approaches", in Proceedings of the 10th International Conference on Business Information Systems (BIS), 2007, Poznan, Poland, Apr. 25-27, pp. 82-94.

[20] C. Yen, W.J. Lin, and J.C Lin, "A web-based collaborative, computer -aided sequential control design tool", IEEE Control Systems Magazine,Apr,.2003,23(2):pp. 14-19.

[21] D. Dori, D. Beimel, and E. Toch, "OPCATeam - Collaborative Business Process Modeling with OPM", in Proceedings of 2nd International Conference on Business Process Management (BPM), 2004, Potsdam, Germany, Jun. 17-18, Springer-Verlag Berlin Heidelberg, pp. 66-81.

[22] A. Agrawal, M. Amend, M. Das, M. Ford, C. Keller, M. Kloppmann, D. Konig, F. Leymann, R. Muller, K. Plosser, R. Rangaswamy, A. Rickayzen, M. Rowley, P.Schmidt, I. Trickovic, A. Yiu, and M. Zeller, "WS-BPEL Extension for People (BPEL4People), Version 1.0", 2007 Jun., Accessed on Available from:
http://download.boulder.ibm.com/ibmdl/pub/software/specs/ws-bpel4people/BPEL4People v1.pdf.

[23] Jia Zhang, Daniel Kuc, and Shiyong Lu, "Confucius: A Tool Supporting Collaborative
Scientific Workflow Composition". IEEE TRANSACTIONS ON SERVICES COMPUTING, MANUSCRIPT ID, 2011.

AUTHORS

Sheksha Vali, he is pursuing M.Tech (CSE) MLRIT, Hyderabad, AP, INDIA. He has received B.TechComputer Science and Engineering in the year 2010. His main research interest includes wireless sensor networks, Web designing and Databases

SrikanthSreerama. He is currently with the Department of Computer Science and Engineering, MLRIT, Andhra Pradesh, India. He is having 7 years of teaching experience**.** His main research interest includes Cloud Computing and Data Mining.