Workflow Management Systems for Grid Computing

By Carmen Bratosin and Wil van der Aalst

The group Architecture for Information Systems from the Technische Universiteit Eindhoven (TU/e) in the Netherlands built up extensive knowledge in the field of Workflow Management Systems and Process Mining. Starting in 2006, the group is trying to apply this knowledge in a new and dynamic research area: Grid Computing. Four research perspectives are currently under investigation.

Software systems are getting more and more complex. To cope with their complexity, one often divides a system in a number of autonomous components whose work is coordinated. Component and service coordination is one of the main challenges in software engineering. Two important application fields of coordination are grid computing and workflow management. Grid computing is mostly used in computational science while workflow management is used for business applications. Therefore, we try to bridge the gap between these areas in order to make further progress in both of them.

Over the last decade we have gathered much experience in process modeling, analysis and enactment. Our workflow patterns have become a standard way to evaluate languages and the workflow management system YAWL is one of the most expressive and mature open-source workflow systems available today. Moreover, we have been specializing in process analysis. Using Petri nets as a theoretical foundation, we have been able to analyze a variety of real-life process models ranging from BPEL and workflow specifications to the entire SAP reference model. In recent years, we focused on the analysis of processes based on system logs. The ProM framework developed at TU/e provides a versatile toolset for process mining, which seems to be particularly useful in a grid environment.

So far the grid computing community focused more on the infrastructure. Grid software has been designed that allows users to submit their ‘problems’ to the grid. Less work has been done to model such a problem in an efficient way. Also, in most applications the correctness properties rest in the hands of the user.

Therefore, we are applying our knowledge related to Petri net modeling and analysis, the workflow patterns, process mining, and concrete workflow technology to grids. We are doing research in the following areas:

(a) Grid modeling

Many definitions exist of grids and often technological aspects and hyped terms are hiding the essence of grids. Therefore, we use a mixture of Petri nets and UML modeling to build formal/conceptual models for grid computing. Here we emphasize the link between the distributed nature of grids (where resources play an important role) and workflow processes. The main purpose is to formalize the concept of a grid and fix a particular interpretation while highlighting the interesting research questions.

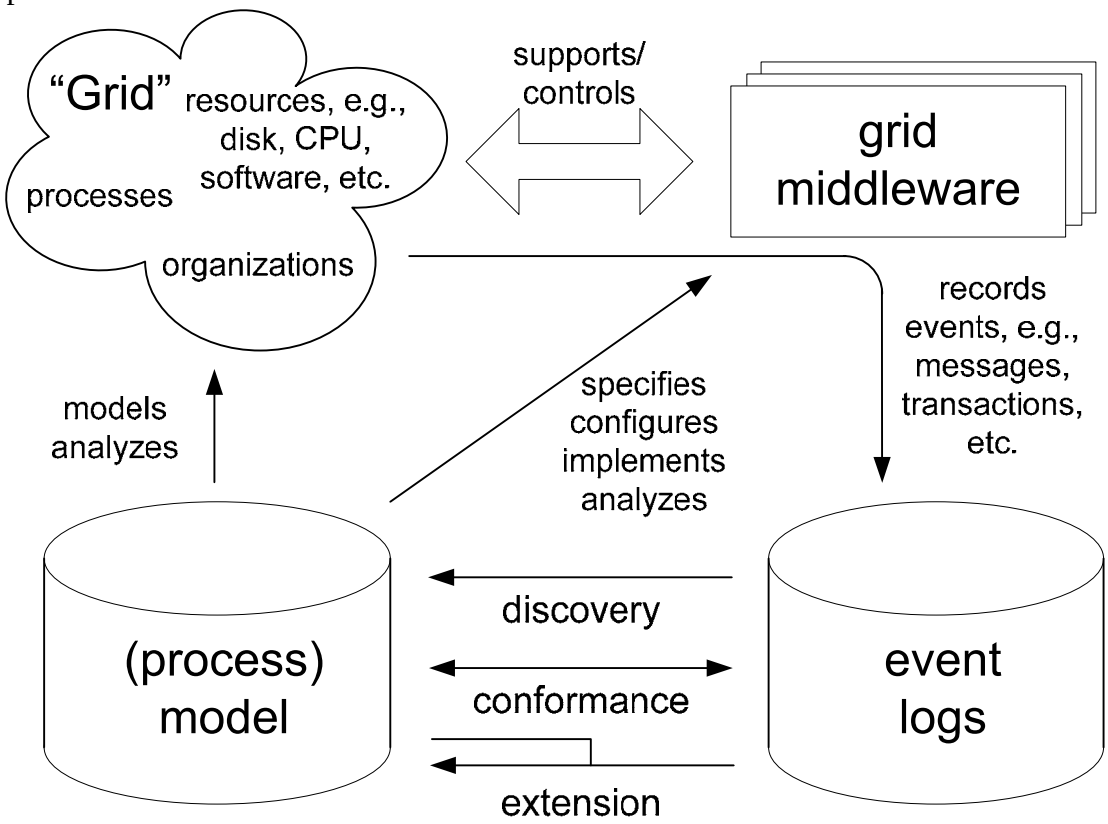
(b) Analyzing grid models

Using Petri-net-based techniques, we analyze different mechanisms used in grid workflows. The goal is to transfer correctness notions such as soundness to grid workflows. Also, we try to find new properties based on the specific grid behavior (e.g. multiple instances of the same process, resource allocation, and distributed management).

(c) Analyzing grid logs

In a grid environment many events are logged and the performance of the system is of the utmost importance. Therefore, it is interesting to apply process mining techniques in this domain. This assists in the configuration of grids and may be used for the on-the-fly optimization of processes.

(d) Building a process-aware grid infrastructure
Using combination of Globus, YAWL, and ProM we want to realize a more "process-aware" grid. By linking a fundamental enabling technology for the grids (Globus) to a powerful process engine (YAWL) and state-of-the-art analysis tools (ProM), we obtain an interesting environment for experimentation.



The diagram above illustrates the scope of the project. On the one hand, we analyze grids by modeling them in terms of Petri nets. Similar models are used for the configuration of the process-perspective of grid middleware (in our case a mixture of Globus and YAWL). On the other hand, we collect event logs via the middleware layer and use this for process mining. We use these event logs for process discovery (i.e., automatically deriving models by observing the grid), for conformance checking (to check whether "the grid" behaving as expected), and model extension (e.g., to project performance indicators onto a process model).

All of the aspects shown in the diagram have extensively investigated in the context of workflow management systems and service oriented architectures using BPEL engines. For example, we have been doing conformance testing in the context of Oracle BPEL and process discovery and process verification in the context IBM WebSphere. We have also evaluated many process engines using the so-called workflow patterns and provided semantics and analysis techniques for process modeling languages ranging from BPEL and YAWL to BPMN and EPCs. The next step is to apply this in a grid environment using both Globus and YAWL.

The research is supported by Netherlands Organization for Scientific Research (NOW) in the context of project *Workflow Management for Large Parallel and Distributed Applications*. In this project TU/e participates with the group of Professor Farhad Arbab from CWI. The project started in 2006 and the duration is 4 years.

Useful Link(s) www.workflowpatterns.com
www.processmining.org
www.yawl-system.com
www.win.tue.nl/ais
www.globus.org

www.nwo.nl/nwohome.nsf/pages/NWOA_6R7FYJ

Please contact:
Carmen Bratosin
Eindhoven University of Technology, the Netherlands
Tel.: +31 40 247 5144
E-mail: c.t.bratosin@tue.nl

Process-aware grid infrastructure