Introducing Risk Management into the Grid*

Karim Djemame, Iain Gourlay, James Padgett
School of Computing
University of Leeds
United Kingdom
{karim,iain,jamesp}@comp.leeds.ac.uk

Georg Birkenheuer, Matthias Hovestadt, Odej Kao, Kerstin Voß
Paderborn Center for Parallel Computing
University of Paderborn
Germany
{birke,maho,okao,kerstinv}@upb.de

Abstract

Service Level Agreements (SLAs) are explicit statements about all expectations and obligations in the business partnership between customers and providers. They have been introduced in Grid computing to overcome the best effort approach, making the Grid more interesting for commercial applications. However, decisions on negotiation and system management still rely on static approaches, not reflecting the risk linked with decisions. The EC-funded project "AssessGrid" aims at introducing risk assessment and management as a novel decision paradigm into Grid computing. This paper gives a general motivation for risk management and presents the envisaged architecture of a "risk-aware" Grid middleware and Grid fabric, highlighting its functionality by means of three showcase scenarios.

1 Introduction

Grid research [17] has been ongoing for a number of years and Grid technologies have reached a high level of development [28]. However, shortcomings relating to risk and dependability remain an obstacle to adoption for commercial applications and services.

Risk management is a discipline that addresses the possibility that future events may cause adverse effects [15] and is important in a diverse range of fields such as statistics, economics, biology, engineering, systems analysis, and operations research. While risk is traditionally seen as a negative force, modern risk management recognises its positive aspects. In contrast to risk avoidance strategies, accepting certain risks could be also benficial. A typical modern-day example would be that of professional poker players, who have to evaluate the chances of win and of loss. They expect to make a long-term profit but with extremely high variance [30]. Analogous are day-traders who aim to profit by making statistically profitable decisions in stock market trading. Current state-of-the-art Grid computing does not incorporate risk management and current Grid infrastructures still follow a best-effort approach. This is insufficient for attracting commercial end-users to use the Grid. Envisage a scenario where a user is looking to pay a Grid resource provider (or providers) to execute his or her application. This application may consist of a single job or a workflow with a number of sub-jobs. Users negotiate for resource usage through a Grid resource broker which queries resource providers on their behalf to find suitable resources. They require a job execution with a desired level of priority and quality. For example they may have a deadline for the completion of their application which, if not met, may lead to financial loss. Consequently, users may want to negotiate for Service Level Agreements (SLAs) to define all aspects of the business relationship between themselves and the Grid resource provider(s). An SLA specifies the performance of the resource provider as well as a penalty fee which the provider has to pay if it does not perform as contracted. Accordingly, contracted performance and penalty fees are of particular importance when the user is paying for the resource usage.

A number of research projects address the issue of SLA negotiation in a Grid environment [29, 13]. However, providers are still cautious on adopting such a system, since

*This work has been partially supported by the EU within the 6th Framework Programme under contract IST-031772 "Advanced Risk Assessment and Management for Trustable Grids" (AssessGrid).

Proceedings of the Second IEEE International
Conference on e-Science and Grid Computing (e-Science'06)
0-7695-2734-5/06 $20.00 © 2006 IEEE

the agreement to meet the objectives specified in an SLA is a business risk. SLA violation can be caused by many events like network or resource failure or even operator unavailability. Without a means of assessing the risk of agreeing an SLA, providers are only able to make uncertain decisions regarding suitable SLA offers. Similarly, end-users would like to know the risk of an SLA violation by a Grid resource provider so that they can make appropriate decisions in relation to acceptable costs and penalty fees. A broker which is acting on behalf of a user to search for suitable resource providers may also require risk assessment mechanisms. Risk assessment enables the broker to evaluate the overall risk involved in mapping a workflow consisting of a set of possibly inter-dependent sub-jobs onto a number of resource providers. Further, the capability to evaluate the reliability of risk assessments presented by resource providers is important. This new functionality will be valuable for avoiding contracts with unreliable providers and significantly enhances the service of a broker. The goal of the AssessGrid project is to address the key problem of risk by introducing a framework for supporting risk assessment and management for all three Grid actors discussed here (end-user, broker and resource provider). The new system components are introduced in the architecture to support risk management. For an accurate risk assessment a consultant service supports the provider with statistical information. The ability to assess the risk associated with a specific SLA offer enables a resource provider to build a schedule according to risk information. In the Grid middleware the broker relies on a confidence service to assess the reliability of risk assessments received by resource providers. This is based on historical statistics relating to previous SLAs. A workflow assessor supports the broker by providing an assessment of risk for an entire workflow. By presenting the AssessGrid's system architecture and describing the future directions of this project, the idea of integrating risk management into the Grid is presented to the Grid community.

The paper is organised as follows: Related work is discussed in the following section. Section 3 presents key risk management concepts and their relationship to the Grid. This is followed in section 4 by a presentation of the system architecture, considering the end-user, broker, and resource provider layers. Section 5 presents three showcase scenarios to illustrate the usage of the system, highlighting requirements and demonstrating component functionality. In section 6, conclusions are presented and the plans for proceeding with the research in AssessGrid are discussed.

2 Related Work

The introduction has described the idea and usage of risk assessment and management. Risk assessment and management are not integrated into any contemporary Grid solution. The idea of estimating the risk of violating an SLA is an obvious consequence of current research topics. The importance of SLAs to Grid commercialisation has lead to a drive to standardise SLA negotiation in the Grid. Within the Global Grid Forum (GGF), this work has been lead by the Grid Resource Allocation Agreement Protocol (GRAAP) working group, resulting in two draft standards for SLA description (WS-Agreement [9]) and SLA negotiation (WS-Agreement Negotiation). Other definitions exist which focus on economics [5] and job submission descriptions [6]. SLAs have been used in the process of negotiation and acquisition of resources and services using the Service Negotiation and Acquisition Protocol (SNAP) [13].

Grid resource scheduling and brokering have received increased attention over the last few years [7]. An overview of some current Grid scheduling efforts is found in [25].

A SNAP based resource broker [19, 13] addresses the problem of reserving resources before job submission, for applications requiring resources on demand. It achieves this by providing a transparent means of meeting end-users' application requirements through the use of SLAs. The Grid Resource Broker [8] provides a portal interface through which users can specify and view resource information and tasks. It does not provide automated resource discovery and decision-making is left to the user. In addition, resources are not secured before task submission.

Condor-G [18] is a system which allows users to take advantage of dedicated and non-dedicated CPU cycles. Resources are allocated using parameters that enhance system utilisation and throughput. Condor/G can manage resources for its jobs but provides no support for co-allocation or site autonomy. In Nimrod/G [10] resource allocation adheres to the user requirements and in preference to system utilisation the utility of computation is enhanced. AppLeS (Application-Level Scheduling) adaptively schedules and deploys applications in dynamic, heterogeneous, multi-user environments [14]. The EZ-Grid project [11] uses Globus services [22] to make Grid resource usage easier and more transparent for the user. This is achieved by developing easy-to-use interfaces coupled with brokerage systems to assist the resource selection and job execution process. The Portable Batch System (PBS) is a batch queueing and workload management system useable for example with the MAUI scheduler. Grid sites can then define their own scheduling policies for running jobs in both time and space [27]. The Load Sharing Facility (LSF), which is currently the most widely used commercial job management system, comprises load sharing and batch queueing software that manages, monitors, and analyses the resources and workloads on a network of heterogeneous computers [4]. None of these systems use risk information in the negotiation, scheduling, or management of Grid resources. The gLite workload man-

Proceedings of the Second IEEE International
Conference on e-Science and Grid Computing (e-Science'06)
0-7695-2734-5/06 $20.00 © 2006 IEEE

agement system is also worthy of mention since it uses historical statistics to generate a reputability ranking for resource providers [1], which can be considered analogous to the confidence service in AssessGrid.

In the Grid fabric, reservation and allocation of corresponding resources are important to achieve the desired QoS. This implies that candidate resources are selected by considering reliability, availability, cost, and performance. Foster et al. [16] present a prototype for the Globus Architecture for Reservation and Allocation (GARA). GARA supports dynamic discovery and reservation of heterogenous (in type and administration) resources. However, it lacks risk management in the resource selection. Accordingly, reliability, availability, cost, and performance are considered but without an estimate of the risk of failure of the SLA. Early work in the AssessGrid project [21], examines a provider's responsibility in providing risk-aware resource reservations. Further, a brief overview of the benefits of risk-aware migrations has been published [31]. The paper presented here aims to give a more general overview of the whole project, considering all actors: the end-user, broker and provider.

If the risk of an SLA violation increases during job-runtime, e.g. because of system outages, the resource provider can initiate precautionary fault-tolerance actions. Precautionary action enhances the fault-tolerance functionalities developed in the EC-funded HPC4U project [2]. Distributed checkpointing and migration for single-node jobs are already realised in order to ensure the fulfilment of SLAs, even in cases of system failures [20]. Combining the HPC4U results with risk management functionalities will enable taking actions *before* an outage. Accordingly, the danger of an SLA violation is significantly reduced.

3 Applying Risk Management in the Grid

Risk is defined as "Hazard, danger, exposure to mischance or peril" [3]. In the present context, risk corresponds to hazardous events. As indicated in the introduction, the importance of evaluating and managing such events is recognised in a wide range of disciplines.

In a Grid environment resource providers aim to assess and manage the risk associated with offering an SLA to an end-user (or broker acting on behalf of a user). Hence the hazardous events in this perspective are any events which potentially adversely affect the provider's ability to ensure that the SLA is fulfilled. The risk associated with such events can be characterised using two key parameters: the probability of occurrence and the impact of occurrence. Consider the example of a node outage affecting a compute resource on which a user's job is running. In order to evaluate the risk of failure, the provider must take into account the possible causes of such an event and their likelihood of occurring. For example, a node outage could be caused by a power cut, a system crash, hardware failure, etc. Each of these events must be accounted in order to enable a calculation of the probability of occurrence. In an ideal world sufficient data and information would be available to enable a probabilistic statistical evaluation of risks. Hence, if a database containing detailed statistics relating to previous node outages and their causes is available, then this could be used to support a probabilistic evaluation of this particular risk. Various approaches to probabilistic risk assessment and discussions of different methods can be found in the literature [24]. However, it is possible that detailed statistical information is not available and cannot be extrapolated for all events that have been identified. For example, the risk of an SLA violation may be higher if special operators are not available because a resource outage could be remedied faster by specialists than by unqualified operators. If information about the availability of specialists cannot be extrapolated from available data, it is necessary to use soft decision analysis techniques, relying on possibility theory and fuzzy logic [12] for the risk assessment.

Different hazardous events have different impacts on an SLA fulfilment. Accordingly, having assessed the probability of hazardous events it is also necessary to evaluate their impact. Consider the example of a node outage discussed above. If the user has a tight deadline, it may be that such an occurrence would result in an SLA violation with no possibility of recovering from the situation. However, if redundancy is built into the SLA (see discussion below) or the job is regularly check-pointed and alternative free resources are available then the impact may not be so severe.

Risk management strategies can also be applied to address unacceptable levels of risk. For example, an SLA request could be rejected (risk avoidance) or an application could be executed redundantly on different compute resources (risk reduction).

It is not only the provider that is concerned with risk assessment. The end-user would like to know the risk associated with agreeing an SLA on their application and may not wish to rely solely on a risk assessment obtained from the provider(s), since this would rely on the accuracy of the resource provider(s) assessment. Consequently, an end-user would prefer to have access to an independent and *objective* evaluation. The user will make use of a resource broker that negotiates with resource providers on his behalf. In this case, it is desirable for the broker to have access to functionality which can estimate the reliability of any resource provider's risk assessment. The end-user may have an application that consists of a number of tasks. In this case, a risk assessment of the entire workflow is needed. Ideally the resource broker would also have access to such assessments.

This research is addressing the needs outlined above by providing mechanisms for risk identification, risk assessment,

Proceedings of the Second IEEE International
Conference on e-Science and Grid Computing (e-Science'06)
0-7695-2734-5/06 $20.00 © 2006 IEEE

risk management, and risk monitoring on all Grid layers. The resource provider will be supported by risk assessment tools and mechanisms to identify causes of previous SLA violations in order to improve its system and reduce future risks. The middleware layer is to be enhanced to enable the broker to evaluate the reliability of risk assessments from resource providers and to determine the risk associated with workflow mappings, onto various resource providers. These mechanisms assist end-users in assessing the risk of agreeing an SLA for an application, thereby enabling them to make informed decisions when submitting their jobs.

4 System Architecture

The system architecture (Figure 1) has to take three actors into account: an end-user, a broker, and a provider. The first layer in the architecture represents the end-user. This layer depicts an interface between the end-user and the provider, broker, and confidence service.

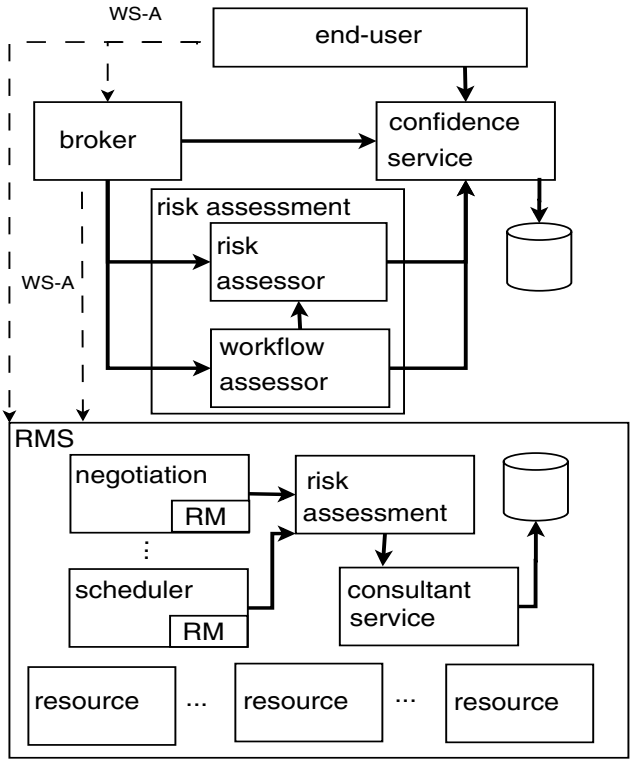

Figure 1. System Architecture of AssessGrid

The second layer illustrates the broker in the Grid middleware. The broker supports the end-user by supplying SLA offers from the provider based on the job requirements. The broker negotiates with the provider on behalf of the end-user. The broker's role is supported by a risk assessment module which contains a risk assessor and a workflow assessor with the functionality to compute risk assessments for workflow orchestrations. In addition the broker has access to a confidence service which provides statistics to enable the reliability of providers' risk assessments to be determined. The confidence service also has access to a database which contains information describing previous SLA offerings and fulfilments. A detailed description is given in section 4.2. The third layer is the Grid fabric in which the provider manages its resources. The consultant service supports the provider's risk assessment methods with statistical information. Figure 1 illustrates communication between the components within the proposed architecture. Dashed lines represent communication which uses an adaptation of the existing WS-Agreement protocol [9]. Solid lines represent communication between components that will require new protocols which will be developed during the project.
The next subsections provide an insight into the three layers of the architecture.

4.1 End-User

AssessGrid supports the end-user with a graphical user interface (GUI). Within this GUI the user can describe prerequisites for their jobs, such as hardware architecture, operating system, amount of memory, and libraries, etc. The user interface modifies the broker's/provider's SLA template based on this input and sends it to the Grid broker or resource provider in order to gather SLA offers. This communication will be realised using the WS-Agreement protocol [9].
The offers are returned in the form of completed SLAs and can be visualised by the end-user. The user can select the different SLAs and read their content, e.g. fee, risk of failure, penalty in case of failure, options to reduce the computed risk, and the operation's costs, etc. WS-Agreement [9] will be used to represent SLA conditions.
To evaluate the provider's reliability the end-user can access a confidence service. It compares the providers' risk assessment with statistics about fulfilled SLAs in the past.
When the customer accepts an SLA, the user interface transfers the job to the provider and supports the user with monitoring information during the job execution. There are a number of established support tools which provide job monitoring within a Grid infrastructure [26].

4.2 Broker

The resource broker is responsible for enabling risk-aware negotiation for resource usage and application execution. This includes both applications which consist of single tasks and those which involve workflow orchestration. The broker is deployed as a Grid service and the WS-Agreement

Proceedings of the Second IEEE International
Conference on e-Science and Grid Computing (e-Science'06)
0-7695-2734-5/06 $20.00 © 2006 IEEE

protocol (WS-Agreement) [9] is used to establish SLAs. The broker has the functionality to:

- negotiate with Grid resource providers on behalf of end-users
- combine, from different resource providers offers into a single package for end-users and map workflows onto a number of resource providers
- rank offers and present them to end-users, based on risk assessments obtained through interaction with a risk assessment module
- agree SLAs with both end-users and resource providers
- submit jobs to resource providers

As indicated above, the broker has access to a risk assessment module which contains functionality to support two key tasks. Firstly, this module contains a risk assessor with risk assessment methods, which are supported by the confidence service in order to adjust any risk assessment received from a resource provider, on behalf of the broker. The confidence service builds statistics about the ratio of offered and fulfilled SLAs to estimate the provider's reliability. The risk assessor uses this historical information as well as static and dynamic data to provide an accurate evaluation of the provider's risk assessment. If a provider typically underestimates the risk, then the confidence service's statistics will mark this provider as unreliable and the risk assessor will return a higher risk value than the provider. If conversely the provider typically overestimates the risk it will return a lower value. The confidence service also provides an interface to enable end-users to access its statistics about the ratio of offered and fulfilled SLAs directly. Secondly, the module enables a risk assessment of a workflow mapping to be obtained, based on the providers' risk assessments, adjusted if necessary by the confidence service. Specifically, the broker receives from each resource provider an offer to perform a single task containing a risk of failure . It is then necessary to combine these into an overall risk for a particular workflow, using a workflow assessor, which makes use of the risk assessment methods provided by the risk assessor.
Both the workflow assessor and the confidence service functionality is supported by risk assessment methods (provided by the risk assessor). These have access to a database containing data relating to all previous SLAs (both accepted and rejected) submitted to the broker. The database contains information such as number of SLA requests, accepted, rejected, failed and (where available) reasons why an SLA has failed. This database is regularly updated by the broker. The risk assessment methods may also access external data such as the Globus Toolkit 4 information services [23] and the Network Weather Service [26].

4.3 Provider

AssessGrid's functionalities for risk assessment and risk management will enhance a state-of-the-art Resource Management System (RMS). In addition to a module for risk assessment, a consultant service (CS) will be developed as an extension to the Grid fabric. The risk management functionalities will be integrated into the RMS modules (scheduler, negotiation module,. . .).
A consultant service has access to all monitoring information obtained by the provider. This data includes static, dynamic, and historic information about the provider's resources and submitted jobs. Examples of such information are current workload, system outages, temporary performance shortages, monitored network traffic, experts' availability, or general information regarding number of incoming jobs. Mechanisms for accurate storage and access will be developed during the project. With the use of tools to compute risk and data mining methods, the provider can assess:

- the risk or probability of fulfilling a given job described by a specific SLA. The RMS at the provider site can use this knowledge to optimise its negotiation and scheduling strategy
- the risk of failure for the overall infrastructure and for individual components (network, compute nodes/resources, experts' support, backup, etc.)
- dynamic changes in risk for an SLA violation during job execution. Monitoring can detect significant changes and therefore the RMS can thereafter decide to use fault-tolerance mechanisms, e.g. increase the checkpoint frequency for the running jobs, block free resources as potential spare resources, or initiate a precautionary replication of data, which enables a more efficient job migration at a later point of time

Based on this information the provider can improve capacity planning, administration, and management of its Grid site. This leads to higher, cost-effective productivity of virtualised resources. Thus performance and quality of a Grid provider depends directly on the underlying consultant service and risk assessment. Enriching the Grid fabric with risk management methods will in this way improve the provider's reliability.

5 Showcase Scenarios

In this section three showcase scenarios are described, representing typical use case scenarios in an AssessGrid en-

Proceedings of the Second IEEE International
Conference on e-Science and Grid Computing (e-Science'06)
0-7695-2734-5/06 $20.00 © 2006 IEEE

abled infrastructure. The scenarios illustrate the functionality that must be supported within its infrastructure. Each of these scenarios are discussed from the differing perspectives of broker and provider.
In the first scenario an end-user wants to submit a single job to a Grid service provider maybe using a Grid broker. In the second scenario, the user plans to submit a workflow orchestration to a Grid environment. For this, the end-user can contact a Grid broker as a mediator for finding suitable resources for his workflow. The third use-case is similar to the second, however here the broker is responsible for the successful execution of the workflow. In contrast to the second case, the broker agrees on an SLA with the end-user, guaranteeing the overall success of the workflow. This means, if the workflow fails due to a failed job, the broker has to pay the penalty fee to the end-user since he is the main contractor. For each sub-job the broker negotiates an SLA with every provider.

5.1 Single-Job Scenario

If the end-user intends to submit a single job to an AssessGrid infrastructure, he can decide to contact a resource provider directly or to ask a broker providing him a ranked list of offers from various providers.
In the first case the contract will be defined through an SLA between the provider and end-user. The end-user can also query a confidence service directly to obtain a reliability measure for the offering of a resource provider.
In the second case the AssessGrid broker will provide a ranked list of offers from several providers to the end-user. The broker only includes offers which fit the requirements of the job, e.g. timing constraints, special hardware such as Pentium CPUs, software like POSIX conform OS or special libraries. In cooperation with the confidence service and risk assessor, the broker will modify the offered risk of the provider. The risk is reduced (increased) if the provider typically overestimates (underestimates) the risk of failure. The end-user will choose a provider based on price, trustworthiness, deadlines or confidence. The contract in this case will be defined through an SLA between the provider and end-user.

5.1.1 Broker View

The broker's role in this scenario is to negotiate on behalf of the end-user, resulting in an SLA between end-user and a single resource provider. The broker passes the user requirements to resource providers and receives offers in return. Each of these offers has an associated provider's assessed risk for failure. The broker then uses the confidence service to determine the reliability of these risk assessments and the risk assessor to determine whether and how providers' assessments are changed. A ranking of the offers is then returned to the end-user to make a selection.

5.1.2 Provider View

From the provider's perspective it makes no difference if the SLA has been negotiated between an end-user or a broker. For the RMS there is a job bound to an SLA. The terms within the SLA have to be fulfilled, regardless of the type of contractor.
During negotiation the RMS has to take the risk of accepting this new SLA request into account. This demand will result in the development of a new scheduler which is used in this negotiation process. When a new SLA request is received by the negotiation module of the RMS, it contacts the scheduler to check for the possibility of realising the new job in the current system condition. If the scheduler is able to realise the new job (for this decision, the scheduler has to contact the CS to determine the impact of the new job on the overall risk situation), the new job can be accepted. The scheduler returns a risk estimation for this new job. This risk estimation is used by the negotiation manager to negotiate on price and fee. The job will only be accepted if price/fee and risk are acceptable to the provider.
During runtime the RMS is responsible for the job execution. It has to ensure that the job will finish according to all terms fixed in the SLA. If necessary, the system has to initiate actions when problems appear. This can either be done reactively (i.e. after the problem has occurred, e.g. a node failure) or pro-actively (i.e. before a problem has occurred). For rating the system condition, the scheduler and the monitoring components will use the services of the CS. In case of node outages, the RMS has to ensure the adherence with SLAs by using its fault tolerance mechanisms. The RMS has the option of migrating a job to free spare resources within its own machine, to migrate the job to another resource within the same administrative domain, or to find suitable resources on another Grid site able to resume the job. The decision as to which action to initiate is taken by fault tolerance mechanisms located in the scheduler of the RMS. With the available mechanisms of AssessGrid the scheduler will contact the CS to take the risk associated with each alternative into account. In the case that a very important job has to be finished, the scheduler may decide to migrate to a remote resource instead of continuing on local spare resources. If these remote resources have significantly lower risk than the internal ones, even additional costs are acceptable.

5.2 Workflow Scenario with Broker acting as Mediator

In this case the end-user intends to submit a workflow consisting of a number of jobs to AssessGrid. The broker will

Proceedings of the Second IEEE International
Conference on e-Science and Grid Computing (e-Science'06)
0-7695-2734-5/06 $20.00 © 2006 IEEE

search for offers for each sub-job within the workflow from various providers. Each provider which fits the requirements of the sub-job and is provided with the computed risk from its own consultant service.
The end-user has to negotiate an SLA for every sub-job with every provider involved. In case of failure of one sub-job the end-user needs to intervene and find a new solution.

5.2.1 Broker View

In this scenario, the broker needs to find an appropriate workflow mapping onto resource providers, taking account of the overall risk associated with the workflow. The broker receives offers from resource providers for each individual sub-job, with their associated risk assessment. As in the previous scenario, the broker uses the confidence service to evaluate the reliability of these risk assessments and the risk assessor to make appropriate changes in the offered risk. The additional step of assessing the overall risk for different possible workflow mappings is required. This is supported by the workflow assessor within the risk management module, as discussed in Section 4.1. The broker may suggest workflow mappings involving some redundancy, where high-risk sub-jobs are carried out by more than one resource provider to minimise risk. The possible workflow mappings are then ranked and returned to the end-user to make a selection. Once the end-user selects a particular workflow, the SLAs are made between resource providers and the end-user for each sub-job, with the broker acting as a mediator. Hence the broker's responsibility for the application ends once the SLAs are agreed, pre-runtime.

5.2.2 Provider View

From the provider's view there is no difference between the first and this scenario since the RMS always negotiates solely on single jobs and does not know about workflows.

5.3 Workflow Scenario with Runtime-responsible Broker

In this case the end-user submits a workflow to the resource broker which will provide the end-user with only one SLA for the whole workflow. For every sub-job the broker will search for appropriate providers and computes the associated risk, then it choose the best providers based on given terms (fastest, cheapest, most trustworthy . . .).

5.3.1 Broker View

The key distinction between this and the previous scenario is that the broker is no longer acting solely as a mediator but is instead offering a higher level of service to the end-user. Once the broker has obtained a risk assessment associated with each possible workflow, it must then make a workflow selection on behalf of the end-user. The broker then agrees a single SLA with the end-user and an SLA with each resource provider that is performing a sub-job. An important consequence of this approach is that the broker is responsible for ensuring the user's requirements are met during run-time. Otherwise the broker will need to pay a penalty fee to the end-user. Hence if a sub-job fails or appears likely to fail, it is in the broker's interest to take steps to reduce the risk of violating its SLA with the end-user. For example if a resource provider informs the broker that it is experiencing node outages and is unable to complete its sub-job the broker needs to re-negotiate with other resource providers in order to agree a new SLA for the sub-job to ensure the workflow can be completed. The broker may also take such a step if the risk of failure for a particular sub-job is high enough to warrant this action, taking into account the cost of submitting the sub-job to a new resource provider and the penalty fee that would need to be paid to the end-user in the event that the SLA between broker and end-user is not met.

5.3.2 Provider View

From the provider's view there is no difference between the first and this scenario, since the RMS always negotiates solely on single jobs and does not know about workflows.

6 Conclusion and Future Directions

The AssessGrid project aims to integrate risk assessment and management into Grid infrastructures. The combination of state-of-the-art Grid technologies and methods of risk assessment will enhance the service of brokers and providers. The resource provider will be able to estimate the risk of agreeing an SLA. It can take into account the business risk and offer adequate price and penalty terms. Furthermore, during job execution the provider can use risk management techniques to fulfil SLAs. Fewer violated SLAs will lead to improved performance. Brokers will be able to estimate the reliability and trustiness of resource providers. In order to handle workflow jobs in a risk-aware manner, they will also be enhanced by risk assessment functionality. End-users will place more trust in Grid technologies if they know the risks involved before job submission. Additionally, they will be supported by the confidence service which measures the reliability and trustworthiness of a providers offer. This will lead to a higher confidence of Grid end-users and a cost-effective productivity of provider's resources. As a consequence we envision an increased use by commercial users and enhancement of Grid economic models.
The AssessGrid project is at an early stage. A more detailed system architecture definition is currently being developed,

Proceedings of the Second IEEE International
Conference on e-Science and Grid Computing (e-Science'06)
0-7695-2734-5/06 $20.00 © 2006 IEEE

based on the recently completed requirements analysis. Furthermore the specification and definition of risk assessment and management functionality is under investigation.

References

[1] gLite: Lightweight Middleware for Grid Computing. http://glite.web.cern.ch/glite/.

[2] Highly Predicatable CLuster for Internet Grids (HPC4). EC-funded project IST-511531.

[3] Oxford English Dictionary. http://dictionary.oed.com.

[4] Platform LSF. http://www.platform.com/products/LSF/.

[5] Grid Economic Services Architecture Working Group, Global Grid Forum, 2003. https://forge.gridforum.org/projects/gesa-wg.

[6] Job Submission Description Language Working Group, Global Grid Forum, 2005. https://forge.gridforum.org/projects/jsdl-wg.

[7] Scheduling and Resource Management Working Group, Global Grid Forum, 2005. https://forge.gridforum.org/projects/srm.

[8] G. Aloisio, M. Cafaro, E. Blasi, and I. Epicoco. The Grid Resource Broker, a Ubiquitous Grid Computing Framework. *Scientific Programming Journal*, 10(2):113–119, 2002.

[9] A. Andrieux, K. Czajkowski, A. Dan, K. Keahey, H. Ludwig, J. Rofrano, S. Tuecke, and M. Xu. Web Services Agreement Specification (WS-Agreement). In *Global Grid Forum GRAAP Working Group*, Sep 2006.

[10] R. Buyya, D. Abramson, and J. Giddy. Nimrod/G: An Architecture for a Resource Management and Scheduling System in a Global Computational Grid. In *Proceedings of the 4th International Conference on High Performance Computing in Asia-Pacific Region (HPC Asia 2000)*, Beijing, China, May 2000. IEEE Computer Society Press.

[11] B. Chapman, B. Sundaram, and K. Thyagaraja. EZ-Grid: Integrated Resource Brokerage Services for Computational Grids, 2001. http://www.cs.uh.edu/ ezgrid/.

[12] H. Cho, H. Choi, and Y. Kim. A Risk Assessment Methodology for Incorporating Uncertainties using Fuzzy Concepts. *Reliability Engineering and System Safety*, 78:173–183, 2002.

[13] K. Czajkowski, I. Foster, C. Kesselman, V. Sander, and S. Tuecke. SNAP: A Protocol for Negotiating Service Level Agreements and Coordinating Resource Management in Distributed Systems. In *Proceedings of the 8th Workshop on Job Scheduling Strategies for Parallel Processing*, 2002.

[14] F. Berman et al. Adaptive Computing on the Grid using AppLeS. *IEEE Transactions on Parallel and Distributed Systems*, 14(4):369–382, 2003.

[15] F. Kloman. Risk Management Standards. March 1995.

[16] I. Foster, C. Kesselman, C. Lee, R. Lindell, K. Nahrstedt, and A. Roy. A Distributed Resource Management Architecture that Supports Advance Reservations and Co-Allocation. In *Proceedings of the International Workshop on Quality of Service*, 1999.

[17] I. Foster, C. Kesselman, and S. Tuecke. The Anatomy of the Grid: Enabling Scalable Virtual Organizations. *International Journal of High Performance Computing Applications*, 15:200–222, 2001.

[18] J. Frey, T. Tannenbaum, I. Foster, M. Livny, and S. Tuecke. Condor-G: A Computation Management Agent for Multi-Institutional Grids. In *Proceedings of the 10th IEEE Symposium on High Performance Distributed Computing (HPDC10)*, San Francisco, CA., Aug 2001.

[19] M. Haji, I. Gourlay, K. Djemame, and P. Dew. A SNAP-based Community Resource Broker using a Three-Phase Commit Protocol: a Performance Study. *Computer Journal*, 48(3):333–346, 2005.

[20] M. Hovestadt. Fault Tolerance Mechanisms for SLA-aware Resource Management. In *Workshop on Reliability and Autonomic Management of Parallel and Distributed Systems (RAMPDS-05), at 11th IEEE International Conference on Parallel and Distributed Systems (ICPADS2005)*, 2005.

[21] M. Hovestadt, O. Kao, and K. Voß. The First Step of Introducing Risk Management for Prepossessing SLAs. In *IEEE International Conference on Services Computing (SCC), Chicago, USA*, 2006.

[22] I. Foster. Globus Toolkit Version 4: Software for Service-Oriented Systems. In *IFIP International Conference on Network and Parallel Computing*, pages 2–13. Springer, 2005. Lecture Notes in Computer Science 3779.

[23] J.M. Schopf et al. Monitoring and Discovery in a Web Services Framework: Functionality and Performance of the Globus Toolkit's MDS4. Technical Report ANL/MCS-P1248-0405, Argonne National Laboratory, Apr 2005.

[24] G. Koller. *Risk Assessment and Decision Making in Business and Industry*. CRC Press, 1999.

[25] J. Nabrzyski, J. Schopf, and J. Weglarz. *Grid Resource Management: State of the Art and Future Trends*. Kluwer Academic Publishers, 2004.

[26] R. Wolski. Experiences with Predicting Resource Performance On-line in Computational Grid Settings. *ACM SIGMETRICS Performance Evaluation Review*, 30(4):41–49, Mar 2003.

[27] R.L. Henderson. Job Scheduling under the Portable Batch System. In D. Feitelson and L. Rudolph, editor, *Proceedings of the 1st International Workshop on Job Scheduling Strategies for Parallel Processing*, pages 178–186. Springer Verlag, 1995. Lecture Notes in Computer Science 949.

[28] D. D. Roure, N. Jennings, and N. Shadbolt. The Semantic Grid: Past, Present, and Future. *Proceedings of the IEEE*, 93:669–681, 2005.

[29] A. Sahai, S. Graupner, V. Machiraju, and A. van Moorsel. Specifying and Monitoring Guarantees in Commercial Grids through SLA. In *Proceedings of the 3rd IEEE/ACM International Symposium on Cluster Computing and the Grid (CCGRID'03)*, Tokyo, Japan., May 2003.

[30] D. Sklansky. *The Theory of Poker*. Gamblers Book Club, 1999.

[31] K. Voß. Risk-aware Migrations for Prepossessing SLAs. In *International Conference on Networking and Services (ICNS), Silicon Valley, USA*, 2006.

Proceedings of the Second IEEE International
Conference on e-Science and Grid Computing (e-Science'06)
0-7695-2734-5/06 $20.00 © 2006 IEEE