Technology Management, Economics and Policy Papers

SEOUL NATIONAL UNIVERSITY
College of Engineering
Technology Management, Economics and Policy Program

Grid Business Models for Brokers Executing SLA-Based Workflows

Dang Minh Quan and Jörn Altmann

TEMEP Discussion Paper No. 2010:46

서 울 대 학 교
공 과 대 학, 기술경영경제정책 대학원과정
151-742 서울시 관악구 신림동 산 56-1

Technology Management, Economics and Policy Program
College of Engineering, Seoul National University
San 56-1, Shillim-Dong, Kwanak-Gu, Seoul 151-742 ,KOREA
Phone: ++82-2-880-9140, Fax: ++82-2-880-8389

Technology Management, Economics and Policy Program (TEMEP) is a graduate program at the Seoul National University. It includes both M.Sc. and Ph.D. programs which are integrated into the graduate program at the College of Engineering.

TEMEP Discussion Papers is intended to serve as an outlet for publishing research about theoretical, methodological and applied aspects of industrial economics, especially those related to the institute's areas of specialization, namely management of technology, information and telecommunication, health and energy industries, as well as development economics infrastructure. In particular submission of studies analyzing current technology and industry related issues and discussion of their implications and possible alternative policies are welcome. The objective is to gain insights into important policy issues and acquiring a balanced viewpoint of policymaking, technology management and economics which enable us to identify the problems in the industries accurately and to come up with optimal and effective guidelines. Another important aim with the series is to facilitate communication with external research institutes, individual researchers and policy makers.

Research disseminated by TEMEP may include views on information and telecommunication, technology management, energy, health and development economics policy, but the institute itself takes no institutional policy positions. When appropriate the policy views of the institute are disseminated in separate policy briefs. Thus, any opinions expressed in TEMEP Discussion Papers are those of the author(s) and not necessarily the institutes.

Editor: Almas Heshmati
Professor of Economics

Technology Management, Economics and Policy Program
College of Engineering
#37-306, Seoul National University
San 56-1, Shilim-dong, Kwanak-gu
Seoul 151-742 Korea

Phone: 0082-(0)2-880-5845
Fax: 0082-(0)2-886-8220
E-mail: heshmati@snu.ac.kr

Grid Business Models for Brokers Executing SLA-Based Workflows

Dang Minh Quan and Jörn Altmann

January 2010

Abstract: In the Grid Computing environment, many users need the results of their calculations within a specific period of time. Those users are willing to pay for getting their work completed on time. However, this requirement must be agreed on in a Service Level Agreement (SLA) by both, the user and the Grid provider, before the application is executed. In order to finish the workflow on time, sub-jobs of the workflow must be distributed to Grid resources. To free users from assigning sub-jobs of the workflow, it is necessary to have a SLA workflow broker. We introduce a set of business models that are based on incentive structure for users to consume the HPCC services, and for brokers to perform efficiently. The business model also leads to several economic issues. In particular, there is asymmetric information regarding the execution environment. The asymmetric information may have negative effect on the business relation between the broker and the user. We will analyze this effect. We also propose some possible solutions for the asymmetric information issue. They include the pricing/fining policy of the broker and the negotiation strategy between users and brokers.

Keywords: Service Level Agreements (SLA), Business Model, Workflow Management, Resource Allocation, Optimization, Information Asymmetry, Negotiation.

JEL Classification Numbers: C02, C61, C63, C78, D46, D82, D86, L11, L86, M15, M21

1. Introduction

In the Grid Computing environment, many users need the results of their calculations within a specific period of time. Examples of those users are weather forecaster running weather forecasting workflows, automobile producer running dynamic fluid simulation workflow [7]. Those users are willing to pay for getting their work completed on time. However, this requirement must be agreed on by both, the users and the Grid provider, before the application is executed. This agreement is called the Service Level Agreement (SLA). In general, SLAs are defined as an explicit statement of expectations and obligations in a business relationship between service providers and customers. SLAs specify the a-priori negotiated resource requirements, the quality of service (QoS), and costs. The application of such an SLA represents a legally binding contract. This is a mandatory prerequisite for the Next Generation Grids.

In order to finish the workflow on time, sub-jobs of the workflow must be distributed to Grid resources. Assigning sub-jobs of the workflow to resources requires considering many constraints such as workflow integrity, on-time conditions, and optimality conditions. To free users from those tedious stuffs, it is necessary to have a SLA workflow broker performing the co-operation task of many entities in the Grid. The business relationship of the SLA workflow broker with the users and the Grid service providers will determine the working mechanism of the broker.

In this chapter, we introduce a set of business models that are based on incentive structure for users to consume the HPCC services, and for brokers to perform efficiently. The business model also leads to several economic issues.

A challenge for the contractual relationships between users and brokers is that users have not as much information about the Grid system as a broker. The broker knows about all aspects of all service providers such as resource configurations, pricing scheme, and past performance. This means there is asymmetric information regarding the execution environment. The asymmetric information may have negative effect on the business relation between the broker and the user. We will analyze the effect with several scenarios and show how the effect can be negative. We also propose some possible solutions for the asymmetric information issue. They include the price/fining policy of the broker and the negotiation strategy between users and brokers.

From the business model, we can see that the user and the broker are business entities. With a workflow, the user is willing to pay up to C_h and the broker will not accept anything below C_l, $C_l < C_h$. The user wants to have a price as low as possible while the broker wants to maximize its profit. There is a bilateral negotiation between the user and the broker to find the appropriate price. Thus, the central problem is how to set the appropriate price that both user and broker can accept. We present a mechanism using fuzzy logic to determine that price.

2. Related works

The literature records many efforts supporting QoS for workflow. AgFlow is a middleware platform that enables the quality-driven composition of Web services [13].

QoS-aware Grid Workflow is a project, which aims at extending the basic QoS support to Grid workflow applications [2]. The work in [12] focuses on mapping the sweep task workflow to Grid resources with deadline and budget constraints. However, all of them do not define business model for the system. Recently, there are many Grid projects working on SLA issue [5,11]. Most of them focus on single job and thus, only the direct relation between user and service provider is considered. The business role of the broker in such systems is not fully evaluated. Moreover, the asymmetric information issue and the bargaining are not considered in all above works.

Related to bargaining, Sim et. al. developed a market-driven bargaining strategy for G-negotiation agents [9]. In which, the authors describe the price setting and adjusting along multiple negotiation steps among agents representing the customers and providers. The key difference between our work and the work of Sim et. al. is that our work attempts to determine a suitable price at the first negotiation round while Sim et. al. try to reach the suitable price after many negotiation rounds.

3. Business model

3.1 Grid-based workflow model

Workflows received enormous attention in the databases and information systems research and development community [6]. According to the definition from the Workflow Management Coalition (WfMC) [4], a workflow is "*The automation of a business process, in whole or parts, where documents, information or tasks are passed from one participant to another to be processed, according to a set of procedural rules.*" Although business workflows have great influence on research, another class of workflows emerged in sophisticated scientific problem-solving environments, which is called Grid-based workflows [1,7]. A Grid-based workflow differs slightly from the WfMC definition as it concentrates on intensive computation and data analyzing but not the business process. A Grid-based workflow has following characters [10]:

- A Grid-based workflow usually includes many sub-jobs (i.e. applications), which perform data analysis tasks. However, those sub-jobs are not executed freely but in a strict sequence.
- A sub-job in a Grid-based workflow depends tightly on the output data from previous sub-jobs. With incorrect input data, a sub-job will produce wrong results and damage the result of the whole workflow.
- Sub-jobs in the Grid-based workflow are usually computationally intensive. They can be sequential or parallel programs and require a long runtime.
- Grid-based workflows usually require powerful computing facilities (e.g. super-computers or clusters) to run on.

Like many popular systems handling Grid-based workflows [3,7], our system is of the Directed Acyclic Graph (DAG) form. The user specifies the required resources needed to run each sub-job, the data transfer between sub-jobs, the estimated runtime of each sub-job, and the expected runtime of the whole workflow.

3.2 Grid service model

The computational Grid includes many High Performance Computing Centers (HPCCs). Sub-jobs of the workflow will be executed in HPCCs as it brings many important advantages for this task.

- Only these HPCCs can handle the high computing demand of scientific applications.
- The cluster or super computer in a HPCC is relatively stable and well maintained. This is an important feature to ensure finishing the sub-job within a specific period of time.
- The HPCCs usually connect to the worldwide network by high speed links, whose broad bandwidth makes the data transfer among sub-jobs easier and faster.

The resources of each HPCC are managed by a software called local Resource Management System (RMS). In this chapter, RMS is used to represent the cluster/super computer as well as the Grid service provided by the HPCC. Each RMS has its own unique resource configuration. A resource configuration comprises the number of CPUs, the amount of memory, the storage capacity, the software, the number of experts, and the service price. To ensure that the sub-job can be executed within a dedicated time period, the RMS must support advance resource reservation such as CCS [5].

3.3 Business model

In order to finish the workflow on time, sub-jobs of the workflow must be distributed to Grid resources. Figure 1 depicts a sample scenario of running a workflow on the Grid environment. Letting users work directly with resource providers has two main disadvantages:

- The user has to have sophisticated resource discovery and mapping tools in order to find the appropriate resource providers.
- The user has to manage the workflow, ranging from monitoring the running process to handling error events.

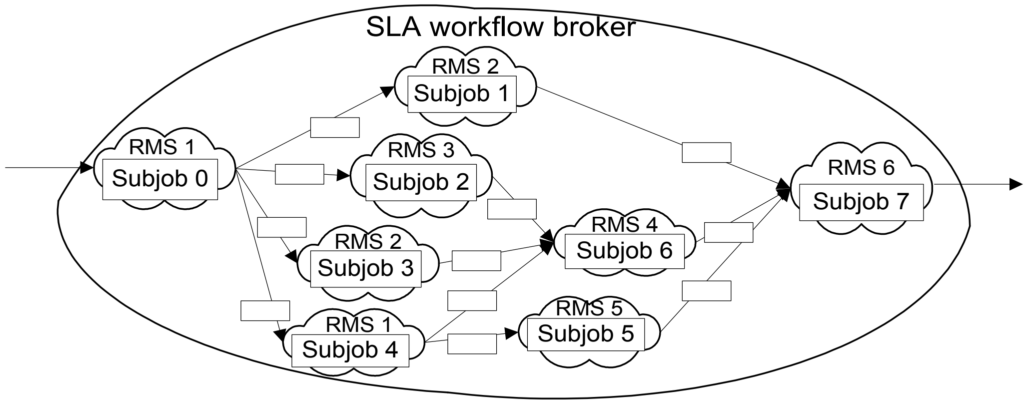

Figure 1. A sample running Grid-based workflow scenario

Thus, it is necessary to have a SLA workflow broker performing the co-operation task of many entities in the Grid. The business relationship between entities in the system running the SLA-based workflow is depicted in Figure 2.

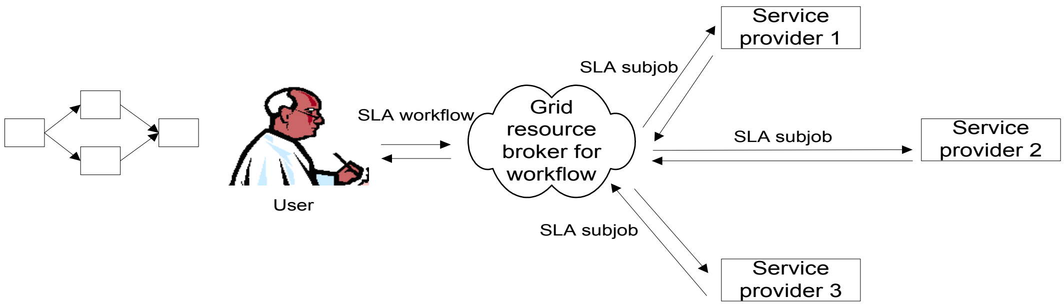

Figure 2. Stakeholders and their business relationship

There are three main types of entities: end-user, SLA workflow broker and service provider:

- The end-user wants to run a workflow within a specific period of time. The user asks the broker to execute the workflow for him and pays the broker for the workflow execution service. The user does not need to know in detail how much he has to pay to each service provider. He only needs to know the total amount. This amount depends on the urgency of the workflow and the budget of the user. If there is a SLA violation, for example the runtime deadline has not been met, the user will ask the broker for compensation. This compensation is clearly defined in the Service Level Objectives (SLOs) of the SLA.

- The SLA workflow broker represents the user as specified in the SLA with the user. It controls the workflow execution. This includes mapping of sub-jobs to resources, signing SLAs with the services providers, monitoring, and error recovery. When the workflow execution has finished, it settles the accounts. It pays the service providers and charges the end-user. The profit of the broker is the difference. The value-add that the broker provides is the handling of all the tasks for the end-user.

- The service providers execute the sub-jobs of the workflow. In our business model, we assume that each service provider fixes the price for its resources at the time of the SLA negotiation. As the resources of a HPCC usually have the same configuration and quality, each service provider has a fixed policy for compensation if its resources fail. For example, such a policy could be that *n%* of the cost will be compensated if the sub-job is delayed one time slot.

Business obligations between broker and service providers includes:

- The broker asks a service provider to execute a sub-job of the workflow. If the provided parameters such as the amount of storage, the amount of memory and the runtime are not correct, the provider has to cancel the sub-job. Otherwise, other sub-jobs would be affected through less-available resources. In this case, the broker has to pay the fee of using resource.

- If the sub-job cannot be finished because of serious computing system failure, the sub-job will be canceled and the provider is fined. This case happens when the whole system is down. For example, the electric power or the network link to the Internet is broken.
- If the runtime of the sub-job exceeds the limit because of computing system failure, the sub-job can continue running and the provider has to pay fining for late time slots. This case happens when there are some small errors inside the computing system. For example, some nodes executing the sub-job are down. To recover from the error, the provider restarts the sub-job from the checkpoint image leading to the small lateness.

Business obligations between user and broker include:

- In order to successfully execute the workflow, the user has to correctly estimate the parameter of workflow's sub-jobs such as the amount of storage, the amount of memory and the runtime. Wrong estimation will lead to the cancellation of sub-jobs and the whole workflow. If this case happens, the user has to pay the following costs: the cost for already executed sub-jobs, the cost for canceling resource reservation of the waiting sub-job, the fine for wrong estimation.
- If the workflow cannot be finished because of computing system failure, the workflow will be canceled and the broker is fined. If this case happens, the broker has to pay the following cost: The cost for already executed sub-jobs, the cost for canceling resource reservation of the waiting sub-job, the fine for not finishing the workflow.
- If the runtime of the workflow exceeds the limit within an acceptable period because of computing system failure, the broker has to pay fining *k$* for each additional time slot.

To distinguish between the case of sub-job delay by wrong estimation or by computing system failure, we have to monitor the system status. If there is a failure of the computing system and the sub-job cannot finish on time, we can deduce that any delay detected is caused by the failure. However, if the runtime of the sub-job exceeds a threshold (which includes the extra time for error recovery), we can say that the user had wrong estimation and the sub-job is canceled.

All activities between end-users, SLA workflow broker and service providers are specified in legally binding contracts (SLAs). We believe that the consideration of business models is important in order to establish incentive structures for users to consume HPCC services at low risk, and for brokers to perform efficiently. That means:

- If the broker does not get any monetary compensation for managing the workflow, there is little incentive for the broker to find a high-quality mapping solution. On contrary, it will encourage the broker to find some unreliable solutions, increasing its income.

- With the proposed business model, the broker takes responsibility for the mapping solution. When a failure happens and the workflow does not finish within the desired period, the user can fine the broker and the broker would not get compensated. This method allows the user to get a guaranteed service.
- The described business model frees the user from the hard work of managing the workflow execution. He only signs an SLA with the SLA workflow broker and then waits for the result.

4. Asymmetric information and negotiation

From the business model, we can see that the broker having more information about the Grid than the user. The broker knows about all aspects of all service providers such as resource configurations, pricing scheme, and past performance. It is difficult for user to know all these information. This situation leads to the asymmetric information issue.

4.1 Asymmetric information issue

In the contractual context between user and broker, asymmetric information could bring negative effects. We look at the following scenarios.

4.1.1 Scenario 1: Asymmetric about Grid state

We assume that the price of provider is fixed at the point of doing mapping. This assumption is suitable as many present resource providers like Sun, Amazon and the Distributed European Infrastructure for Supercomputing Appications (DEISA) use this model. The cost of running a workflow depends mainly on the cost of the mapping solution. The cost of a workflow mapping solution depends on the Grid state. When the Grid is free, there are many free resources and the broker could have big chance to assign many sub-jobs of the workflow to cheap providers. More over, when the resource in each provider is free, there is a great chance to exist a solution that dependent sub-jobs of the workflow are executed on the same RMS. Thus, the cost of data transfer among those sub-jobs is neglected. This leads to low cost mapping solution. In contrast, when the Grid is busy, there are little free resources and the broker may have to assign many sub-jobs of the workflow to expensive providers. The busy state of the Grid also leads to the big chance that sub-jobs of the workflow must be executed in different RMS. In this case, the cost of data transfer could be a significant part of the total workflow running cost. Therefore, the cost of running the workflow could be high.

The user does not know beforehand whether the Grid is free or busy. Therefore, the user's best guess for a mapping solution is that the mapping is done in the average state of the Grid. The user is willing to pay a cost correlated to the average state. Thus, when the Grid state is busy, the high cost of the mapping solution may annoy the user.

4.1.2 Scenario 2: Asymmetric about quality of mapping solution

The deadline of the workflow has different meanings for different people. The importance of a deadline depends on the urgency of the workflow. For example, the result of the weather forecasting workflow is very important especially in storm prediction. The lateness of the weather forecasting workflow in this case may lead to the dead of many people. Thus, the urgency is very high. In contrast, the small lateness of a dynamic fluid workflow in a scientific research project does not have great effect to the progress of the project. In this case, the urgency is very low.

With different urgency level, the user requires different levels of ensuring the deadline. This requirement equals to running the workflow with different risk levels. The risk is defined here as the ability of not finishing the workflow on time. Among many factors affecting the risk level of a workflow mapping solution, the fail probability is the most important. The fail probability includes small-scale failure and large-scale failure. The small-scale system failure is mainly caused by the failure of computing nodes. The large-scale system failures could affects the whole computing system of the provider. Those failures can be large hard ware failure, network connection failure and security holes.

With the workflow with high urgency level, sub-jobs of the workflow should be assigned to RMSs having low failure probability. With the workflow with low urgency level, the demanding of mapping sub-jobs to high reliability is not so high. In general, the price of the RMS having higher reliable level is higher than the price of the RMS having lower reliable level. Thus, the cost of running a high urgency workflow could be higher than the cost of running the lower one.

However, the user does not know beforehand about the failure probability of RMSs and how to evaluate the risk of the mapping solution. If the user requires a high level of ensuring the deadline and is asked to pay a high price, the user may suspect that the broker found an unreliable mapping solution to get higher revenue.

In both described scenarios, if the broker does not have suitable ways to resolve the asymmetric information problem, the broker may loose customers.

4.2 Possible solutions for the asymmetric information issue

4.2.1 Pricing and guarantee as the signal

From the obligation description between user and broker, we referred to the fine. If the broker cannot finish the workflow execution at the due time, he will be fined. There is a question that what the suitable fining rate is. In this part, we present a way to answer the question.

The user and the broker form a contract to execute the Grid-based workflow. The workflow can be finish on time or late. If the workflow finishes on time, its monetary value to the user is $b1$. If the workflow is late, its monetary value to the user is $b2$. Assume that the late probability of the workflow mapping solution is q. We can also assume that the broker is risk-neutral and the user is risk-averse. The broker proposes a contract that the cost to execute the workflow is p and a guarantee g. If the workflow is late, the broker has to pay the user g. The utility of the buyer is $u(b1-p)$ if the workflow is not late and $u(b2-p+g)$ if it is. As the user is risk-averse, $u''<0$. The user will accept the contract when his expected utility is greater or equal to $u(0)=0$. The utility of the broker is presented in Formula 1.

$$B(p,g) = (1 - q)*p + q*(p - g) \quad (1)$$

The utility of the user is presented in Formula 2.

$$U= q*u(b2 - p + g) + (1 - q)*u(b1 - p) \quad (2)$$

The optimal contract must satisfy following conditions:

Max $\{B(p,g)\}$

S.t $U >= u(0)$

Using Lagrange multipliers, we have following results:

$$g = b1 - b2 \quad (3)$$

$$p = b1 - u^{-1}(u(0)) = b1 \quad (4)$$

$$\frac{g}{p} = 1 - \frac{b2}{b1} \quad (5)$$

It is possible to say that with low urgency level, the different in monetary between late and on time result is not so big. For example, if dynamic fluid workflow in a scientific research project is late one or two hours, it has little effect to the progress of the whole project. This means that the $\frac{b2}{b1}$ value is high. Thus, from the Formulation 5, the $\frac{g}{p}$ value would be low. In contrast, with the high urgency level, for example with the weather forecasting workflow, if the result is late 1 hour, many ships may not be able to return the harbour to avoid the storm. Thus, the different in monetary is great. This means that the $\frac{b2}{b1}$ value is low and the $\frac{g}{p}$ value is high. From Formula 5 and the analyzed aspect of price and guarantee, we apply to our case as following.

- The broker provides a menu of contract. Each contract contains urgency level and the appropriate guarantee. The guarantee is computed in percent of the total cost. The guarantee is higher when the urgency level is higher.
- The user chooses a contract from the menu. Based on this requirement, the broker will do mapping solution and negotiate the SLA.

The higher guarantee rate with higher price will persuade the user to believe on the quality of the service provided by the broker.

4.2.2 Signalling during the SLA negotiation

From the business mode, there are three different types of sub SLA negotiation using three different kinds of SLA text. User - Broker negotiation focuses on the definition of the submitted SLA. Broker – Provider negotiation considers the workflow sub-jobs and uses analyses the sub-job SLAs. Provider - Provider negotiation deals with the data transfer between sub-jobs (and also between providers) so the SLA part for data transfer is used.

Although there are three types of SLA negotiation, the negotiation procedure remains the same, only the service attributes differ. In a first step, the client creates a template SLA with some preliminary service attributes and sends those to the server. The server parses the text and checks the client requirements. In case of conflicts, a new SLA version is compiled and sent back. Here, we focus on the negotiation process between user and broker. When receiving an SLA from customer, the broker parses it to get all information about the general SLA, sub-jobs, SLO, data transmission, the dependency among sub-jobs and the structure of the workflow. From the information of sub-jobs and the structure of workflow, the broker does mapping to determine the appropriate provider and the time period to run each sub-job. During the negotiation process, the broker could provide the user following information to avoid the asymmetric information issue.

- The number of feasible solutions in the reference set created by H-Map algorithm [8].
- Mapping information. The mapping information include the start, stop time of the workflow and the RMS for each sub-job. Depending on the state of the Grid, a mapping module can find a feasible solution in the expected time period or not. If not, it will find the earliest solution and ask for the consumer's approval.

This information contains many signals for the user.

- Firstly, the number of feasible solutions in the reference set created by H-Map algorithm can say about the Grid state. H-Map algorithm created a reference set that distributed over the search space. If the number of feasible solution is low, this means the Grid is busy and vice versa.
- The second is the start, stop time of the workflow. If it is not within the preferred period from the user, it means that the Grid is very busy and the user should prepare for a high execution cost.
- The third is the RMS for the sub-jobs of the workflow. By providing this information, the broker signals the user about the cost of the mapping solution. With those information, the user can do query himself to know the price from each provider. Since then, he can evaluate the cost of the mapping solution.

It is noted that the broker should not provide detail information about start, stop time of sub-job as well as data transfer. It is because that this information does not signal the user about the mapping solution. It may help user bypass the broker to work directly with providers.

5. Bargaining game in negotiation

One workflow has many parameters such as number of input, output data, required CPU speed, required number of memory, required number of storage, QoS, execution duration and so on. With each user, one or some of those parameters are modified to fit with the specific requirement. As far as we know, there is rarely more than one user having the identical workflow.

In the early phase of business Grid, the scale of the Grid is small and the number of SLA workflow brokers is also small. Even when the Grid has large scale, some local Grid organizations such as the Distributed European Infrastructure for Supercomputing Appications (DEISA) still exist. When the local Grid organization establishes an SLA workflow broker service, this broker could have many supports about technique and pricing policy from the local Grid community. It is obvious that those users of this Grid should choose this broker as default.

The analyzed situation above describes a market for a workflow with one user and one broker. This is the ideal condition for bilateral bargaining to happen. The user wants to have a price as low as possible while the broker wants to maximize its profit. Thus, the central problem is how to set the appropriate price that both user and broker can accept.

5.1 Bilateral bargaining game

Assume that after many times running the same workflow, the user knows the highest cost C_h of running the workflow. Thus, C_h is the highest price that the user can pay for the workflow execution. The broker does mapping to determine the based price C_l of running the workflow. C_l includes the cost paying for service providers and the cost for the broker service. The broker will not accept to run the workflow if the price is less than C_l. If $C_l > C_h$ the trading will not happen. Assume that $C_l < C_h$, the price scenario is presented in Figure 3. As the broker is a business entity, he also wants to gain profit. Thus, the actual price of the service C will satisfy $C_l < C < C_h$. As presented in the part about asymmetric information, it is noted that the broker signals the user about the based cost C_l and the Grid state.

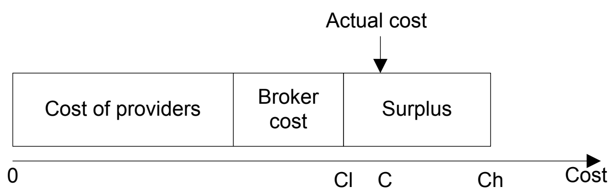

Figure 3. Pricing schema of the game

During the negotiation phase, the broker and the user will do bargaining over the division of the surplus C_h - C_l using alternate offers. The negotiation process cannot take place forever. There is a defined start time of the workflow execution while each negotiation round needs about one time slot for user to recognize the change of the SLA content or for broker to find a mapping solution. With this feature, we say that both broker and user are impatient. The level of user's impatience is different from broker's impatience. This is discussed in more detail in the below section. With this feature, we can say that this is a bargaining game with asymmetric impatience. The participants will accept the offer when they are exactly indifferent about accepting or rejecting it.

We assume that the cost of delaying agreement by one round reduces the gains from trade by s_u% with the user and s_b% with the broker. This cost relates to the omitted opportunity when the negotiation time extends. s_u and s_b are common knowledge of both user and broker because all related information is revealed. The *1-s* percent is defined as the discount factor. We call d_u% (d_u =1-s_u), d_b% (d_b =1- s_b) as the discount factor of the user and the broker respectively.

Assume that the game can be taken place in N rounds. Considering the sub-game beginning with the final round N^{th}, the broker offers a price C_h and the user accepts it. The reason is that if the user does not accept, his utility will be reduced seriously because of the lateness. Thus, at this sub-game, the share surplus of the broker Sb_N=100% and the share surplus of the user Su_N =0%. Now, we move back the sub-game at round $(N-1)^{th}$ beginning with user's offer. As the user knows that delaying the negotiation one more round can waste the broker s_b % of his share surplus, the user offers the broker a share surplus as in Formula 6 and the broker should accept it.

$$Sb_{N-1} = Sb_N * d_b \tag{6}$$

The share surplus of the user is presented in Formula 7.

$$Su_{N-1}= 100(\%) - Sb_{N-1} \tag{7}$$

At the subgame at round $(N-2)^{th}$ beginning with broker's offer, the broker also knows that delaying the negotiation one more round can waste the user s_u % of his share surplus, the broker offers the user a share surplus as in Formula 8 and the user should accept it.

$$Su_{N-2}= Su_{N-1}* d_u \tag{8}$$

The share surplus of the broker is presented in Formula 9.

$$Sb_{N-2} = 100(\%) - Su_{N-1} \tag{9}$$

Using the backward deduction like above we can compute the price at the first round and it is the ideal proposed price for both user and broker. From the analysis, we can see that the discount factor is the key issue in the price determination process. The discount factor depends on many variables: the remaining time period, the urgency of the workflow and the Grid state. We will apply fuzzy logic to determine the discount factor as described in the following Section.

5.2 Fuzzy logic to determine the discount factor

The overall architecture of using fuzzy logic in the price negotiation process is presented in Figure 4.

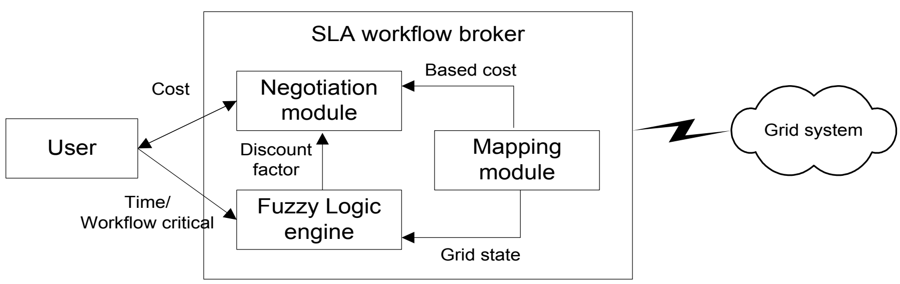

Figure 4. The architecture of using fuzzy logic in the price negotiation process

As can be seen in Figure 4, the fuzzy logic engine receives the remaining time period and the workflow urgency level from the user, the Grid state from the mapping module. Based on those input parameters, the fuzzy logic engine will compute the discount factor d_u and d_b and provides it to the negotiation module. The negotiation module uses d_u, d_b, C_l, C_h to compute the cost of execution the workflow and then proposes to the user.

5.2.1 Membership function of the input parameters

The Grid state

The Grid state has five levels: Very Busy (VB), Busy (B), Medium (M), Free and Huge Free (HF). The level is determined based on the number of feasible solutions that the Grid can propose to the workflow. This number is reflected through the number of feasible solutions in the reference set created by H-Map algorithm [8]. H-Map algorithm created a reference set that distributed over the search space. If the number of feasible solution is low, this means the Grid is busy and vice versa. Call n_f is the number of feasible solutions, n_t is the total number member of the reference set. The level of Grid state is mapped to the rate $\frac{n_f}{n_t}$ as presented in Figure 5.

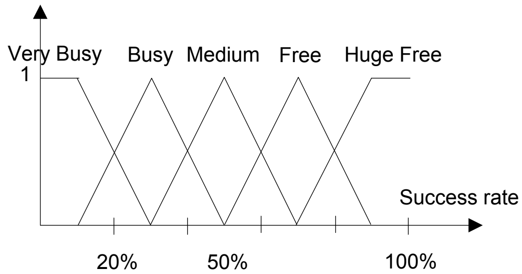

Figure 5. Membership function of the Grid state

The time remaining period

The time remaining period has five levels: Very Short (VS), Short (S) Medium (M), Long (L) and Very Long (VL). The level is determined based on the ability to re-negotiate with the broker. As each negotiation round needs about 1 time slot, 9 time slots is a safe period for the user to change a lot of things. Thus, if the remaining time period is greater than or equal to 9, it is considered as Very High. In contrast, if the remaining time period is 1 time slot, the user has 1 time slot to recognize what should be change and 0 time slot to negotiate with the broker. Thus, the period of 1 time slot is quite dangerous for the user and is considered as Very Low. The level of the time remaining factor is mapped to the real value as presented in Figure 6.

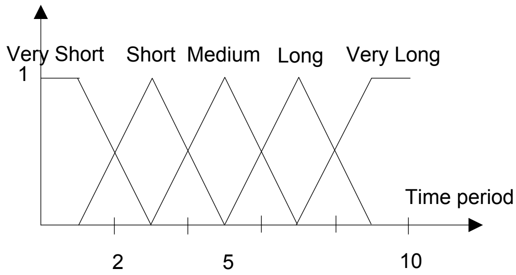

Figure 6. Membership function of the time remaining period

The urgency of the workflow

The urgency of the workflow has five levels: Very Low (VL), Low (L) Medium (M), High (H) and Very High (VH). As the user selects the discrete value for the urgency of the workflow, the membership function is presented in Figure 7.

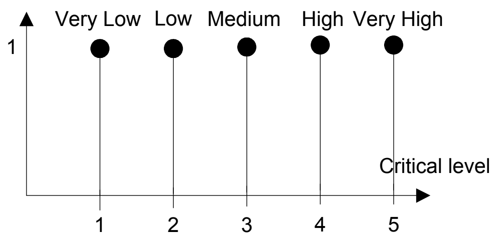

Figure 7. Membership function of the urgency of the workflow

5.2.2 Effect of input parameters to the discount factor

In this section we will analyze the effect of input parameters to the user discount factor and the broker discount factor. These effects are then expressed as fuzzy inference rules. With the broker, the discount factor is determined based on the ability of attracting user to use the service. If the ability of attracting user is big, the discount factor is low and vice versa. With the user, the discount factor is mainly determined based on the utility of ensuring the deadline for the workflow. If the probability of delaying the deadline is big, the discount factor is low and vice versa.

State of the Grid

Effect on the broker: The Grid state affects the policy of attracting user to use the Grid. When the Grid is free, this means that only a small number of customers use the Grid at that time. Thus, the broker wants to attract more users using the Grid. If the negotiation takes a long time, the negotiation could annoy the user and, therefore, could negatively effect the success of attracting users. Thus, the discount factor of the broker is low. The discount factor is high when the Grid is busy and low when the Grid is free. The assumption about effect of the Grid state on the broker discount factor is shown in Table 1.

Table 1. Effect of Grid state on the broker discount factor

Grid state	HF	F	M	B	VB
Discount	VL	L	M	H	VH

Effect on the user: The Grid state affects the ability of finding a feasible mapping solution for the workflow. When the Grid is busy, at time slot *t*, the broker finds out a feasible mapping solution. If the negotiation process extends longer to time slot $t+1$, the probability for the broker to find out a feasible mapping solution is small. Because the number of free resources is small, just another resource demand within $[t,t+1]$ can eliminate greatly the available resource space of the workflow. The situation is reversed when the Grid is free. Thus, the effect of the Grid state on the user discount factor is summarized in Table 2.

Table 2. Effect of Grid state on the user discount factor

Grid state	HF	F	M	B	VB
Discount	VH	H	M	L	VL

The urgency of the workflow

Effect on the broker: The high or low urgency does not affect the broker's ability of attracting user.

Effect on the user: The urgency of the workflow affects the ability of finding a feasible mapping solution for the workflow. When the urgency of the workflow is high, it will require high quality resources. If the negotiation process extends longer, the probability of loosing the feasible found solution is very big. Because the space of suitable Grid resources is small, just another resource demand can greatly eliminate the feasible resource space of the workflow. The situation is reversed when the urgency level is low. Therefore, the discount factor is low when the urgency of the workflow is high and vice versa as summarized in Table 3.

Table 3. Effect of the urgency on the user discount factor

Urgency	VL	L	M	H	VH
Discount	VH	H	M	L	VL

Remaining time period

Effect on the broker: The big or small remaining time period does not affect the broker's ability of attracting user.

Effect on the user: The remaining time period will affect the utility of the user through the probability of ensuring the deadline for the workflow. When the remaining time period is small and the negotiation time extends longer, the user will face the big probability of not starting the workflow on time. The situation is reversed when the period is long. The effect of the remaining time period on the discount factor of the user is summarized in Table 4.

Table 4. Effect of the remaining time period on the user discount factor

Time	VS	S	M	L	VL
Discount	VL	L	M	H	VH

5.2.3 Membership function of the output

The output of the fuzzy logic system is the discount factor for user and broker. With both of them, we use the membership function as presented in Figure 8. To determine the crisp value of the discount factor, we use the popular RSS (Root Mean Square) and centroid methods.

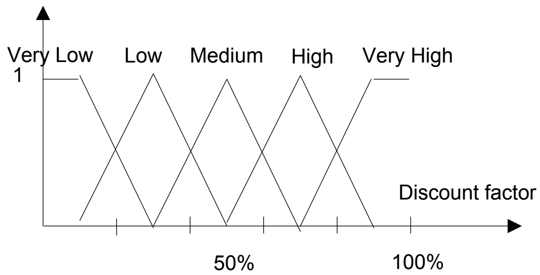

Figure 8. Membership function of the output

5.3 Algorithm to determine the price

The algorithm to calculate the proposed price include following steps:

- Step 1: Gets the urgency level of the workflow from user, computes the remaining time period, determines the highest cost of executing the workflow C_h.
- Step 2: Uses the H-Map algorithm to get the Grid state and the cost for providers.
- Step 3: Computes the broker service cost. This cost could be computed in a simple way for example *k%* of the cost for providers.
- Step 4: Computes the surplus share between broker and provider with procedure presented in Figure 9. At each remaining time period value, we compute the

discount factor and then determine the share. As the Grid state could only be remained or become busier along the time, this means that the broker discount factor becomes higher and the user discount factor becomes lower. Thus, the share for user will be slightly bigger than the equilibrium. However, in the procedure, we use the Grid state determined at the negotiation time for all remaining time period value. It is an incentive for user to accept the proposed price.

```
Share_broker=1
Share_user=0
For I=1 To (time Remaining /2)
{
    Compute The Disc_brk,  Disc_user Using Fuzzy
     With Remaining Time = 2*i -1 And Grid State In Step 2
    Share_broker= Share_broker* Disc_brk
    Share_user=1- Share_broker
    Compute The Disc_brk,  Disc_user Using Fuzzy
     With Remaining Time = 2*i And Grid State In Step 2
    Share_user= Share_broker* Disc_user
    Share_broker=1- Share_user
}
```

Figure 9. Computing the surplus share procedure

- Step 5: Computes the final cost and propose to the user.

5.4 Validation

The goal of this task is to validate and visualize the impact of the input parameters to the discount factor and the surplus share. In the first place, we set all the parameters to medium value. After that, we change each input parameter from very low value to very high value. The discount factor and the surplus share of both user and broker are recorded.

5.4.1 Grid state

From Figure 10, we can see the discount factor of the user does not start from 0 and does not reach 1 because beside the Grid state, it also depends on the urgency of the workflow and the remaining time period. The discount factor and the surplus share of the user increases while the broker's discount factor and surplus share decreases along the increase of the Grid state. In the practical business, when the good is sold slowly, the seller usually reduces price to motivate the trade. Here, we can see the same picture. This means if the Grid is free, the broker reduces its share surplus to attract more users.

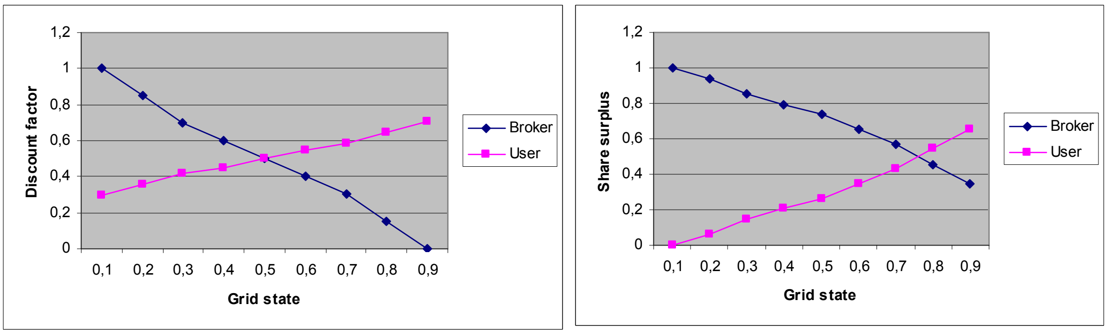

Figure 10. Effect of the Grid state on the discount factor and the surplus share

5.4.2 Time remaining period

From Figure 11, we can see the discount factor of the broker does not depend on the remaining time period while the discount factor of the user increase along the increase of the remaining time period. The surplus share of the user increases while the broker's decreases along increase of the remaining time period. In the practical business, we usually see a technique of negotiation is that the seller tries to postpone the negotiation until the deadline coming very closed. Then, the seller imposes a high price to the customer. We can also see the same situation with the case of the broker and the user. The broker gets a larger part of the surplus when the time remaining period is small.

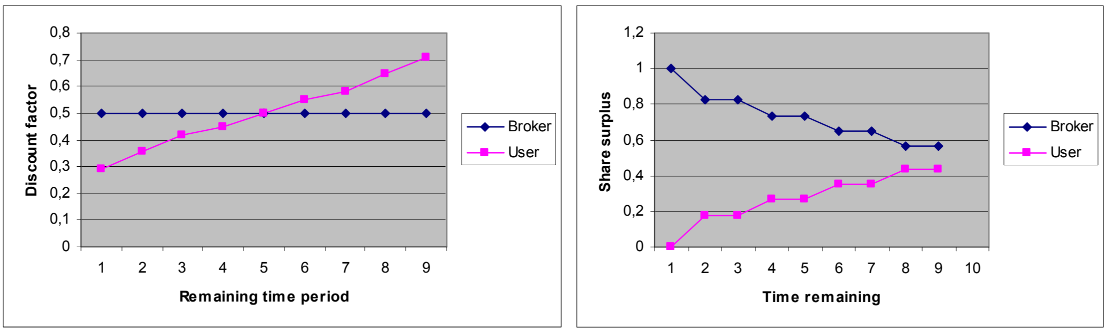

Figure 11. Effect of the remaining time period to the discount factor and the surplus share

5.4.3 Workflow urgency

Figure 12 shows that the discount factor of the broker does not depend on the urgency of the workflow. The discount factor of the user decreases along the increase of the urgency level. Thus, its share surplus also decreases. In the real life, when the seller knows that a customer really like/need a good, he will impose a high price on the customer. The result in Figure 12 reflects well this behavior. This means, if the urgency of the workflow is high, the broker will require a larger part of the share surplus.

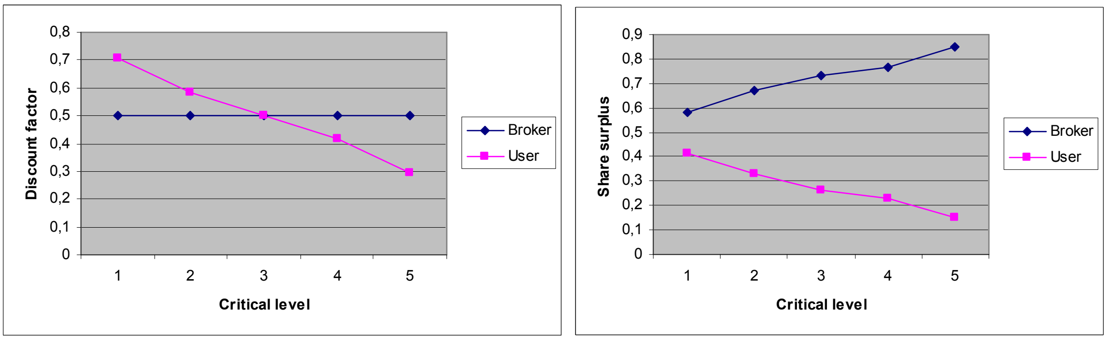

Figure 11. Effect of the urgency level on the discount factor and the surplus share

6. Future research directions

In this chapter, we considered the bilateral relation between the user and the SLA workflow broker. However, when the side of the Grid increases, the trading on the business increases. Many SLA workflow brokers will appear. One problem to the broker is that how to set the appropriate price at the first negotiation round in the competitive context. With the competitive factor, the game between the user and broker becomes more complicated and need to be exploited care fully.

When the Grid market becomes larger, many new business components may appear such as Insurance broker or Risk broker. The functionalities as well as the business roles of those new components are still the questions at the present. The answers for those questions could enrich the business Grid structure and motivate the development of the Grid.

7. Conclusion

This chapter presents a business model and related economic issues for the SLA-based workflows broker. In the business model, user, broker and providers are business entities. In particular, we consider the broker also as a service provider which provides workflow execution service. A set of business obligations among user, broker and provider is defined to ensure a clear and legal relationship. The business model leads to asymmetric information issue between the user and the broker. The user has less information about the cost of executing the workflow and the quality of mapping solution than the broker. The user may suspect that the broker makes benefit from this information. To avoid the negative effects, the broker should have suitable guarantee policy and signal information. Another issue derived from the business model is the bargaining game between user and broker. We proposed a method using fuzzy logic to determine the price that both user and broker can accept at the first negotiation round. The validation results show that the solution reflects correctly the behavior of the practical business.

8. References

1. G. B. Berriman, J. C. Good, and A. C. Laity, Montage: a Grid Enabled Image Mosaic Service for the National Virtual Observatory, in *Proceedings of the Astronomical Data Analysis Software and Systems XIII*, Volume 13, pp. 145-167, 2003.
2. I. Brandic, S. Benkner, G. Engelbrecht, and R. Schmidt, QoS Support for Time-Critical Grid Workflow Applications, in *Proceedings of the 1st IEEE International Conference on e-Science and Grid Computing (e-Science 2005)*, Melbourne, Australia, December 5 - 8, 2005.
3. E. Deelman, J. Blythe, Y. Gil, C. Kesselman, G. Mehta, S. Patil, M. Su, K. Vahi, and M. Livny, Pegasus : Mapping Scientific Workflows onto the Grid, in *Proceedings of the 2nd European Across Grids Conference*, Nicosia, Cyprus, January 28-30, 2004.
4. L. Fischer, *Workflow Handbook 2004,* Future Strategies Inc, Lighthouse Point, FL, USA, 2004.
5. M. Hovestadt, Scheduling in HPC Resource Management Systems: Queuing vs. Planning, in *Proceedings of the 9th Workshop on JSSPP at GGF8*, Washington, USA, June 24, 2003.
6. M. Hsu (ed), *Special Issue on Workflow and Extended Transaction Systems*, IEEE Data Engineering, Volume 16 (2), IEEE-CS, CA, USA, 1993.
7. R. Lovas, G. Dózsa, P. Kacsuk, N. Podhorszki, D. Drótos, Workflow Support for Complex Grid Applications: Integrated and Portal Solutions, in *Proceedings of the 2nd European Across Grids Conference*, Nicosia, Cyprus, January 28-30, 2004.
8. D.M. Quan, Mapping heavy communication workflows onto grid resources within SLA context, in *Proceedings of the 2nd International Conference of High Performance Computing and Communication (HPCC06)*, Munich, Germany, September 12-14, 2006.
9. K. M. Sim, G-Commerce, Market-driven G-Negotiation Agents and Grid Resource Management, *IEEE Transactions on Systems, Man and Cybernetics, Part B*, 36 (6), pp.1381-1394, 2006.
10. M. P. Singh, and M. A. Vouk, Scientific Workflows: Scientific Computing Meets Transactional Workflows, in *Proceedings of the NSF Workshop on Workflow and Process Automation in Information Systems: State of the Art and Future Directions*, Athens, USA, 1996.
11. M. Surridge, S. Taylor, D. De Roure, and E. Zaluska, Experiences with GRIA, in *Proceedings of the 1st IEEE International Conference on e-Science and Grid Computing (e-Science 2005)*, Melbourne, Australia, December 5 - 8, 2005.
12. J. Yu, and R. Buyya, Scheduling Scientific Workflow Applications with Deadline and Budget Constraints using Genetic Algorithms, *Scientific Programming Journal*, 14 (3-4), pp. 217-230, 2006.
13. L. Zeng, B. Benatallah, A. Ngu, M. Dumas, J. Kalagnanam, and H. Chang, QoS-Aware Middleware for Web Services Composition, *IEEE Transactions on Software Engineering,* 30(5), pp. 311-327, 2004.

TEMEP Discussion Papers

2009-31: Nils Karlson, Marcus Box and Almas Heshmati, "Generality, State Neutrality and Unemployment in the OECD"

2009-32: Chang Seob Kim, Yoonmo Koo, Ie-jung Choi, Junhee Hong, and Jongsu Lee, "Consumer Preferences for Automobile Energy Efficiency Grades"

2009-33: Jongsu Lee, Jae Young Choi and Youngsang Cho, "A forecast simulation analysis of the next-generation DVD market based on consumer preference data"

2009-34: Jungwoo Shin, Chang Seob Kim, and Jongsu Lee, "Model for Studying Commodity Bundling with a Focus on Consumer Preference"

2009-35: Yuri Park, Hyunnam Kim, and Jongsu Lee, "Empirical analysis of consumer preferences for mobile phones and mobile contents in the presence of indirect network effects"

2009-36: Jörn Altmann and Juthasit Rohitratana, "Software Resource Management Considering the Interrelation between Explicit Cost, Energy Consumption, and Implicit Cost: A Decision Support Model for IT Managers"

2009-37: Marcel Risch, Ivona Brandic, and Jörn Altmann, "Using SLA Mapping to Increase Market Liquidity"

2009-38: Amit K. Bhandari and Almas Heshmati, "Willingness to Pay for Biodiversity Conservation"

2010-39: Marcel Risch, Jörn Altmann, Li Guo, Alan Fleming, and Costas Courcoubetis, "The GridEcon Platform: A Business Scenario Testbed for Commercial Cloud Services"

2010-40: Junseok Hwang, Jörn Altmann, and Ashraf Bany Mohammed, "Determinants of Participation in Global Volunteer Grids: A Cross-Country Analysis"

2010-41: Ruzana Davoyan, Jörn Altmann, and Wolfgang Effelsberg, "Intercarrier Compensation in Unilateral and Bilateral Arrangements"

2010-42: Ruzana Davoyan, Jörn Altmann, and Wolfgang Effelsberg, "Exploring the Effect of Traffic Differentiation on Interconnection Cost Sharing"

2010-43: Ruzana Davoyan and Jörn Altmann, "Investigating the Role of a Transmission Initiator in Private Peering Arrangements"

2010-44: Ruzana Davoyan, Jörn Altmann, and Wolfgang Effelsberg, "A New Bilateral Arrangement between Interconnected Providers"

2010-45: Marcel Risch and Jörn Altmann, "Capacity Planning in Economic Grid Markets"

2010-46: Dang Minh Quan and Jörn Altmann, "Grid Business Models for Brokers Executing SLA-Based Workflows"