PhD Proposal
Economically enhanced autonomic resource allocation for datacenters on Cloud computing markets

Mario Macías
Advisor: Jordi Guitart

June 4, 2010

Departament d'Arquitectura de Computadors
UNIVERSITAT POLITÈCNICA DE CATALUNYA

Abstract

Cloud Computing markets arise as an efficient way to allocate resources for the execution of tasks and services within a set of geographically dispersed providers from different organisations. Client applications and service providers meet in a market and negotiate for the sales of services by means of the signature of a Service Level Agreement that contains the Quality of Service terms that the Cloud provider has to guarantee by managing properly its resources. Current implementations suffer from a lack of information flow between the negotiating agents that sell the resources, and the resource managers that allocate the resources for fulfilling the agreed SLAs and provide to the client the requested Quality of Service. This thesis proposes to establish a communication channel between the market agents and the resource managers, so agents can perform accurate negotiations by considering the status of the resources in their negotiation models, and providers can perform resource management by considering not only the technical parameters, but also how their decisions can influence in the achievement of the Business Objectives. By using this bidirectional communication, this thesis defines and evaluates several Business-Oriented models and policies for both negotiation and enforcement of SLAs to achieve important Business Objectives of the provider, such as revenue maximisation or classification of clients.

Acknowledgement

This work is supported by the Ministry of Science and Technology of Spain and the European Union (FEDER funds) under contract TIN2007-60625, by the Generalitat de Catalunya under contract 2009-SGR-980.

Contents

Chapter 1

Introduction

In the recent past years, academic and scientific entities as well as some companies owned big mainframes, clusters, or even supercomputers that had to be shared across their users for satisfying their computing requirements. Since they were expensive machines, the resources had to be managed having basically into account performance metrics: throughput, response time, load-balancing, etc. These systems were relatively easy to manage in a central way, since they had a single access point, and their usage was restricted to users of the same organization (or a few ones).

The big mainframes paradigm [1], where users own their computing resources, is being progressively transiting to a more utility-driven paradigm [2], where users do not own their resources and pay for the usage of remote resources. Utility computing has the advantages of other public utilities, like water or electricity. This is, the clients do not require spending neither an initial expenditure for the hardware nor maintenance costs (hardware, employees, electricity, physic space, cooling...).

The Grid [3] and, more recently, Cloud Computing [4] are the most promising current implementations of Utility Computing, the first in scientific and academic environments, and the second in the world of the enterprise. This new evolution has made the classical Resource Management mechanisms very inefficient because some reasons:

- The complexity of finding an optimum resource allocation is exponential, in huge systems like big clusters, data centres, or Grids, is growing very quickly [5]. Here are enumerated some causes:
  - A set of computers must be integrated via software and networking technologies.
  - Propelled by the transition from single- to multi-core processors, the average size of the clusters is being multiplied in few years. This implies more capacity and, as a consequence, more applications to manage.
  - New processing units are being added: GPGPUs, FPGAs, etc...
  - There is a transition from homogeneous resources and applications (mainly CPU-intensive) to heterogeneous resources and applications (web services, I/O intensive).
  - With the explosion of the implantation of Virtualization technologies [6], a single physical computer now becomes multiple logical computers, phys-

    ical CPUs can be split into multiple logical CPUs in runtime, Memory can be resized dynamically, etc...

  - Grid and Cloud systems are not static: they are dynamically growing and decreasing, since new resources are hot plugged or removed.

- Multiple, remote users can now access the systems, and every one has its own preferences, which can be in conflict with the preferences of the other users.

- The idea of business is being introduced: some providers will sell their resources to the clients, which are willing to pay for accessing them. This introduces new high-level metrics: Quality of Experience, Quality of Business [7]... It is very difficult to manage resources having into account these metrics because they can be different for every provider and client, and the central resource manager does not have to know what good Quality of Experience of a user is.

- If the central resource manager breaks, the whole system gets useless. This is a big waste of resources in large systems.

Having into account these arguments, large systems seem to be too complex to be managed centrally. This thesis defends the usage of decentralized resource management as a paradigm to deal with the complexity. Concretely, market-based resource management is proposed by the next reasons:

- In utility computing, the possibility of doing business will motivate service providers to offer their resources in the system and give a Quality of Service (QoS) according to their real capacity.

- We can let the users reserve a spatial and temporary portion of the system, and we know that market mechanisms will obligate them to adjust their allocation to their real requirements.

- It is relatively easy to implement in a decentralized architecture.

- The complexity is reduced, because participants enter in the market looking for the satisfaction of their own necessities, and they do not need to know the global status of the system to maximize their utility.

Decentralized market-based resource management is starting to be implemented in open scenarios, such as Grid or Cloud computing, but this paradigm could also be implemented in closed systems, like Server Farms of a single organization, even when providers do not have business objectives.

Figure 1 shows the generic logical architecture of a market-based resource allocation system. In this architecture, Brokers that represent Service Providers or Clients participate in a market to sell or buy their services. When the clients find there their necessities, a **negotiation process** is started to establish the terms of the contract (QoS, price, time slots...). If both parts reach an agreement, the terms of the contract are specified in a Service Level Agreement (SLA) and the client application can acquire the bought resource. During the usage of the resources, the

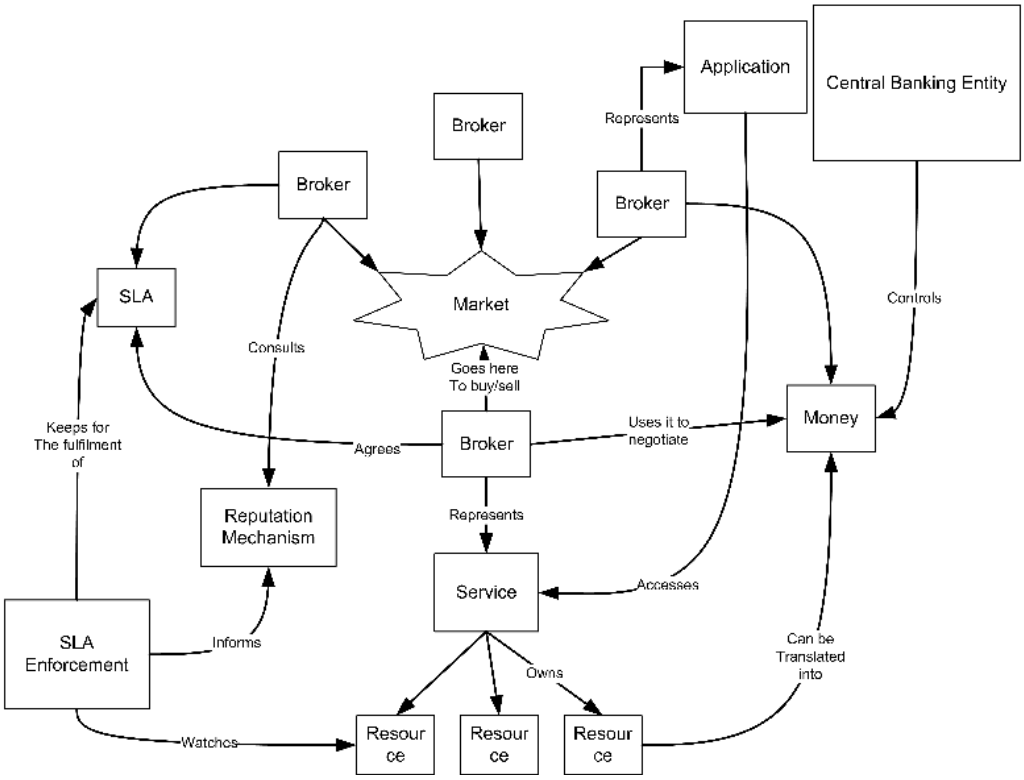

Figure 1.1: Market-based Resource Self-Organisation entities.

component called SLA Enforcement watches for the correct fulfilment of the terms of the SLA, and penalises the buyers or the sellers if they violate the SLAs.

From the provider side, this model has two main limitations caused by the same problem: the isolation between the Provider Broker and its representing entity, the service provider that owns the resources and services to sell. From the side of the negotiating brokers, the problem is the difficulty to perform accurate negotiations, because they do not have knowledge about the status of the production fabrics. Current implementations of Provider Brokers base their prices in function to some economic information, such as the status of the demand or the prices that are being paid in the last sales. This causes that the provider only considers the past data, and cannot take decisions based in future predictions about the market or the resources to propose accurate prices or predict changes in the market or the system before they happen.

On the other hand, the Resource Manager tries to enforce the agreed SLAs only by considering the QoS requirements, without considering the economic implications of its decisions. For example, maximising the number of SLAs fulfilled can be dramatic for the revenue of the provider if the SLAs that have been violated (by, for example, a system failure) are the SLAs that report more economic profit to the provider.

The technical decisions of the Resource Managers have direct influence in the achievement of the Business Objectives (e.g. violating SLAs will have impact in the revenue of the provider). Likewise, the business decisions of the provider Brokers may be inaccurate if they do not consider the information about the resources that they are selling. This thesis proposes to establish a bidirectional communication

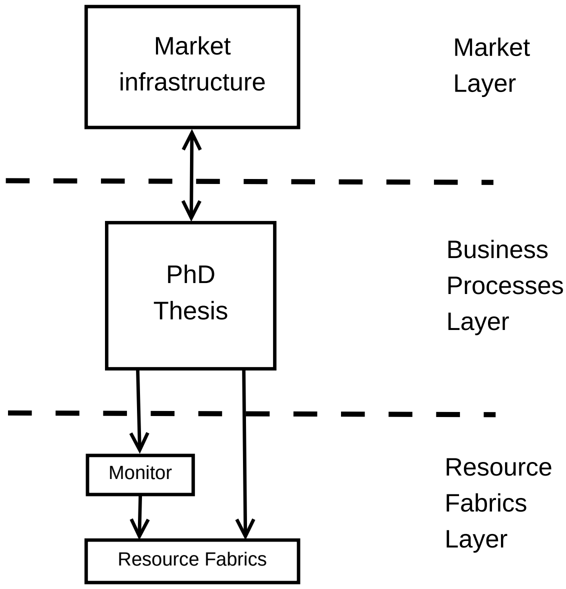

Figure 1.2: Multi-layered architecture of the system of this Thesis

flow between business and resources layers in order to improve the technical and economical performance of both: business decisions can be more precise if the brokers have information about the resources, and resource management decisions can help achieving the business objectives of the provider if the Resource Managers consider economic models and policies in their decisions. The gap between the Market and the Resource Fabrics (see Figure 1.2) is filled by providing bi-directional communication between the two layers by adding an intermediate layer: the Business Processes Layer.

This layer model is inspired by the one proposed by Moura et al. [8] and is divided in three main layers:

Resource Fabrics layer Provides the hardware infrastructure to run the jobs and services that will be offered to the clients.

Business Process layer Those processes which depend on Resource Fabrics, directly or indirectly. This layer captures how technical metrics or events affect changes in business metrics or vice-versa [9, 10, 11]. This layer also will assist in the calculation of business metrics for the Market Layer.

Market Layer Models business entities and behaviours, such as the negotiating Brokers of client and provider.

The work in Business Processes Layer will combine the purely economic knowledge (because is in direct contact with the economic layers of the marketplace) and the plain resources management (because it manages directly the resource fabrics) to help economic brokers to perform better negotiations and enforce the resource management, not only having into account performance but also economic goals.

This Layer must help maximising the achievement of the Business-Level Objectives of a particular provider. These can vary across the different providers. Most of providers that enter in a market have as a main objective to earn money from their participation in the market. So it is really important that the Business Process layer helps to manage the Revenue adequately for maximising the economic benefit. However, this may not be the only objective. Corporations that use their internal clusters and data centers might want to offer their spare resources to cushion their infrastructure expenses. In this scenario, the provider must implement a Client Classification mechanism prioritized over Revenue Maximisation, that guarantees that preferent users will have the maximum guarantees of Quality of Service over the external users.

Both Revenue Maximisation and Client Classification must consider the concept of reputation: maximising some of both objectives can lead to an impact in the reputation that a provider has in the market. The short-term maximisation of these objectives can affect their achievement in the mid-term. For example, cancelling a SLA for providing enough resources to the fulfilment of another preferent SLA will decrease the reputation of the provider and, in consequence, make the potential clients less willing to use the offered resources.

Considering all the explained earlier, the main requirements at Business Process Layer are [9]:

- Decide whether incoming tasks are accepted or not, based of the usage of the resources for a given time slot, the client priority, and the sale price.
- Calculate prices for offered jobs and services based in current market status, current resource usage and predictions about the impact of a job in the usage of resources.
- Check that the accepted SLAs can be kept. If one or more SLAs can not be fulfilled it takes the decision to suspend or cancel jobs to ensure the fulfilment of the other SLAs and maximise overall revenue.
- Be responsible of the communication with the local resource managers and influences the local resource management to achieve a more efficient global resource use.
- Offer quality of service and be able to deal with situations in which parts of the resources fail.
- Allow the Business Administrator of a particular provider to tune up the behaviours enumerated here by means of Business Rules definition.

Taking into account the previously defined requirements, this thesis defines five research topics to improve the economic feasibility of Cloud Computing by establishing an intermediate layer that uses the information flow between Market and Resource layers:

SLA Negotiation Implementation of existing economic models for negotiation and application of them to the negotiation of services and resources between computing agents. Dealing of both economic and technical information is crucial here for performing accurate negotiations that report benefits to the provider and whose resulting SLA can be fulfilled correctly.

SLA Enforcement This thesis will provide the Resource Manager with economic models and business knowledge in order to make economically sound decisions when allocating, migrating, pausing or cancelling resources. This aspect will not only be useful for maximising the revenue of the provider, but also to achieve other possible Business-Level Objectives such as Client Classification.

Rule Support The objectives that an EERM pursuits depends on the policies of its organisation. So it is required that the EERM component include support for custom rules, defined by the Business Administrator. It is a key fact, because not all the providers that enter in a market have the same objectives and solutions for custom problems. However, classic Rule Systems are pretty static since a market environment is continuously changing. The main contribution of this thesis in this aspect is that the Resource Manager will be able to continuously tune-up their rules to adapt them to the changes in the market and keep a good ratio of fulfilment of SLAs.

Revenue Management This thesis pretends to be a step forward is the management of the revenue. This can be performed in two moments: negotiation time, by providing the more suitable prices to acquire the sales despite of the competence in the market, and execution time, by managing the resources for the maximisation of the economic benefit. Results will show how implementing new dynamic methods for negotiation and enforcement increase the economic profit of the provider.

Trust Management Here are included all the mechanisms to force both clients and providers to fulfil their agreed SLAs, by providing in negotiation time some previous information about the negotiating parties that will help decide if a concrete client or provider is trustworthy or not. This thesis show how these aspects have a strong correlation with the achievement of Business Objectives, and propose methods for implementing some of them, such as Client Classification or maximisation of the reputation of the provider.

All the enumerated features are described in detail in the next chapter.

Chapter 2

Research Goals

The implementation of the functions mentioned at the end of previous chapter involves five research fields: three of them are the main topics, and the other two which are transversal to the entire thesis (this is, their results are used by the three main research goals).

The three main research goals are: *BLO-Driven SLA Negotiation*, *BLO-Driven SLA Enforcement* and *Rule support for Management of both Business and Resources*. The transversal topics are *Revenue Management* and *Trust management*.

2.1 BLO-Driven SLA Negotiation

Since brokers that negotiate for the sales of services are autonomous agents (this is, they communicate between them and take decisions without human intervention), it is needed to provide them some economic models and intelligent behaviour so they are able to take the best decisions for their represented actors in the market (client applications or service providers) and maximise their utility.

Examples of this economic behaviour are client classification, dynamic pricing, negotiation strategies, revenue maximisation, and so on. This thesis does not intend neither to evaluate them from an economic perspective nor invent new economic models and strategies; but some of these models will be explained and used for implementing negotiation scenarios and evaluate their performance.

This thesis uses existing economic models for negotiation and applies them to the negotiation of services and resources between computing agents: when the resource broker negotiates an SLA with the broker of a client, it has into account some economic terms, such as price, penalties for contract violation, time, etc. But there are other terms in the SLA, which are more technical, and also can have influence in the economic terms, specially those related with the Quality of Service (QoS) (e.g. throughput, response time...) or those related with the sales of plain resources (number of CPUs, speed, memory...).

2.1.1 Using Multiple BLOs within Non-additive Utility Functions

Traditional negotiation models for Utility computing are based in the models proposed by Raiffa [12] and Faratin [13]. This model is pretty easy to manage and calculate the maximum and minimum utilities. However, it is an *additive model* which assumes that all the factors are independent from the others.

Let S be the SLA under negotiation, this thesis proposes a *nonadditive utility function* U in the form of the shown in Equation 2.1.

$$U(S) = \sum_{i=1}^{m} o_i u_i(S) \tag{2.1}$$

Where m is the number of goals for the Provider, such as revenue maximisation, reputation, performance maximisation, high occupation of resources, or satisfaction of certain type of users. u_i is the sub-utility function that defines how much will be the objective i satisfied, and o_i is a number between 0 and 1 that defines the priority that the Provider assigns to the particular objective. It must be considered that $\sum_{i=1}^{m} o_i = 1$.

Although Equation 2.1 is similar to an additive function, actually it is not. Instead of calculating each of the sub-utility functions as a function of a single SLA term and finally add them up, Equation 2.1 calculates all the sub-utilities as a function of the whole SLA, because the different objectives are not independent from the others and, for example, revenue maximisation can affect negatively the Client satisfaction.

2.1.2 Resource-level Support

For a purely-economic resource broker, it is very difficult to quantify the terms of the SLAs, since it has not enough technical knowledge about the status and punctual capacities of the resources. This can be seen also in real world sales, where brokers need to know not only the main characteristics of the products but also, for example, the capacity of production for a specified good in a concrete time, or the associated costs to the production and distribution process for a product that is customized for a concrete client. To illustrate this, imagine a factory of cookies: there are several sellers that look for potential clients, offering their products to be sold, for example in small shops, and also offering their production resources for creating cookies for other companies, for example those generic brands for big supermarket chains. When the sellers of the factory negotiate the terms of the contracts, they should talk first with the people that controls and schedule the production, for example:

- When the amount of production and the deadlines are negotiated, brokers must have knowledge about the production capacity of the factory, and about the reserved production for other current clients, in order to not order too many cookies than they actually are able to produce.

- When the price is negotiated, brokers have more liberty of election, but they also need to know the costs of production in function of the ingredients, the energy, the number of workers involved in the production, their salary, etc. And these costs can vary in time (due to the fluctuations in food and energy prices).

This example illustrates how a sales broker need to communicate with the agents involved in the production of goods in order to negotiate accurate contracts and get profitable sales. This thesis defends the idea that Market-Based Utility Computing is not an exception.

2.1.3 Dynamic Pricing

Economic markets are composed of a set of tradable goods (supply) offered by the market, and a set of buyers (demand) who ask for a subset of the supplied goods. In a Cloud Market the supply are the computer resources, such as storage systems, web servers or clusters, and the demand are the users and applications which are willing to pay for accessing these resources. The goal of dynamic pricing is to find an efficient allocation of the supply among the demand.

Law of Demand states that the higher price of a good, the less people will demand this good, and Law of Supply states that the higher price has a good, the higher quantity will be supplied, because selling higher amounts of goods at higher prices will increase the revenue. Having this into account, the supply and demand can be controlled by changing the prices of the goods: at lower prices, supply is low and demand is high, and at higher prices, supply is high and demand is low. This leads the market to a disequilibrium point.

In a market with disequilibrium, if the prices are too low the demand can not be satisfied because the supply is too scarce, and if the prices are too high the supply can not be sold completely because demand is too low. The efficient allocation that dynamic pricing pursues is achieved when it reaches the equilibrium point, and both demand and supply are satisfied.

Once shown the importance of pricing for finding the equilibrium of a market, two approaches to find the optimal pricing can be used: tâtonnement and auctioning [14].

2.1.4 Resource Overprovisioning

A classical RM will refuse an incoming reservation request from a client if there are not enough free resources for the requested time period. *Free resources* are those resources that are not reserved by any other client. However, these resources are often idle in concrete time periods (such as the early morning). If the RM allows to reserve resources that are already reserved by other clients but not used, the revenue could be noticeably increased.

When resources are overprovisioned for future tasks, it is needed an accurate predictor that is able to estimate the future usage of the resources with a small error rate, based on historical data. Based on this prediction, the possibility of overprovision a resource is calculated and incoming jobs are accepted or refused.

2.1.5 Risk-aware Pricing

Risk is defined in ISO 31000 [15] as the effect of uncertainty on objectives (whether positive or negative). Risk management can therefore be considered the identification, assessment, and prioritization of risks followed by coordinated and economical application of resources to minimize, monitor, and control the probability and/or impact of unfortunate events or to maximize the realization of opportunities. Risks can come from uncertainty in financial markets, project failures, legal liabilities, credit risk, accidents, natural causes and disasters as well as deliberate attacks from an adversary. Risk Management is a key transverse topic in this thesis, since the most of the negotiation and management decisions of a resource manager have an inherent risk associated. The risk must be managed in order to minimise the economic and technical impact of uncertain situations.

One of the usages of Risk in this thesis will be in pricing decisions. The idea is to translate to the pricing process some risks such as system-failures, client dishonest behaviors, etc. that can affect negatively the revenue of the provider. Having such hazards or opportunities into account in the pricing decisions can lead to a maximisation in the revenue and the avoiding of economic losses.

2.2 BLO-Driven SLA Enforcement

It is demonstrated that maximising utility in negotiation time can lead to a better achievement of the business objectives [16, 17, 18].

However, some unforeseen events can happen in execution time that are not taken into account in negotiation [10, 11], for example:

- The Resource Manager does not have the ability to decide which SLAs must be accepted or rejected. It is only used for consultative purposes. Even if the Resource Manager advises that it cannot fulfil an incoming SLA, economic agents could decide to send it to the Resource Manager because they consider that it is strategically necessary.

- The Resource Manager uses a predictive model to calculate the impact of a task execution. Any prediction system has a margin of error, and the system could accept a task which cannot be fulfilled, resulting in system overload.

- An adverse situation could reduce the number of available resources; for example, some nodes of an available cluster could crash. The Resource Manager must manage the situation to minimize the economic losses as consequence of having too much jobs for the available resources.

In the cases described above, the service provider would have a reduced number of resources and the system would become overloaded. In consequence, some SLAs of accepted jobs might be violated and the service provider would pay some penalties to the clients whose SLA is not fulfilled. It is necessary that the Resource Manager take reactive actions in order to maximise the achievement of BLOs in runtime.

However, BLO-Driven Resource Management do not have to apply necessarily only in error scenarios such as those described previously. Actually resources can be managed to maximise some BLOs even when everything works normally, but to take advantage over traditional resource management. For example, if client applications do not use all the resources that they have purchased for, the Resource Manager could try to resell part of these purchased (but unused) resources to other potential clients in order to maximise its economic profit.

2.2.1 Selective SLA Violation and Cancellation

When system is overloaded, many SLAs will be violated, so the Resource Manager will have to pay an uncertain amount of money to its clients. Arrived to this scenario, this thesis proposes the idea of selecting which SLAs to violate temporarily or cancel permanently in order to minimize the economic impact of such violations [10, 11].

2.2.2 Task Redistribution

Current works in virtualisation show the low cost of migrating tasks across a virtualized environment [19]. Migration of virtual machines is a key tool that will allow the creation of policies to be used in addition to the one explained in previous subsection: redistribution of tasks across the pool of resources, in order to minimize the probability of violations [10, 11].

2.2.3 Resource Reassignation

Services are negotiated in terms of high-level metrics, such as response time, throughput, quality, etc. The Resource Manager only manages low-level metrics such as CPU or memory. That implies that decomposition from high-level to low-level metrics must be performed in negotiation time so that the provider-side negotiating part can decide if a requested service is suitable.

However, this SLA-decomposition has some uncertainty and it is possible to not allocate enough resources for accomplishing a task. In this case, it is necessary to detect errors on SLA decomposition and reassign in real time some free resources to allocated task [20].

2.3 Rule Support for Managing both Business and Resources

The BLOs that an Resource Manager pursuits depend on its organisation policies. So it is required that the Resource Manager component includes support for custom rules, defined by the Business Manager of each organisation.

This thesis will contribute with the description and evaluation of rules for both negotiation and resource management. Another contribution, which is more qualitative, is the creation and evaluation of a framework for supporting dynamic rules, which are self-adapted and self-optimised in function of the current status of the market and the resources.

2.3.1 Rule-based SLA Negotiation Framework

Negotiation models such as those described in section 2.1.1 have been demonstrated to be effective for achieving objectives that can be easily expressed in an utility function. However, there are some situations where the negotiation must be expressed in a more procedural way. For example, when defining highly differentiated ways of treating the negotiations depending on the status of the system or the market.

This thesis will provide a framework for integrating rule engines within an Utility Computing Market negotiation broker, and the creation and evaluation of rules for performing efficient negotiations.

2.3.2 Rule-based Control Theory for SLA Enforcement

Once a SLA is agreed in the negotiation phase and the tasks are submitted to the resource fabrics, it is needed that the Resource Manager surveys for the good fulfilment of the contracts. However, there are some situations that potentially can lead for the breakage of the SLAs, such as errors when predicting situations in the negotiation phase, system failures, etc. In these situations, the SLAs and resources must be managed for minimizing the penalties derived from those setbacks. Some standard measures can be done, such as resource redistributions, but there are some other actions that must be done in function of the business preferences of the provider, such as selective SLA violations in function of the client or the revenue. These actions can be expressed as rules and configured by the business administrators.

This thesis will provide a framework for integrating rule engines within an Utility Computing Market negotiation broker, and the creation and evaluation of rules for performing efficient negotiations.

2.3.3 Self-adapting and Self-optimising Business Rules

The rules and utility functions defined to pursue the BLOs might not be 100% accurate, or can become deprecated when time goes on and the scenario changes. The proposal of this thesis is to tune up the rules progressively, having into account the inputs, the monitoring data, and the error in the expected objectives versus the actual objectives. This approach is based on control theory.

First, a set of situations those act as triggers to launch the policies must be defined, for example:

- Resources usage exceeds a threshold (e.g. the 80% of its capacity).
- A SLA is violated by the provider or the client.
- An inefficient workload balancing is detected.

When the previously defined triggers are activated, the Resource Manager launches some policies to take reactive measures, such as resource redistribution. In addition, and that is where the contribution of this thesis relies, these rules can refer to the rules themselves, and describe how to change them in function of their effectiveness: change priorities, thresholds, weights of utility functions, reservation prices, etc...

2.4 Revenue Management

From Gabor Forgacs [21]: Revenue Management has become part of mainstream business theory and practice over the last fifteen to twenty years. Whether we call it an emerging discipline or a new management science (it has been called both), revenue management is a set of revenue maximization strategies and tactics meant to improve the profitability of certain businesses. It is complex because it involves several aspects of management control, including rate management, revenue streams management, and distribution channel management, just to name a few. Revenue management is multidisciplinary because it blends elements of marketing, operations, and financial management into a highly successful new approach. A revenue manager frequently must work with one or more other departments when designing and implementing revenue management strategies.

Since Revenue is the main BLO in the most of the providers within Utility Computing Markets, all the research topics within this thesis must address the revenue management and maximisation as a key aspect. Next are explained some subtopics that will be applied in this thesis. It can be seen their close relation to most other topics such as SLA Negotiation or SLA Enforcement.

2.4.1 Modeling Revenue in Service Level Agreements

This part will be in charge of evaluating and improving the current ways of expressing revenue within Service Level Agreements, such as maximum revenue, maximum penalty (in case of SLA violation), transitions from maximum revenue to maximum penalty, deadlines for finishing tasks, etc.

2.4.2 Maximising Revenue in Negotiation Time

Once the revenue is correctly modelled for its expression within SLAs, it will be possible to apply this model for optimise the negotiation rules and policies. Having into account risks, uncertainties, but also strengths of the provider, this thesis provide negotiation strategies for maximising the revenue of the provider, such as different pricing algorithms, risk management applied to pricing, reputation management, resource overprovisioning, etc.

2.4.3 Maximising Revenue in Execution Time

When resources are scarce and the system gets overloaded, some reactive actions must be performed in order to minimize the economic impact of penalties in the revenue of the provider. Revenue Management takes the role of possible objective when defining policies and behaviours for managing resources in execution time, such as those explained in section 2.2.

2.5 Trust Management

Utility computing markets allow the participation of multiple agents from different organisations: any provider can join the market and sell its own resources or services to market clients. Before the sale is performed, both parts must negotiate and agree the terms of a contract: the Service-Level Agreement (SLA) [22]. But the fact is that this SLA can be violated by the service provider as a consequence of a system failure, an error in the negotiation (e.g. provider did not calculate well the expected system workload and it gets overloaded), or simply because the service provider is dishonest with the customers.

To avoid this, a mechanism for putting pressure on providers to fulfil the agreed SLAs is needed. The most usual mechanism is adding terms in the SLA for describing penalties, specifying the amount of money that the provider must pay to the customer if any of the SLA terms is violated [23]. This paper focuses on a complementary mechanism for enforcing SLA fulfilment: reputation and trust [24]. If a provider violates an agreed SLA its reputation decreases, and this information is accessible to market participants in future negotiations.

This thesis proposes the creation and evaluation of mechanisms for enabling Trust and Reputation within Business and Resource Management. This data can be taken into account for performing better decisions, for example, when negotiating for SLAs and the reputation or the clients are taken into account for evaluating the risks of SLA violations.

2.5.1 Reputation

This thesis uses reputation as a mechanism for providing trust capabilities within Utility Computing Markets. In the work performed, new reputation mechanisms are created and evaluated. It is important to know how reputation has influence in several BLOs, such as revenue [25, 16, 18], and take into account it in order to minimize the risk of harmful consequences in the achievement of BLOs due to the decrement in the reputation of service providers.

2.5.2 Client Classification

Some corporative clusters that most of the time have more resources than the amount that they actually need would like to sell their spare resources in a market for minimising the economic expenses of maintaining their own resources.

However, it sounds reasonable to differentiate the business objectives and rules within different types of users, such as external users, corporative users, intradepartmental users, etc.

This classification can be performed in various manners, such as price discrimination, different priorities on job acceptation, management or cancellation, different Qualities of Service, etc.

2.5.3 Trust-aware Policies

This thesis also will work in the way of integrating trusting within the rules and policies, to use it as an objective that coexist with other business objectives.

Rules for calculating the influence of current decisions in the future reputation or rules that prioritize Client Classification over other objectives will be implemented and its effectiveness will be evaluated through several experiments.

Chapter 3

State of the Art

3.1 Research projects on Market-Oriented Utility Computing

This section describes some research projects related with the market allocation paradigm applied to Utility Computing.

3.1.1 Fingrid

Fingrid [26] is a German consortium that studies the feasibility of applying Grid computing to business. Their main tasks are divided in several interconnected tasks:

- Evaluate the market and compile empirical recommendations and investigate service-oriented Grid cases.
- Come up with different prototypes that are used to demonstrate the feasibility of our concepts in terms of security, accounting, monitoring and pricing.
- Evaluate different types of pricing mechanisms that seem to be applicable for the financial service Grid. The most promising pricing mechanisms are then implemented in a prototype. It is calculated the willingness to pay for different Grid services and determined an optimal tariff structure.
- To deal with the problem of accounting which is a preceding step for a professional pricing and billing of Grid services.
- Discuss how a financial on-demand Grid should utilize both unused resources within a department as well as allow the spontaneous discovery and use of computational resources in other departments or even other organizations.
- Investigate the issues involved for providing support for service level agreements in financial applications. The challenges to be addressed are to leverage on the work on monitoring and managing infrastructure in order to enable an SLA management framework that can use this information, thereby enabling an autonomous SLA management framework.

3.1.2 GridEcon

GridEcon [27] is an European Community-funded project that offers market place technology to allow many small providers to offers their resources for sale. It designs the technology that is needed to create an efficient market place for trading commoditized computing resources as standardized Virtual Machines (VM). The market mechanism used has been designed to be simple for participants and also economically sound. The later is concerned with inducing the right economic incentives to participants and avoiding unwanted strategic behaviour leading to market dominance with large players. The GridEcon project also designs a series of value-added services on top of the market place (e.g. insurance against resource failures, capacity planning, resource quality assurance, etc...), ensuring quality of the traded goods for Grid users.

When a buyer wants to acquire a resource, it sends a bid to the market by specifying its requirements according to the next terms:

- The type of resource (VM) required.
- The quantity of resources required.
- The start time of the interval for using the resources.
- The time duration of using the resources.
- The price expressed in €/min/unit.
- The time limit until which the bid is valid. If this time limit is reached without the bid being matched, the bid is removed from the system.

In order to keep the definition of the bid as general and flexible possible, instead of allowing only fixed values for the number of VMs and the time duration, it is allowed to specify whether these constraints should be met with equality, or $\leq$ or $\geq$.

Resource providers can also send its spot and futures offers to the market, in order to describe what resources they are selling:

- The type of resource (VM) offered.
- The quantity of resources offered.
- The start and end time of the interval when the resources are available
- The price expressed in €/min/unit.
- The time limit until which the offer is valid. If this time limit is reached without the offer being matched, the offer is removed from the system.

Once bids and offers are in the market, a matching algorithm is used in order to help buyers to find the most suitable provider for them, and *vice-versa.* The trading is performed by means of a continuous double action, where both bids and offers that are submitted by traders are placed in queues and ordered in decreasing order of price for bids, and in increasing order of price for offers. This will guarantee that the buyers that are available to pay more and the sellers that put their resources at lower prices will be the first on finding their matches.

3.1.3 SORMA

The Self-organising ICT Resource Management (SORMA) [28] is an EU IST [29] funded project aimed at developing methods and tools for efficient market-based allocation of resources, using a self-organising resource management system and market-driven models, supported by extensions to existing grid infrastructure. Topics addressed include Open Grid Markets, economically-driven middleware, and intelligent support tools.

Unlike traditional grid environments, jobs submitted to SORMA are matched with available resources according to the economic preferences of both resource providers and consumers, and the current market conditions. This means that the classic grid job scheduler, which is based on performance rules, is replaced by a set of self-organising, market-aware agents that negotiate Service Level Agreements (SLAs), to determine the 'best' resource allocation to fulfil both performance and business goals. In SORMA, an Economically Enhanced Resource Manager (EERM) [11] exists at each resource provider's site, and acts as a centralised resource allocator to support business goals and resource requirements.

3.1.4 Grid4All

The Grid4All [30] project promotes the concept of a democratic Grid, accessible to modest groups of end-users such as schools, families, non-governmental organizations, or small businesses. It enables to put together people and computing resources to form Virtual Organizations (VO): a virtual collection of users or institutions that pool their resources into a single virtual administrative domain for a common purpose. Virtual Organizations can also trade resources among different VO on a decentralized market place.

In the area of data services, compared to previous Grids, the Grid4All architecture provides with a minimal administration enhanced support for content sharing and collaboration within groups. Semantic search and ontologies are used to locate and select among diverse resources and services.

3.1.5 Catnets

The Catnets project proposes a market-based approach based on the Catallaxy concept [31]: a market order without planned ends, characterized by the *spontaneous order* which emerges when individuals pursue their own ends within a framework set by law and tradition. The function of government is to maintain the rule of law which guarantees fair and equal procedures, but is neutral as to goals.

The advantage of Callaxy is that does not need to support for centralized brokers: it uses a "free market" self-organisation approach, which enables prices within the market to be adjusted based on particular demands being placed on particular scarce services. To implement this decentralized and highly chaotic market, Catnets adopts a P2P approach, which allows establishing a symmetric interaction between peers, and allocate dynamically the communication paths in function of the changes in the network topology.

3.1.6 Nimrod-G

Nimrod-G [32] is a tool for automated modelling and execution of parameter sweep applications over global computational Grids. It provides a simple declarative parametric modelling language for expressing parametric experiments. A domain expert can easily create a plan for a parametric experiment and use the Nimrod system to submit jobs for execution. It uses novel resource management and scheduling algorithms based on economic principles. Specifically, it supports user-defined deadline and budget constraints for schedule optimisations and manages supply and demand of resources in the Grid using a set of resource trading services.

Nimrod-G provides a persistent and programmable task-farming engine (TFE) that enables plugging of user-defined schedulers and customised applications or problem solving environments (e.g., ActiveSheets) in place of default components. The task-farming engine is a coordination point for processes performing resource trading, scheduling, data and executable staging, remote execution, and result collation. In the past, the major focus of our project was on creating tools that help domain experts to compose their legacy serial applications for parameter studies and run them on computational clusters and manually managed Grids. It is focused on the use of economic principles in resource management and scheduling on the Grid in order to provide some measurable quality of service to the end user.

3.2 BLO-Driven SLA negotiation

Howard Raiffa established and compiled the mathematical basis of the negotiation models in his book *The Art and Science of Negotiation* [12]. This book classifies the different negotiation models in base to the characteristics of the environment and the negotiated goods. It will be widely referenced in this thesis, and in the most of the other works about negotiation.

Faratin et al. [13] applied and extended some existing models for service-oriented decision functions in bilateral negotiations between autonomous agents. It concentrates in many-parties, many-issues, single-encounter negotiations with an environment of limited resources, which is a variation of the *bilateral negotiation model* introduced by Raiffa: let $i \in \{a, b\}$ the negotiating agents and $j \in \{1, \ldots, n\}$ the issues under negotiation, let $x_j \in [min_j, mmax_j]$ be a value for issue j in the range of its acceptable values. Let ω_j^i the importance of issue j for agent i, and $\forall i, \sum_{1 \leq j \leq n} \omega_j^i = 1$. If $V_j^i : [min_j, max_j] \rightarrow [0, 1]$ gives the score that agent i assigns to a value of issue j in the range of its acceptable values, it is possible to define a scoring function of an agent for a contract, that is, for a value $x = (x_1, ..., x_n)$:

$$V^i(x) = \sum_{1 \leq j \leq n} \omega_j^i V_j^i(x_j) \tag{3.1}$$

Since computing services are qualitative in nature rather than quantitative, Faratin extends this model by adding qualitative values and associates fuzzy sets to them [33] in order to express better the quality in the negotiations.

Once the agents have determined the set of variables over which they will negotiate, the negotiation process between two agents consists of an alternate succession of offers and counter offers of values for the x, until an offer or counter offer is accepted by the other side or one of the parties terminates the negotiation. Faratin et al. demonstrated what this thesis affirms: negotiation tactics must be responsive to changes in the environment. Several tactics must be applied and dynamically changed in a same negotiation. They define *time-dependent tactics* (the acceptance value depends on the remaining negotiation time), *resource-dependent tactics* (the scarcer is the resource, the more urgent is the need for an agreement), *resource-estimation tactics* (the agent becomes progressively more conciliatory as the quantity of resource diminishes) and *behaviour-dependent tactics* (the next offer is computed based on the previous attitude of the negotiation opponent, useful for co-operative problem-solving negotiation settings).

By some experimental simulations, they proved that agents negotiating use their model were guaranteed to converge on a solution in a number of well defined situations and, with respect to tactics, they also discovered that: (i) irrespective of short of long term deadlines it is best to be a linear type tactic, otherwise and imitative tactic; (ii) tactics must be responsive to changes in their environment; and (iii) there is a trade-off between the number of deals made and the utility gained which is regulated by the initial offers.

The work in this thesis tries to extend the model of Faratin by extending the information extracted from the resources and used in the negotiation, and by having into account other economic factors, such as reputation, risk management, etc. The other main difference is that the work of Faratin was limited to a concrete scenario: client and provider brokers meet to negotiate for a concrete type of resource. The work in this thesis must consider the service discovery (this is, a market place) and the fact that agents can negotiate for a huge range of services.

Ouelhadj et al. [34] introduce a protocol for robust scheduling in Grid Computing based in the Contract Net Protocol [35]. The described architecture is similar to the current Grid Market systems, but it has the particularity that the SLAs are negotiated at two levels:

Meta-SLA negotiation The User Agent (UA) requests the execution of jobs on the Grid and negotiates a Meta-SLA with the Super Scheduler (SS) agents. There is one SS agent per each entity that owns a resource pool. The Meta-SLA contains high-level description of jobs supplied by the user and may be refined or additional information may be added in the final agreed SLA at the end of the negotiation. Uncertainty in user requirements such as time and cost constraint is represented by Fuzzy numbers.

Sub-SLA negotiation Each resource in a resource pool has a Local Scheduler (LS) agent, which is the responsible for scheduling the jobs that arrive. In the Sub-SLA negotiation level, the SS agents negotiate sub-SLA with the LS agents. The SS agents decompose the meta-SLA into its low level resource attributes, sub-SLAs which contain low level raw resource description such as processes, memory, processors, etc.

Other interesting feature in the work of Ouelhadj et al. is the possibility of a re-negotiation of the SLAs. Re-negotiation is useful when considering some uncertainties: presence of high-priority jobs, changes in the QoS requirements, resource failures, etc.

This thesis took some ideas from the Meta-SLAs of Ouelhadj's work and from WS-Agreement [22]: SLA negotiations between client and provider brokers are performed by using high-level QoS metrics. In a lower level, an Economically Enhanced Resource Manager (see section 4.2.1) helps in the negotiation of the SLAs by decomposing the high-level SLOs into low-level metrics, to calculate if a particular service can fit in the resources, given their status.

Vulkan et al. [36] evaluate the efficiency of English Auctions in the negotiation for services in multi-agent environments. Like in this thesis, they assume that the negotiations are initiated by the client. In addition, they introduce a *pre-auction* protocol for allowing the provider to initiate an auction when the client does not do. The winner of this auction is offered to the client as "take it or leave it". The difference with this thesis is that they use an English Auction instead of a direct negotiation and the presence of pre-auction protocols. However, the way they represent the negotiation terms is very similar to this thesis: a price and a set of SLOs.

3.3 BLO-Driven SLA Enforcement

Yeo et al. [37] propose a model for market-based cluster computer with some elements common to EERM. The cluster nodes are connected to a central manager which incorporates other sub-components for performing Pricing, job scheduling, monitoring and dispatching, user admission control, and so on. The main difference with EERM is that EERM is conceived to be integrated into a higher level grid market, so it does not implement some market functions such as accounting, billing, or identity provisioning.

Dube et al. [38] establishes different ranges of prices for the same resource and analyse an optimisation model for a small number of price classes. Their proposal is similar to the proposal of this paper of establishing Gold, Silver and Bronze ranges and optimising their QoS performance giving priority to the contracts that report the highest economic profit.

Freedman et al. [39] focus on Peer-to-Peer (P2P) content distribution by identifying explicitly highly demanded files and rewarding most those peers sharing highly demanded content. They use a market-based mechanism for allocating network resources across multiple files and play with the Law of Offer and Demand for encouraging providers to sell most scarce high-value resources (only files, but not CPUs or Memory). Their system is also designed so that the buy client chooses files consistent with its best interests, since it seeks to download at the current minimum price. These policies are

Menasce et al. [40] demonstrated the importance of managing the resources having in mind the business objectives. They maximise the revenue of e-Commerce servers by calculating dynamically Customer Behaviour Model Graphs and prioritising the workloads of the users that are willing to spend more money. However, this model is not applicable to the scenario of this paper, since the Cloud Provider supports more generic types of workloads, not restricted to online shops, and does not interact with human customers, but with other client machines that automatically buy resources in a market. Poggi et al. [41] introduce a similar approach, where QoS for user sessions is shaped based on the commercial intention of the users.

Pueschel and Neumann [42] also use the figure of an EERM for optimising the revenue of a Cloud Manager. It applies policies as a heuristic and demonstrates their achievements on revenue maximisation or client classification when applying economic enhancements. Compared to work of Pueschel and Neumann, this paper intends to be a step forward in the use of policies, by allowing not only heuristic policies or utility functions to maximise, but also procedural policies that not only say to EERM what is the best solution, but also what to do under certain conditions.

Sulistio et al. [43] proposed overbooking strategies for mitigating the effects of cancellations and no-shows for increasing the revenue. The overbooking policies implemented in this paper, in addition, considers the possibility of under-usage of the reserved resources from the client.

Previous work from Macías and Guitart demonstrated that managing the resources according to economic objectives could lead to an increase of the economic profit. That management could be in execution time, by dynamically redistributing the allocation of resources [10, 11], or in negotiation time, by deciding the prices or scheduling the time slots to allocate the resources in function of some business objectives like Revenue Maximisation, Client Classification, or to keep the reputation [17, 16, 18]. This paper is intended to be a step forward from previous work by implementing and testing the validity of these business-oriented policies within an EERM with a rule engine [44] customizable y the system administrator, and by introducing new rules and policies, such as Selective SLA Violation or Resource Overprovisioning.

3.4 Rule Support for Management of both Business and Resources

The first part of the research on this topic is related to the definition and integration of rules within current Resource Management systems.

Herrero et al. [45] propose a rule-based Collaborative Awareness Management (CAM) environment for autonomic Grid Resource Management that manages awareness of interaction by means of a set of rules, optimising the resources collaboration, promoting the resources cooperation in the environment, and responding to the specific demanded circumstances by means of a set of pre-established rules. However, their system does not consider neither High-Level nor Business metrics as feedback information for their autonomic system.

Schiefer et al. [46] introduce a Business Rules Management System which is able to sense and evaluate events in order to respond to changes in a business environment or customer needs. Users can define the rule set and the events in which the rules will be triggered. Our work tries to extend their approach, which is mainly focused on Business Management, to the management of both Business and Resources of Cloud Computing providers, by allowing the flow of information between Business and Resources layers for their mutual optimisation.

Since one of the contributions of this thesis is also about the self-optimisation of the rules, it is important to know the related work about dynamic policies and its specification.

Several authors have been proposed and axiomatized logic rule languages for dynamic policies. Pucella and Weissman [47] introduce a logic based on Propositional Dynamic Logic (PDL) [48], that assume a set of primitive actions and provide combinations for building complex actions from the primitive ones. To establish the truth of the formulas, they store which transitions are assumed to be permitted in sets of policy transitions. Their logic, called DLP_{DYN}, compares policy sets and allows the reasoning about changing policies. Benthem and Liu [49] present a complete logic of knowledge update plus preference upgrade that can model changing obligations or conflicting commands. They model a preference logic that, instead of expressing statements such as “A knows B”, it allows the expression of “A prefers B”. Yamada [50] introduces a language for commanding updates in logic and obligations, $\mathcal{L}_{MDL^+}$, and introduces some operators, such as promises, as dynamic updaters of obligations.

Although the recently cited papers give a strong basis for modelling dynamic rules, they come from the world of philosophy. In consequence, there are not implementations of these logics in computer rule management systems. Since the creation of a system that implement logics such as PDL, DLP_{DYN}, or $\mathcal{L}_{MDL^+}$ would require a big effort that is out of the scope of this thesis, other methods for adding dynamicity to Business Rules Must be considered.

Genetic Algorithms have been proved to be a valid method for dynamically learn and self-adapt the behaviour of a system to the changes in the environment [51, 52]. Because finances are chaotic and noisy, they use Genetic Algorithms for several problems, such as auctioning [53, 54, 55], model agent behaviour for evaluating the validity of offer sets [56], or sales and pricing of options [57].

Machine Learning and Data Mining have been used successfully in the adaptive management of energy resources in Cloud Computing environments [58]. Creating models and discovering patterns in examples from the past experience, can help to tune up a system in function of the changes in the environment. Machine learning models could be inserted and adapted within rules to help modify themselves as dynamic programs.

3.5 Trust and Reputation

One of the first analyses of trust in virtual communities was introduced by Abdul-Rahman and Hailes [59]. However, their proposed model was more stuck to social communities than to Utility Computing. The proposed model is based in the subjective feedback of users, like in other systems such as the eBay reputation feedback [60] that reported a strong positive feedback to the marketplace despite its drawbacks and positive nature [24]. To compensate the inconvenience of subjectivity, some objective reputation systems have been implemented in P2P systems [61] where users can upload or download credit in function to their collaboration with the network. Hwang et al. [62] also propose an approach to integrating virtual clusters, security-reinforced datacenters, and trusted data accesses guided by reputation systems by implementing a hierarchy of P2P reputation systems to protect clouds and datacenters at the site level and to safeguard the data objects at the file-access level. These reputation models could be interesting for this thesis. However, its distributed implementation could add too much complexity to a Cloud Computing market such as the described in this thesis.

Azzedin and Maheswaran [63] studied a Trust-aware Resource Management System (TRMS) that allocates resources considering a trust relationship between the resource provider and the resource consumer. Their proposed Trust Model divides the overall Grid Domains: autonomous administrative entities consisting of a set of resources managed by a single administrative authority. Each Grid Domain is mapped to several Resource and Client Domains, whose trust relations are established in a Trust-Level Table. Their experiments show how the performance of a system can be improved up to 40% by incorporating trust in the scheduling heuristics. This thesis will take some ideas from this work, but it must simplify the general architecture for eliminating domains that can be mapped to Virtual Organizations of Grid Computing, since Cloud Computing markets usually do not implement them.

Song et al. [64] models the risk and insecure conditions in Grid job scheduling. They suggest that, in a wide-area Grid environment, it is more resilient for the global job scheduler to tolerate some job delays due to the trust mechanisms instead of resorting to pre-emption or replication, or taking a risk on unreliable resources. This thesis must extend this work to not only support batch jobs, but also Web Services.

Kim et al. [65] present a trust model for efficient reconfiguration and allocation of computing resources satisfying various user requests. Their model collects and analyzes reliability based on historical information of Cloud providers. It prepares the best available resources for each service request in advance, providing the best resources to users. They demonstrate how Cloud providers can utilize their resources efficiently and also provide highly trusted resources and services to many users. This thesis defends that this work can be more accurate by providing new prediction models [66, 67], in addition to statistical data.

Chapter 4

Working Plan

4.1 Schedule

Figure 4.1 shows the Gannt diagram of the research topics enumerated in chapter 1.

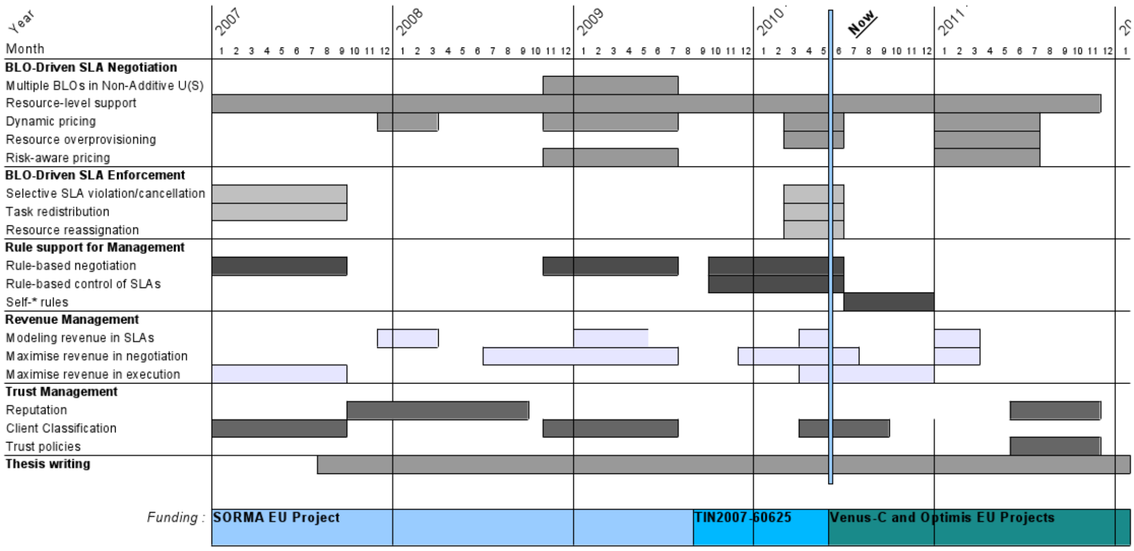

Figure 4.1: Gannt diagram of the work performed in the PhD thesis

4.2 Work performed

4.2.1 Economically Enhanced Resource Manager

The problem raised in this thesis is faced through a component Economically Enhanced Resource Manager (EERM), whose architecture is depicted in Figure 4.2. Next are described its main components:

Negotiation

This component interacts with the clients in order to perform the resource allocation and pricing that fits best within its own objectives.

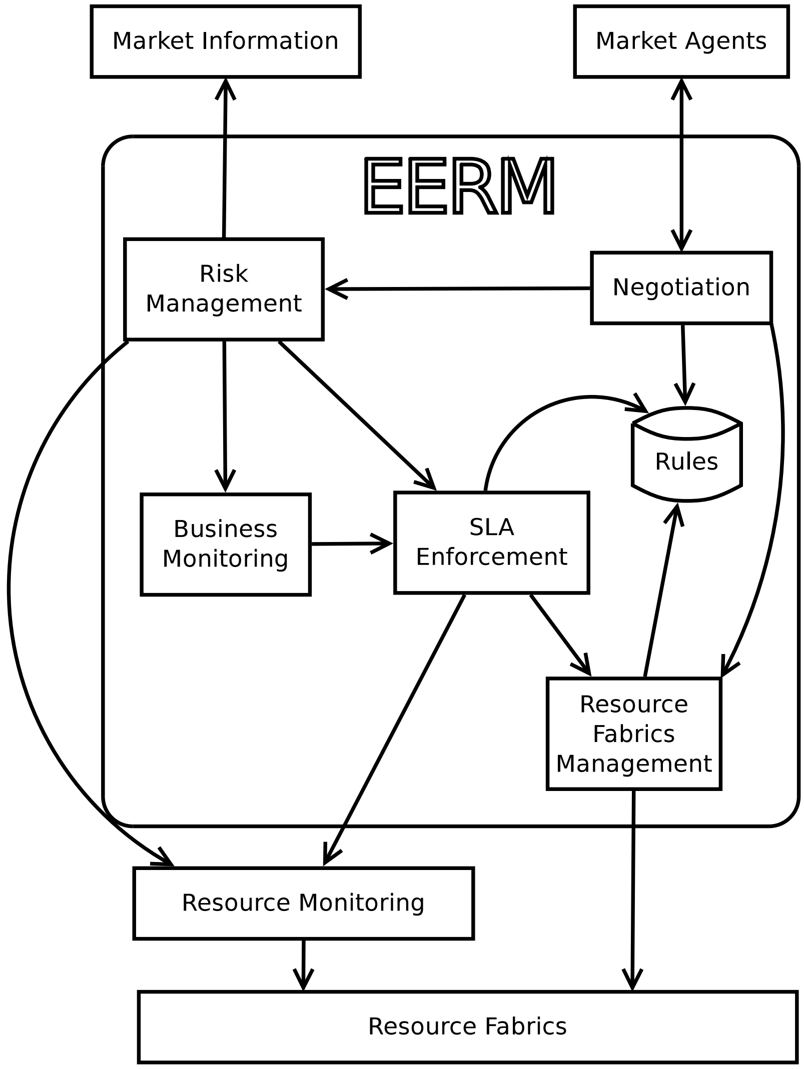

Figure 4.2: General architecture of the EERM

Risk Management
: Taking into account both market and resources information, manages the risk for the EERM of doing determined actions, such as resource allocations, negotiation confirmations, etc. This risk is taken into account when negotiating and managing resources.

SLA Enforcement
: Keeps track of tasks executed in the system, and continuously watches the status of the agreed SLAs. If SLAs are being violated, takes reactive measures in order to minimize the economic and technical impact of these violations.

Resource Fabrics Management
: Performs basic operations in the resource fabrics, such as resource creation, cancellation or migration, and tasks reallocation. Basically, it is the interface between the Resource Fabrics and the EERM.

Business Monitoring
: Keeps track about the status of the achievement of the BLOs. It can change the business rules for ensuring the maximum compliance of the established objectives.

Rules Manager
: A configurable Rules Repository that includes all the rules that describe the economic behaviour of the EERM. The implementation of the EERM uses the Drools Expert Rule Engine [44].

4.2.2 BLO-Driven SLA Negotiation

Resource-level Support

Previous work on this area defends the idea that efficient negotiations require of the usage of resource-level information for increasing the accuracy of negotiated Service Level Agreements and facilitating the achievement of both performance and business goals. A negotiation model based on the maximisation of nonadditive utility functions that considers multiple objectives is defined, and its validity is demonstrated through the experiments performed [17, 18, 16].

The benefits of applying knowledge about resources to Market negotiations are shown. In Cloud Computing Markets, this will lead to the Market Broker to perform better business decisions. It is demonstrated how the figure of an Economically Enhanced Resource Manager can benefit Providers by providing resource data to the Brokers and by considering economic policies in resource management.

Dynamic Pricing and Using Multiple BLOs within Non-additive Utility Functions

Previous contributions done in the framework of this thesis are the intention of being a step forward in the modelling and evaluation of utility functions for negotiations in Cloud Computing Markets. Simulations show how a Provider can perform complex actions by only maximising a multi-dimensional utility function. The contribution of these experiments is based in the usage of nonadditive utility functions, more difficult to treat, but needed when assuming that the terms under negotiation are not independent between them. For example the utilities for Client classification and revenue maximisation were related by the price: maximising the price would lead to maximise the revenue but to minimise the affinity of Clients.

The proposed nonadditive utility function considers the possibility of having multiple objectives in a same entity, such as revenue maximisation, Client classification, reputation or load-balancing in time. In order to keep the efficiency both in business and performance terms, most of the parameters that compose the utility function are collected dynamically from the resource-level information. It is demonstrated the high importance of having this information available in negotiation time. The simulations performed demonstrate how the objectives can be partially achieved by balancing correctly their weights in the utility function.

As an alternative approach, genetic algorithms have been used and tested their validity. A paper with the results of this topic is written and submitted [68]. A generic, inexact pricing function that evolves to a pricing function that offers suitable prices in function of the system status is shown and its results are compared with other pricing strategies, demonstrating its validity.

Resource Overprovisioning

Policies that perform overprovisioning in negotiation time have been implemented and evaluated in an EERM simulation environment. Experiments demonstrate how revenue and resources use can be multiplied in time slots where the workload is low, such as the early morning. The consideration is that some SLAs are violated by the provider, since the risk calculations and predictions have a small error rate. However, the impact of these violations can be minimized by the BLO-Driven SLA Enforcement.

4.2.3 BLO-Driven SLA Enforcement

Selective SLA violation and cancellation

Previous works describe techniques used by the EERM to support revenue maximisation across multiple Service Level Agreements and provide application scenarios to demonstrate its usefulness and effectiveness [10, 11].

In particular, previous work focuses on revenue maximisation using SLAs and describes how a strategic approach to managing SLAs can be used to secure optimal profit in situations where resources are scarce. Using methods for selective SLA fulfilment and violation, the resource provider can determine which tasks should be pre-empted in favour of freeing up resources for more lucrative SLAs. For example, it may be more profitable to violate an existing SLA, and pay the associated penalty, than it is to checkpoint and redistribute existing tasks, so that all SLAs can be fulfilled.

The performed simulations demonstrated that using economic enhancements in a resource manager can be more cost effective for service providers, especially when such providers try to fulfil the SLAs in scenarios with insufficient resources.

Task redistribution

Rules for task redistribution are implemented in the EERM, and the results show how the redistribution of tasks when the load of a system node exceeds a threshold (e.g. the 90% of its capacity) minimizes the violation of SLAs and, in consequence, maximise its benefit. A paper with the results of this topic is being written and submitted soon.

4.2.4 Rule support for Management of both Business and Resources

Drools Rule Engine [44] has been integrated within EERM. Under certain situations such as resources overload, negotiation requests or violation of SLAs, rules specified by the business administrators are triggered and make the EERM taking decisions and reactive actions under certain situations. These rules implement a big part of the research topics in this thesis. A paper with the results of this topic is being written and submitted in a near time.

Rule-based negotiation framework

Several rules have been implemented and tested for its usage in negotiation time [20]:

- Resource allocation rules that look for the most suitable time slot and, in case it is needed, perform resource overprovisioning. Results show how sales can be increased if EERM allocate batch workloads in the early morning (and, in consequence, prices are cheaper) and if it performs overprovisioning for real-time tasks that must be allocated in peak hours, such as public Web Services.

- Pricing rules that, given a concrete SLA, try to decide dynamically the best price for winning the auction in a competitive market. Experiments performed show how a dynamic pricing provider earns more money than fixed-pricing providers by offering lower prices in off-peak hours and high prices in peak hours.

- When the revenue is a secondary goal, price discrimination rules have been implemented, as commented in section 4.2.6.

Rule-based control theory for SLA Enforcement

Many rules have been implemented and tested for its usage in execution time [20]:

- As a **preventive** measure against system overload scenarios, task redistribution has been implemented, as explained in section 4.2.3.

- As a **reactive** measure under system overload scenarios, selective SLA violations and cancellations have been implemented, as described in section 4.2.3.

4.2.5 Revenue Management

Work for maximising revenue in both negotiation and execution time has been performed. Actually, revenue maximisation is the objective of other works in this thesis, such as dynamic pricing (section 4.2.2), resource overprovisioning (section 4.2.2), selective SLA violation and cancellation (section 4.2.3) and, in an indirect way, all of the policies that are intended to minimize the violations of SLAs, like task redistribution 4.2.3.

4.2.6 Trust Management

Reputation

The trustworthiness of commodity providers is a key issue to motivate customers to participate in a market. Several Grid Market frameworks are being developed and they are not an exception in the need of demonstrate their trustworthiness to clients. This thesis proposes a simple reputation mechanism for auctions within Grid Markets, which is easy to embed in old Grid solutions that do not provide reputation yet. Effectiveness of this mechanism is proved through simulation in different market scenarios, by showing the direct correlation between reputation and revenue of Grid service providers.

Previous work [25] introduced a proposal for a very simple reputation mechanism that can be easily embedded in Grid scenarios which do not provide reputation. This mechanism will help legacy Grid clients to acquire better resource providers in a transparent way, without need of modification. This affirmation is demonstrated through a simulation experiment and the analysis of its result data.

It is demonstrated that a market in equilibrium state is the optimal scenario, because almost all the entities obtain benefit from their participation in the market. It is also shown the importance of reputation when negotiating for resources and how the reputation makes providers to adjust their category and prices to their true reliability.

Client Classification

Work currently performed defines Price Discrimination methods and rules for providers that, for example, want to offer free (or cheap) resources to its affine organisations. Results demonstrate that this kind of policies attract most of internal users or users from affine organisations (e.g. from other headquarters of the same company).

4.3 Future work

Next are enumerated the future research until finish this thesis. A paper will be written and submitted for each of the next topics.

4.3.1 Business Rules and Policies for Client Classification

In comparison to rules for revenue maximisation, this research work intends to establish policies for another Business-Level Objective: to attract as much users from an specified type or organisation as possible. In addition to current work, which performs Client Classification by price discrimination, some other rules can be established, similar to the explained in sections 4.2.2 and 4.2.3, but prioritizing Client Classification instead of Revenue Maximisation. All of these rules will be experimentally evaluated to demonstrate its validity.

4.3.2 Self-adapting and Self-optimising Business Rules

Current rules and its supporting engine will be extended for allowing the dynamic evolution of the rules to allow its adaptation to the changes in the environment. New policies will be defined and evaluated, and compared to the rules already defined in this thesis for showing how Cloud Providers that dynamically adapt their rules without the human intervention obtain a better achievement of their BLOs.

4.3.3 Improved Revenue Modelling and Management

Current negotiation models will be extended by taking into account some extra economic concepts when modelling the revenue, for example:

Risk The Cloud provider must identify the risks and hazards that can incur penalties in the achievement of the BLOs. When trying to maximise the reputation of the provider, these risks must be considered when SLAs are negotiated, and impact them in the price.

Return of Investment A performance measure used to evaluate the efficiency of an investment or to compare the efficiency of a number of different investments. To calculate ROI, the benefit (return) of an investment is divided by the cost of the investment; the result is expressed as a percentage or a ratio. Prices of Cloud Computing services could be influenced by the status of the return of the investment expenses in Cloud Infrastructures.

Inventory Turnover With the exponential evolution of computing power, Cloud Computing providers must renovate their machines for being competitive against other providers. The costs of this inventory turnover must be considered, in addition to Return of Investment.

Work will consist also in evaluating some other economic variables and considering their inclusion in the modelling and management of the revenue, such as time from order to fulfillment, time-to-market, market share, frequency of delayed deliveries, average time for collection, age of accounts receivables, etc [69].

4.3.4 Deeper integration of trust within existent policies

In addition to Client Classification, some other trust policies must be integrated in the Cloud provider. From the client-side trustworthiness, the market broker and the resource manager must consider it when negotiating for a SLA and consider the potential risks when offering prices, giving access to some resources, etc.

From the provider-side trustworthiness, both brokers and resource managers must consider the consequences of their actions in the reputation of the provider, and how can affect them in the future trustworthiness. For example, being too aggressive when overprovisioning resources, or perform SLA Cancellations can have a short-term economic benefit, but a mid-term impact in the reputation.

Chapter 5

Publications

Journal Articles

- [11] Maximizing Revenue in Grid Markets using an Economically Enhanced Resource Manager M. Macías, O. Rana, G. Smith, J. Guitart, and J. Torres Concurrency and Computation: Practice and Experience, Vol. XX (XX), pp. XX-XX, September 2008 ISSN: 1532-0634 Impact Factor: 1.154 (SCI 2007)

Conference Proceedings

- [68] A genetic model for pricing in Cloud Computing markets
  M. Macías, J. Guitart
  Submitted to 7th Intl. Workshop on Economics of Grids, Clouds, Systems, and Service (Gecon 2010)
  Ischia - Naples, Italy. August 30-31, 2010. Co-located with Euro-Par 2010
  Acceptance pending

- [20] On the usage of Dynamic Policies within Resouce Management of Cloud Providers
  M. Macías, J. Guitart
  Submitted to 6th International Conference on Network and Service Management (CNSM 2010)
  Niagara Falls, Canada. October 25-29, 2010.
  Acceptance pending

- [18] Using Resource-level Information into Nonadditive Negotiation Models for Cloud Market Environments
  M. Macías and J. Guitart
  12th IEEE/IFIP Network Operations and Management Symposium (NOMS'10)
  Osaka, Japan, April 19-23, 2010, pp. 325-332
  ISBN: 978-1-4244-5367-2, ISSN: 1542-1201
  54 Accepted of 201 Submissions: 26.9%
  Conference ranking at CORE list: B

- [17] A Non-Additive Negotiation Model for Utility Computing Markets M. Macías and J. Guitart

XX Jornadas de Paralelismo (JP 2009)
A Corua, Spain, September 16-18, 2009, pp. 695-700
ISBN: 84-9749-346-8

- [25] Influence of Reputation in Revenue of Grid Service Providers
M. Macías and J. Guitart
2nd International Workshop on High Performance Grid Middleware (HiPer-GRID 2008)
Bucharest, Romania, November 21, 2008, pp. 9-16
ISSN: 2065-0701
(Also as Research Report number: UPC-DAC-RR-2008-61, October 2008)

- [10] Enforcing Service Level Agreements using an Economically Enhanced Resource Manager
M. Macías, G. Smith, O. Rana, J. Guitart, and J. Torres
1st Workshop on Economic Models and Algorithms for Grid Systems (EMAGS 2007)
In conjunction with the 8th IEEE/ACM International Conference on Grid Computing (Grid 2007)
Austin, Texas, USA, September 19, 2007
(Also as Research Report number: UPC-DAC-RR-2007-49, September 2007)

Collaborations in other publications

- [70] SLA-based Resource Management and Allocation
J. Guitart, M. Macías, O. Rana, P. Wieder, R. Yahyapour, W. Ziegler
Chapter 12 in: R. Buyya, K. Bubendorfer (Eds.), "Market-Oriented Grid and Utility Computing"
Wiley (publisher). June 2009. ISBN: 978-0-470-28768-2

- [9] Economically Enhanced Resource Management for Internet Service Utilities
T. Pueschel, N. Borissov, M. Macías, D. Neumann, J. Guitart, and J. Torres
Lecture Notes on Computer Science (LNCS), Vol. 4831, pp. 335-348
ISBN: 978-3-540-76992-7, ISSN: 0302-9743
8th International Conference on Web Information Systems Engineering (WISE'07)
Nancy, France, December 3-7, 2007
40 Accepted of 200 Submissions: 20%
Conference ranking at CORE list: A
(Also as Research Report number: UPC-DAC-RR-2007-51, September 2007)

- [71] Extended Resource Management Using Client Classification and Economic Enhancements
T. Pueschel, N. Borissov, D. Neumann, M. Macías, J. Guitart, and J. Torres
Information and Communication Technologies and the Knowledge Economy, Volume 4
Expanding the Knowledge Economy; Issues, Applications, Case Studies; Part 1

ICT for Networked Enterprise, Applications, pp. 65-72
ISBN: 978-1-58603-801-4, ISSN: 1574-1230
17th eChallenges e-2007 Conference & Exhibition (e-2007)
The Hague, The Netherlands, October 24-26, 2007
(Also as Research Report number: UPC-DAC-RR-2007-50, September 2007)

Bibliography

[1] J. Basney and M. Livny, "Deploying a high throughput computing cluster," in *High Performance Cluster Computing: Architectures and Systems, Volume 1*, R. Buyya, Ed. Prentice Hall PTR, 1999.

[2] M. A. Rappa, "The utility business model and the future of computing services," *IBM Syst. J.*, vol. 43, no. 1, pp. 32–42, 2004.

[3] D. Thain and M. Livny, "Building reliable clients and servers," in *The Grid: Blueprint for a New Computing Infrastructure*, I. Foster and C. Kesselman, Eds. Morgan Kaufmann, 2003.

[4] R. Buyya, C. S. Yeo, and S. Venugopal, "Market-oriented cloud computing: Vision, hype, and reality for delivering it services as computing utilities," *High Performance Computing and Communications, 10th IEEE International Conference on*, vol. 0, pp. 5–13, 2008.

[5] S. Conway, R. Walsh, and E. C. Joseph, "Hpc management software: reducing the complexity of hpc cluster and grid resources," Platform Computing Co., Tech. Rep., 2008.

[6] I. Goiri, "Towards virtualized service providers," Master's thesis, Technical University of Catalonia, 2008.

[7] A. V. Moorsel, "Metrics for the internet age: Quality of experience and quality of business," HP, Tech. Rep. HPL-2001-179, 2001.

[8] A. Moura, J. Sauve, and C. Bartolini, "Research challenges of business-driven it management," in *Business-Driven IT Management, 2007. BDIM '07. 2nd IEEE/IFIP International Workshop on*, 2007, pp. 19–28.

[9] T. Püschel, N. Borissov, M. Macías, D. Neumann, J. Guitart, and J. Torres, "Economically enhanced resource management for internet service utilities." in *WISE*, ser. Lecture Notes in Computer Science, B. Benatallah, F. Casati, D. Georgakopoulos, C. Bartolini, W. Sadiq, and C. Godart, Eds., vol. 4831. Springer, 2007, pp. 335–348. [Online]. Available: http://dblp.uni-trier.de/db/conf/wise/wise2007.html

[10] M. Macías, G. Smith, O. Rana, J. Guitart, and J. Torres, "Enforcing service level agreements using an economically enhanced resource manager," in *Proceedings of hte 1st Workshop on Economic Models and Algorithms for Grid Systems (EMAGS 2007)*, Austin, Texas, USA, September 2007.

[11] M. Macías, O. Rana, G. Smith, J. Guitart, and J. Torres, "Maximizing revenue in Grid markets using an Economically Enhanced Resource Manager," *Concurrency and Computation: Practice and Experience*, vol. 9999, no. 9999, p. n/a, September 2008. [Online]. Available: http://dx.doi.org/10.1002/cpe.1370

[12] H. Raiffa, *The art and science of negotiation.* Cambridge, Mass: Belknap Press of Harvard University Press, 1982.

[13] P. Faratin, C. Sierra, and N. R. Jennings, "Negotiation decision functions for autonomous agents," *International Journal of Robotics and Autonomous Systems*, vol. 24, pp. 3–4, 1998.

[14] D. F. Ferguson, "The application of microeconomics to the design of resource allocation and control algorithms," Ph.D. dissertation, New York, NY, USA, 1989.

[15] ISO 31000 2009 Risk management. Principles and guidelines (last visited: April 2010). [Online]. Available: $http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=43170$

[16] M. Macías and J. Guitart, "On the use of resource-level information for enhancing sla negotiation in market-based utility computing environments," Master's thesis, Technical University of Catalonia, 2009.

[17] M. Macías and J. Guitart, "A non-additive negotiation model for utility computing markets," in *Proceedings of Jornadas del Paralelismo 2009*, A Coruña, Spain, September 2009.

[18] M. Macías and J. Guitart, "Using resource-level information into nonadditive negotiation models for cloud market environments," in *12th IEEE/IFIP Network Operations and Management Symposium (NOMS'10)*, Osaka, Japan, April 2010, pp. 325–332.

[19] I. Goiri, F. Julia, and J. Guitart, "Efficient data management support for virtualized service providers," in *17th Euromicro Conference on Parallel, Distributed and Network-based Processing (PDP'09)*, Weimar, Germany, February 2009, pp. 409–413.

[20] M. Macías and J. Guitart, "On the usage of dynamic policies within resouce management of cloud providers," in *Submitted to 6th International Conference on Network and Service Management (CNSM 2010)*, Niagara Falls, Canada, October 2010.

[21] G. Forgacs, *Revenue Management. Maximizing Revenue in Hospitality Operations.* American Hotel and Lodging Educational Institute, 2010.

[22] Web Services Agreement specification (http://www.ogf.org/documents/GFD.107.pdf). [Online]. Available: http://www.ogf.org/documents/GFD.107.pdf

[23] C. S. Yeo and R. Buyya, "Service Level Agreement based Allocation of Cluster Resources: Handling Penalty to Enhance Utility," in *Proceedings of the 2005 IEEE International Conference on Cluster Computing*, Boston, Massachusetts, USA, September 26–30, 2005, pp. 1–10.

[24] A. Josang, R. Ismail, and C. Boyd, "A Survey of Trust and Reputation Systems for Online Service Provision," in *Decision Support Systems*, Elsevier Science Publishers B. V., Ed., vol. 43, no. 2, Amsterdam, The Netherlands, 2007, pp. 618–644.

[25] M. Macías and J. Guitart, "Influence of reputation in revenue of grid service providers," in *Proceedings of the 2nd International Workshop on High Performance Grid Middleware (HiPerGRID 2008)*, Bucharest, Romania, November 2008.

[26] Fingrid - financial business grid. [Online]. Available: http://www.fingrid.de

[27] J. Altmann, C. Courcoubetis, G. D. Stamoulis, M. Dramitinos, T. Rayna, M. Risch, and C. Bannink, "Gridecon: A market place for computing resources," in *GECON*, 2008, pp. 185–196.

[28] Self-organizing ICT Resource Management (SORMA). [Online]. Available: http://www.sorma-project.eu

[29] Information Society Technologies Programme. [Online]. Available: http://www.cordis.lu/ist

[30] Grid 4 All. [Online]. Available: http://www.grid4all.eu

[31] F. A. Hayek, W. Bartley, P. Klein, and B. Caldwell, *The collected works of F. A. Hayek*. University of Chicago Press, 1989.

[32] R. Buyya, "Economic-based distributed resource management and scheduling for grid computing," *CoRR*, vol. cs.DC/0204048, 2002.

[33] L. Zadeh, K. Fu, K. Tanaka, and M. Shimura, *Fuzzy sets and their applications to Cognitive and Decision Processes*. Academic Press, New York, 1975.

[34] D. Ouelhadj, J. Garibaldi, J. MacLaren, R. Sakellariou, and K. Krishnakumar, "A multi-agent infrastructure and a service level agreement negotiation protocol for robust scheduling in grid computing," in *Proceedings of the European Grid Conference*, ser. Lecture Notes in Computer Science. Springer-Verlag, 2005.

[35] R. G. Smith, "The contract net protocol: High-level communication and control in a distributed problem solver," *Transactions on Computers*, vol. C-29, no. 12, pp. 1104–1113, 1980.

[36] N. Vulkan and N. R. Jennings, "Efficient mechanisms for the supply of services in multi-agent environments," in *ICE '98: Proceedings of the first international conference on Information and computation economies*. New York, NY, USA: ACM, 1998, pp. 1–10.

[37] C. S. Yeo and R. Buyya, "A taxonomy of market-based resource management systems for utility-driven cluster computing," *Softw. Pract. Exper.*, vol. 36, no. 13, pp. 1381–1419, 2006.

[38] P. Dube, Y. Hayel, and L. Wynter, "Yield management for it resources on demand: analysis and validation of a new paradigm for managing computing centres," *Journal of Revenue and Pricing Management*, vol. 4:1, pp. 24–38, 2005.

[39] M. J. Freedman, C. Aperjis, and R. Johari, "Prices are right: Managing resources and incentives in peer-assisted content distribution," in *Proc. 7th International Workshop on Peer-to-Peer Systems (IPTPS08)*, Tampa Bay, FL, Feb. 2008.

[40] D. A. Menascé, V. A. F. Almeida, R. Fonseca, and M. A. Mendes, "Business-oriented resource management policies for e-commerce servers," *Perform. Eval.*, vol. 42, no. 2-3, pp. 223–239, 2000.

[41] N. Poggi, T. Moreno, J. Berral, R. Gavalda, and J. Torres, "Self-adaptive utility-based web session management," *Computing Networks*, vol. 53:10, pp. 1712–1721, 2009.

[42] T. Pueschel and D. Neumann, "Management of cloud infrastructures: Policy-based revenue optimization," in *International Conference on Information Systems (ICIS 2009)*, Phoenix, Arizona, December 2009. [Online]. Available: http://aisel.aisnet.org/icis2009/178

[43] A. Sulistio, K. H. Kim, and R. Buyya, "Managing cancellations and no-shows of reservations with overbooking to increase resource revenue," *Cluster Computing and the Grid, IEEE International Symposium on*, vol. 0, pp. 267–276, 2008.

[44] "Drools expert." [Online]. Available: http://www.jboss.org/drools/drools-expert.html

[45] P. Herrero, J. L. Bosque, M. Salvadores, and M. S. Perez, "A rule based resources management for collaborative grid environments," *Int. J. Internet Protoc. Technol.*, vol. 3, no. 1, pp. 35–45, 2008.

[46] J. Schiefer, S. Rozsnyai, C. Rauscher, and G. Saurer, "Event-driven rules for sensing and responding to business situations," in *DEBS '07: Proceedings of the 2007 inaugural international conference on Distributed event-based systems.* New York, NY, USA: ACM, 2007, pp. 198–205.

[47] R. Pucella and V. Weissman, "Reasoning about dynamic policies," in *Proceedings of the Sixth International Conference on Foundations of Software Science and Computation Structure (FOSSACS-04)*, 2004, pp. 453–467.

[48] D. Harel, D. Kozen, and J. Tiuryn, *Dynamic Logic.* The MIT Press, 2000.

[49] J. V. Benthem and F. Liu, "Dynamic logic of preference upgrade," *Journal of Applied Non-Classical Logics*, vol. 17, 2005.

[50] T. Yamada, "Acts of commanding and changing obligations," in *CLIMA VII*, ser. Lecture Notes in Computer Science, K. Inoue, K. Satoh, and F. Toni, Eds., vol. 4371. Springer, 2006, pp. 1–19.

[51] S.-H. Chen, Ed., *Genetic Algorithms and Genetic Programming in Computational Finance.* Norwell, MA, USA: Kluwer Academic Publishers, 2002.

[52] S.-H. Cheng, Ed., *Evolutionary Computation in Economics and Finance*, ser. Studies in Fuzziness and Soft Computing. Springer-Verlag, 2002, vol. 100, no. XII.

[53] D. Cliff, "Evolution of market mechanism through a continuous space of auction-types," *Computational Intelligence, Proceedings of the World on Congress on*, vol. 2, pp. 2029–2034, 2002.

[54] D. Cliff, "Evolution of market mechanism through a continuous space of auction-types ii: Two-sided auction mechanisms evolve in response to market shocks," in *International Conference on Internet Computing IC02*, H. Arabnia and Y. Mun, Eds., vol. III. CSREA Press, 2002, pp. 682–688. [Online]. Available: http://eprints.ecs.soton.ac.uk/12133/

[55] D. Cliff, "Evolution of market mechanism through a continuous space of auction-types iii: Multiple market shocks give convergence toward cda," Hewlett-Packard Laboratories, Tech. Rep. HPL-2002-312, 2002.

[56] M. B. E. Fayek, I. A. Talkhan, and K. S. El-Masry, "Gama (genetic algorithm driven multi-agents)for e-commerce integrative negotiation," in *GECCO '09: Proceedings of the 11th Annual conference on Genetic and evolutionary computation.* New York, NY, USA: ACM, 2009, pp. 1845–1846.

[57] N. K. Chidambaran, C.-W. J. Lee, and J. R. Trigueros, "An adaptive evolutionary approach to option pricing via genetic programming," New York University, Leonard N. Stern School of Business-, New York University, Leonard N. Stern School Finance Department Working Paper Seires, 1998. [Online]. Available: http://econpapers.repec.org/RePEc:fth:nystfi:98-086

[58] J. Berral, I. Goiri, R. Nou, F. Julia, J. Guitart, R. Gavalda, and J. Torres, "Towards energy-aware scheduling in data centers using machine learning," in *1st International Conference on Energy-Efficient Computing and Networking (e-Energy'10)*, University of Passau, Germany, April 2010, pp. 215–224.

[59] A. Abdul-Rahman and S. Hailes, "Supporting trust in virtual communities," in *HICSS '00: Proceedings of the 33rd Hawaii International Conference on System Sciences-Volume 6.* Washington, DC, USA: IEEE Computer Society, 2000, p. 6007.

[60] "eBay." [Online]. Available: http://www.ebay.com/

[61] M. Gupta, P. Judge, and M. Ammar, "A reputation system for peer-to-peer networks," in *NOSSDAV '03: Proceedings of the 13th international workshop on Network and operating systems support for digital audio and video.* New York, NY, USA: ACM, 2003, pp. 144–152.

[62] K. Hwang, S. Kulkareni, and Y. Hu, "Cloud security with virtualized defense and reputation-based trust mangement," *Dependable, Autonomic and Secure Computing, IEEE International Symposium on*, vol. 0, pp. 717–722, 2009.

[63] F. Azzedin and M. Maheswaran, "Integrating trust into grid resource management systems," in *ICPP '02: Proceedings of the 2002 International Conference on Parallel Processing*. Washington, DC, USA: IEEE Computer Society, 2002, p. 47.

[64] S. Song, K. Hwang, and Y.-K. Kwok, "Risk-resilient heuristics and genetic algorithms for security-assured grid job scheduling," *IEEE Trans. Comput.*, vol. 55, no. 6, pp. 703–719, 2006.

[65] D. Slezak, T. hoon Kim, S. S. Yau, O. Gervasi, and B.-H. Kang, "A trust evaluation model for cloud computing," in *Grid and Distributed Computing*, ser. Communications in Computer and Information Science, vol. 63. Springer Berlin Heidelberg, 2009, pp. 184–192.

[66] G. Reig and J. Guitart, "Prediction of job resource requirements for deadline schedulers in virtualized environments," Master's thesis, Technical University of Catalonia, 2009.

[67] G. Reig, J. Alonso, and J. Guitart, "Deadline constrained prediction of job resource requirements to manage high-level slas for saas cloud providers," Technical Univerity of Catalonia, Tech. Rep. UPC-DAC-RR-2010-9, April 2010.

[68] M. Macías and J. Guitart, "A genetic model for pricing in cloud computing markets," in *Submitted to 7th Intl. Workshop on Economics of Grids, Clouds, Systems, and Service (Gecon 2010)*, Ischia - Naples, Italy, August 2010.

[69] J. N. Morgan, "A roadmap of financial measures for it project roi," *IT Professional*, vol. 7, pp. 52–57, 2005.

[70] J. Guitart, M. Macías, O. Rana, P. Wieder, R. Yahyapour, and W. Ziegler, *Market-Oriented Grid and Utility Computing*. Wiley, 2009, no. 12, ch. SLA-based Resource Management and Allocation.

[71] T. Püschel, N. borissov borissov, D. Neumann, M. Macías, J. Guitart, and J. Torres, "Extended resource management using client classification and economic enhancements," in *Proceedings of eChallenges e-2007 Conference*, The Hague, Netherlands, October 2007.