VIVIANE TAVARES NASCIMENTO

ENERGY MANAGEMENT FOR CLOUD COMPUTING ENVIRONMENT
GERENCIAMENTO DE ENERGIA PARA AMBIENTE DE COMPUTAÇÃO EM NUVEM

São Paulo

2017

VIVIANE TAVARES NASCIMENTO

ENERGY MANAGEMENT FOR CLOUD COMPUTING ENVIRONMENT
GERENCIAMENTO DE ENERGIA PARA AMBIENTE DE COMPUTAÇÃO EM NUVEM

Dissertação apresentada à Escola Politécnica da Universidade de São Paulo para obtenção do título de mestre em Ciências

São Paulo

2017

VIVIANE TAVARES NASCIMENTO

ENERGY MANAGEMENT FOR CLOUD COMPUTING ENVIRONMENT
GERENCIAMENTO DE ENERGIA PARA AMBIENTE DE COMPUTAÇÃO EM NUVEM

Dissertação apresentada à Escola Politécnica da Universidade de São Paulo para obtenção do título de mestre em Ciências

Área de concentração:
Engenharia de Computação

Orientadora:
Tereza Cristina de Melo Britto Carvalho

Co-orientador:
André Luiz Veiga Gimenes

São Paulo

2017

Este exemplar foi revisado e corrigido em relação à versão original, sob responsabilidade única do autor e com a anuência de seu orientador.

São Paulo, _____ de _____ de _____

Assinatura do autor: _____

Assinatura do orientador: _____

Catalogação-na-publicação

Nascimento, Viviane
Energy Management for Cloud Computing Environment / V. Nascimento - versão corr. -- São Paulo, 2017.
121 p.

Dissertação (Mestrado) - Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia de Computação e Sistemas Digitais.

1.Computação em Nuvem (Gerenciamento) 2.Energia (Gestão/Gerenciamento) 3.Cloud I.Universidade de São Paulo. Escola Politécnica. Departamento de Engenharia de Computação e Sistemas Digitais II.t.

DEDICATORIA

Para Sheila Tavares Nascimento, minha irmã,
amiga e inspiração.

AGRADECIMENTOS

Agradeço aos meus pais, Lucia Tavares Nascimento e José Carlos do Nascimento, pelo apoio nessa jornada. O carinho de vocês fez tudo ser mais especial.

Agradeço também à minha irmã, Sheila Tavares Nascimento, companheira musical de uma vida.

Obrigada à minha orientadora, Profa. Dra. Tereza Cristina M. B. Carvalho, pela oportunidade de realizar esse trabalho. Especialmente, obrigada por ter aberto a porta dessa nova vida.

Obrigada também à Catalin Meirosu (Ericsson Research - Suécia) por toda a ajuda no desenvolvimento desse trabalho.

Também agradeço ao co-orientador, Prof. Dr. André Luiz V. Gimenes, por ter me ajudado com tantas "caixinhas"no decorrer desse trabalho.

Gostaria de agradecer a todos do LASSU-PCS-EPUSP (Laborat´ orio de Sustentabilidade do Departamento de Engenharia de Computação e Sistemas Digitais da Escola Politécnica da USP) por todo o aprendizado e todas as experiências compartilhadas nesses anos.

Um obrigada mais que especial à Felipe Madureira Carneiro, Gilberto dos Santos Meneses, Karla de Lima Mota, Renata Dias dos Anjos: agradeço o apoio (mesmo que distante) em cada etapa desse projeto!

Um obrigada e muitos abraços para as amigas que ganhei nessa etapa. Vivian Fernandes Marinho Ferreira e Raquel Machado de Sousa: minhas Patricias, vocês moram no meu coração! Camila Bonassa, obrigada por sua amizade e carinho!

ACKNOWLEDGMENTS

Thanks to my parents, Lucia Tavares Nascimento and Jose Carlos do Nascimento, for the support through this journey. All your made everything more special.

Thanks to my sister, Sheila Tavares Nascimento, my whole life musical companion.

Thanks to my adviser, Prof. Dr. Tereza Cristina M. B. Carvalho, for the opportunity to this work and to open up the door to a whole new life.

Also, I would like to thank Catalin Meirosu (Ericsson Research - Sweden) for all his help in this work development.

Thanks to my co-advisor, Prof. Dr. Andre Luiz V. Gimenes, for helping me with uncountable "little boxes"during this research.

I would like to thank my colleagues of to everyone of Lassu-PCS-EPUSP (Laboratory of Sustainability of Computer Engineering and Digital Systems Department of the Polytechnic School of USP) for all the learning and all the experiences shared through this years.

My special gratitude to Felipe Madureira Carneiro, Gilberto dos Santos Meneses, Karla de Lima Mota, Renata Dias dos Anjos: thanks for your support (even with the distance) in each stage of this research!

Thanks and many hugs to my gained friends: Vivian Fernandes Marinho Ferreira and Raquel Machado de Sousa: my "Patricias", you live in my heart! Camila Bonassa, thanks for your friendship and all the caring!

Sounds of laughter, shades of love are ringing
through my opened ears
Inciting and inviting me
Limitless undying love, which shines around
me like a million suns
And calls me on and on across the universe

(*Across the Universe*- The Beatles)

RESUMO

Como um dos maiores consumidores de energia do mundo, o setor de Tecnologia da Informação e Comunicação (TIC) busca por maneiras eficientes para lidar com o consumo de energia da infraestrutura. Uma das áreas que tende a crescer nos próximos anos, os provedores de serviço de Computação em Nuvem procuram por abordagens para mudar o padrão de consumo de energia, ao mesmo tempo reduzindo custos operacionais. A estratégia mais comum para lidar com o consumo de energia é relacionada à sua eficiência. No entanto, há a oportunidade para incentivar um novo padrão de demanda por serviços de Computação em Nuvem, baseado na variação do fornecimento e preços da energia. Uma solução que considera a flutuação da energia para negociar a alocação é proposta. Termos de serviços contratáveis referentes a energizar os serviços são estabelecidos para permitir a solução de gerenciamento proposta. Também, uma nova camada de serviço capaz de lidar com requisitos da energia é definida como um elemento do ambiente de Computação em Nuvem. A literatura existente não lida com os diferentes termos do fornecimento da energia e com o gerenciamento de contratos simultaneamente. O método proposto inclui descrição dos termos de serviço, a definição da camada de serviço relacionada à energia e uma metodologia de implementação. Um modelo foi construído para validar a proposta através de um Caso de Uso que simula uma quantidade de Data Centers (DCs) espalhados pela região metropolitana de São Paulo. Os resultados obtidos mostram a capacidade de gerenciar a alocação dos serviços buscando o melhor aproveitamento da energia auto-gerada pelo ambiente. Utilizando do critério de variação dos custos de alocação, tanto para o usuário quanto para o provedor de serviços, o método negocia a alocação mais favorável para os contratos em razão da variação do fornecimento de energia.

Palavras-chave: Computação em Nuvem;Gerenciamento;Gestão da Energia;Níveis de Serviço.

ABSTRACT

As one of the major energy consumers in the world, the Information and Communication Technology (ICT) sector searches for efficient ways to cope with the energy expenditure of the infrastructure. One of the areas that tend to grow in the coming years, the Cloud Computing services providers look for approaches to change the energy expenditure pattern, concurrently reducing the operational costs. The most common strategy to cope with the energy consumption is related to its efficiency. However, there is the opportunity to encourage a new demand standard, based on the energy supply and price variation. A management approach that takes into account the fluctuation of the energy to negotiate the contracts allocation is proposed. Contractible service terms regarding powering the services are established to enable the proposed management approach. Also, a new service layer able to deal with energy requirements is defined as an element of the Cloud Computing environment. Existing literature does not cope with the different terms of the energy supply and does not apply a management of the contracts simultaneously. The proposed method includes a service terms description, the energy-related service layer definition, and a framework for its implementation. A model designed to validate the approach applies a Use Case that simulates Data Centers (DCs) spread through the metropolitan area of São Paulo. The obtained results show the ability of the model to manage the contracts allocation in accordance to the best exploitation of the self-generated energy. Taking into account the assignment costs range, to both user and services provider, the method negotiates the most affordable contracts assignment regarding the energy supply variation.

Keywords: Cloud Computing;Management;Energy Management;Service Levels

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

CPP Critical Peak Pricing

CPU Central Processing Unit

CORD Central Office Re-architected as a Datacenter

CMP Capacity Market Programs

DCs Data Center

DLC Direct Load Control

DSM Demand-Side Management

DVFS Dynamic Voltage and Frequency Scaling

DR Demand Response

E2C Energy Efficient Cloud

EaaS Energy as a Service

GLB Geographical Load Balancing

IaaS Infrastructure as a Service

ICT Information and Communication Technology

IRP Integrated Resources Planning

IT Information Technology

PaaS Platform as a Service

QoS Quality of Service

RTP Real Time Pricing

SLA Service Level Agreement

SLS Service Level Specification

SLO Service Level Objective

ToU Time-of-Use

VM Virtual Machine

CONTENTS

1 INTRODUCTION

Information and Communication Technology (ICT) sector is one of the biggest power consumers in the world. The power consumed by the industry enables to place it among the major consuming countries of the world (COOK et al., 2014). Studies estimate that the sector consumed almost 7.4% of all the generated energy in 2012, and this amount is expected to reach up to 12% until 2017 (POMERANTZ G. COOK, 2015). The fast pace of the energy consumption by the sector implies in high operational costs (WITHNEY; DELFORGE, 2014). Only in the United States, Data Centers (DCs) electrical consumption costs achieved $ 13 billion in 2013 (WITHNEY; DELFORGE, 2014), which foster owners to pursue a more efficient usage of the infrastructure (POMERANTZ G. COOK, 2015).

As the Cloud Computing environment usage tends to grow, following the demand for digital content, big data, e-commerce, and Internet traffic (WITHNEY; DELFORGE, 2014), the energy expenditure of this sector tends to increase as well. Although the Cloud Computing providers typically run the infrastructure in an efficient way (WITHNEY; DELFORGE, 2014), energy consumption reduction is a primary concern to the sector (ARROBA; BUYYA et al., 2015).

The relative flexibility of the energy demand is yet potential points to the effective functioning of the infrastructure (BERL et al., 2013). Dealing with the energy demand sets conditions to manage the infrastructure based on capacity-powering strategies. The power-aware management of the infrastructure (SCHRÖDER; NEBEL, 2013; NARAYAN; RAO, 2014; BUNSE; KLINGERT; SCHULZE, 2012), deployed by Cloud Computing services provider, sets a quality negotiation regarding energy provisioning.

Demand-Side Management (DSM), first defined as utility energy usage coordination (COOKE, 2011), incentives its consumers to alter the consumption behavior, based on energy sector supply information, environmental and infrastructural concerns (TORRITI, 2015). Deploying the DSM concept from the Cloud Computing services provider perspective, the relative computing flexibility presents an opportunity to manage the contracts based on the energy fluctuation during established time periods.

This work develops an energy management approach, focusing on the allocation of contracts on a Cloud Computing environment based on the energy deployment and purchasing. The deployment of dynamic pricing strategies, weighing the attributed contracted terms, allows developing the EaaS (Energy as a Service), an infrastructure layer to manage the powering of the capacity as a separated and contractible resource of the Cloud Computing environment.

1.1 OBJECTIVES

This section presents the main and specific objectives of this work.

1.1.1 General Objectives

The main objective of this work is to develop an energy management approach for Cloud Computing environment that enables to combine the computing infrastructure usage, energy prices and availability to set negotiation terms in the energy capacity allocation contracts.

1.1.2 Specific Objectives

The specific objectives of this work are detailed in the further itemized terms.

(a) Establish a demand management strategy, for the Cloud Computing services provider, by setting contractible energy service conditions.

(b) Propose a scheduling method that enables to manage the energy deployment and search for the most favorable contracts assignment, from the service provider perspective.

(c) Set energy service-level terms that allow the service provider to negotiate the demand assignment.

(d) Include a new service layer to the Cloud Computing model to manage the information concerning the energy data.

1.2 RESEARCH METHOD

The method adopted is the development of a new workflow to solve a known problem, presenting a comparative evaluation with different existing techniques (WAZLAWICK, 2014). The deductive approach required the study of previously stated works to define the research gap and develop a new method to solve a specified problem. The following stages were set to achieve the established objectives:

- **Literature Review:** recently published surveys and articles related to the subject were applied as comparison milestones for existing solutions. A comparison among the different energy deployment approaches for the ICT infrastructure is employed to define the strategy to be used by the proposed management method.

- **Establish the research goal:** the established techniques and concepts developed to manage the ICT infrastructure were used to handle with a broader context of the renewable sources deployment and energy market integration.

- **Evaluate a method:** develop an energy management plan that must be able to negotiate contractible energy service requirements and set strategies on how to power the infrastructure and cope with electricity fluctuation during stated time periods. The energy management deals with the energy as an infrastructure resource apart from the computing resources of the Cloud Computing, allowing stating negotiable terms with the electricity sector and DCs infrastructure.

- **Design a proof model:** it was designed a test model that enables to verify the energy management for a set of infrastructure resources and electrical energy prices and availability variation. The model must manage the terms of the contracts during established time periods from the electrical energy provisioning perspective.

- **Writting stages:** develop articles and thesis that encompass the stages of the energy management approach development.

1.3 WORK ORGANIZATION

This work is organized into six different sections, including this first one. This first section describes the motivation and objectives of the present work and the deployed research method. Section 2 presents state of the art for the Cloud Computing services and the resources management for this environment. Also in section 2, the Demand-Side Management concept is shown. Demand Response programs are defined, from the energy utility approach, and its concept as an approach to incentive energy usage pattern changes for the Cloud Computing and DCs infrastructure is presented.

The research gap identified through the study of previously presented methods to cope with the energy management of the physical infrastructure of a Cloud Computing environment is primarily described in section 3. The motivation of this research development, followed by its functional requirements are also detailed on the referred section.

Section 4 shows the proposed energy management approach. The concept of electrical energy as a provisioned resource of the Cloud Computing environment and the development of service terms for coping with the demanded resources are described. The section presents services contracting stages, setting conditions for service management and the developed method for contracts assignment. The workflow developed for the implementation of the

energy-management details how the information is received and managed along the allocation process. The section ends with the presentation of the designed model proposed by this work.

Section 5 presents the Use Case developed to show the management proposed for a near real-life situation. Specifically for the São Paulo Metro Area, a numerous DCs scenario is modeled to validate the method proposed. The test cases addressed to validate the energy management approach, and obtained results are discussed in this section. The section presents the graphs for the test case 1 of each scenario proposed.

Section 6 concludes the work. The section details the considerations of this research, from the problem, described and the motivation of this work. The contributions of this research, considering the energy management gains to the service provider and the statement of energy-service levels, are also detailed in this final section. Finally, the future steps of this work and further contributions conclude the section.

Appendix A contains the graphs for test cases 2 to 6, for the self-sufficient scenario and Appendix B contains the graphs for the second scenario proposed, from test case 2 to 6.

2 BACKGROUND

Initially, the chapter defines Cloud Computing and its contractual agreements. Different works that include energy-related service levels and resources management are referenced through this section. Also, the section refers to works that deal with the energy management and set energy strategies for the Cloud Computing model.

The following section brings the concept of Demand Management, from the electric utilities and regulatory sector definitions, and related works to DC energy management. Different authors evaluate DSM programs as a solution for energy costs reduction in ICT sector, apart from energy conservation strategies that focus on the energy consumption reduction.

2.1 CLOUD COMPUTING

Cloud Computing is a computing model that provides dynamic and on-demand access to computing resources as services that can be provisioned and discharged with minimal management effort or interaction with the services provider (MELL; GRANCE, 2011; FURHT; ESCALANTE, 2010). Services provisioning is a key feature for Cloud Computing, considering that their supply demands minimal interaction between provider and user, access via a broadband network and standard mechanisms. Also, capabilities are provisioned and released automatically, the services must be measured, according to its type, and the user pays according to the consumed resources (MELL; GRANCE, 2011).

As Cloud Computing resources management, Jennings and Stadler (JENNINGS; STADLER, 2015) mentions it as the process to allocate storage, allocating, computing and network according to demanded services. The authors include the energy provisioning as a manageable infrastructure of the Cloud Computing environment, related to its effective deployment. Despite the absence of standard contract composition, the interested parts agree on service levels and performance metrics (JENNINGS; STADLER, 2015; LONGO; ZAPPATORE; BOCHICCHIO, 2015).

Service Level Agreement (SLA) is the contract claim that enables to guarantee Quality of Service (QoS), service guarantees and compliance, and sets Service Level Objectives (SLO), and sustainability terms required by the user (FURHT; ESCALANTE, 2010; LONGO; ZAPPATORE; BOCHICCHIO, 2015). SLA establishes metrics and standards to measure the service performance and tolerance to services faults (LONGO; ZAPPATORE; BOCHICCHIO, 2015); SLO measures the performance of the service provider, to assure that the SLA agreed is fulfilled (ZHANG et al., 2014).

SLA establishes how to supply services in unambiguous terms, specifying their quality, time

and response requirements (HILES, 2002). The prediction of resources utilization and deadlines are deployed to negotiate the contract terms. SLA relies on business directives, and it can be negotiable regarding the contractors' needs and timescale, which is variable and can be different for determined time intervals. The agreements must be bilateral, and the constant negotiation of these terms is applied to consider costs and usage deployment (HILES, 2002).

For a DC functioning, the energy consumption pattern cannot be modified without taking into account performance conditions, such as availability and security (BUNSE; KLINGERT; SCHULZE, 2012); establishing SLA energy terms enables to define quality requirements related to the energy consumption on the Cloud Computing environment. The contracted levels can determine the deployment of aggressive energy efficiency techniques, the time intervals for the services allocation and even the acceptance of renewable energy sources. The GreenSLA is a SLA-type that enables to negotiate energy efficiency parameters by dealing with the QoS loosening; the possible relaxation of the services levels arrangement in exchange for incentives to the users (BUNSE; KLINGERT; SCHULZE, 2012).

Although the authors define the SLA loosening to energy efficiency strategies, they enclose its flexibility to cope with the renewables intermittent behavior, such as wind and solar sources. The consumption of renewable energy sources is established as an implementation of the energy-related SLA, although not limited to it. The use of renewable energy source is described in (HAQUE et al., 2013) as a deployment of GreenSLA for the cloud-computing environment. The work cites how to distribute and control the usage of renewable sources, and a framework for the workload allocation without compromising the contracted GreenSLAs.

Hasan et al. (HASAN et al., 2015) evaluated an SLA supporting language to promote GreenSLA parameters. The work presents the virtualization of renewable energy, named as green energy, based on energy availability during specified time intervals. The work establishes the virtualization of the provisioned renewable energy sources considering the surplus and sources non-availability. A GreenSLA strategy sets the balance between energy provisioning and demanded services. The monitoring of the resources available in different time slots defines a scheduling strategy for the workload allocation. The GreenSLA guarantees that the contracted SLA and the services are going to be provided physically by the Cloud provider.

Bunse et al. (BUNSE; KLINGERT; SCHULZE, 2012) proposed work to support GreenSLA through contracts. The authors propose the deployment of energy efficiency metrics to negotiate the contracted terms between services providers and final users. The work puts shows how to define the energy efficiency metrics that must be hired in three steps: measurement, infrastructure monitoring, and resource scheduling.

Initially, the environmental measurement of service usage defines parameters as optimization tools, and the build the schedule approach based on metrics criteria. The second step deploys

the energy efficiency metrics to monitor and optimize the system, through behavioral aspects, and the third point refers to the usage of the parameters to schedule the DCs resources provisioning.

The authors justify this definition based on the fact that provided services nature is not altered and that additional metrics must compare the environmental impact of the new contract. Both works cite how the SLA can be applied to provide services guaranteeing an efficient use of energy in the physical infrastructure. However, they evaluate the concepts of energy usage as a negotiable term; both works do not detail the management of contracts terms as decision requirements.

2.2 SCHEDULING THE CONTRACTS

Resources are managed according to established performance objectives that enable to deal more efficiently and effectively with the Cloud Computing resources and maintain the agreed service level with the users (JENNINGS; STADLER, 2015). Resources management is defined as the process of resource allocation and fulfilling the performance requirements specified by providers and users, considering scheduling of cloud resources, demand profiling, resource estimation, pricing and profit maximization, scaling and provisioning, workload administration and management systems (JENNINGS; STADLER, 2015).

Cited approaches that enable the management of the physical infrastructure of the Cloud Computing environment are resource provisioning, resource allocation, resource adaptation, resource mapping, resource modeling, resource estimation, resource discovery and selection, resource brokering, and resource scheduling (MANVI; SHYAM, 2014). Manvi and Shiam (MANVI; SHYAM, 2014) define resources scheduling as a management of the resources considering time variation of events and resources, setting the services functioning depending on duration, forerunner activities and resource allocation.

The electrical energy deployment is defined as a resource of the Cloud Computing environment, and its consumption depends on the management of physical resources (MANVI; SHYAM, 2014), (JENNINGS; STADLER, 2015). Especially regarding efficient use of the electrical energy, its deployment controls the physical resources. Taking into account the timing variability of the power, including generation and price fluctuation, the scheduling of the services regarding the energy supply is one strategy to manage the resources.

The power-aware services schedule proposes the assignment of contracts according to power profiles, electricity generation rates, and price range. Therefore, scheduling the services from an electricity deployment perspective enables to establish both energy efficiency scheduling strategies and energy peak rates and generation resources assignment.

Therefore, works cite the implementation of power-aware scheduling techniques to take advantage of the prices, focusing on electrical energy consumption (HSU; LIU; WU, 2011), (REN; HE; XU, 2012), (WU; CHANG; CHAN, 2014), (KIM; CHO; SEO, 2014), (LUCANIN; BRANDIC, 2013), and to optimize the infrastructure deployment of renewable energy sources (MINH; SAMEJIMA, 2015), (GOIRI et al., 2011), (MÄSKER et al., 2016). These works present different approaches to managing the workload according to the electrical energy consumption perspective. The cited techniques apply contract requirements, electricity availability and prices, and the monitoring of the physical infrastructure, under specified operational conditions and contracted service levels.

Lucanin and Brandic (LUCANIN; BRANDIC, 2013) define an optimized scheduling based on the energy grid situation and the energy prices but include the renewable sources supply to establish allocation priority. The authors describe green instances like the ones that allow the resource allocation prioritization considering the energy prices and renewable energy source availability.

Applying a similar approach, Goiri et al. (GOIRI et al., 2011) establish a job distribution strategy oriented by solar panels energy generation. The authors define the energy generated as virtual energy provisioning, and, considering the energy surplus, determine the amount of power to process most of the jobs, reducing the grid energy consumption. Although both works define strategies to deploy renewable sources, they lack costs and service level establishment to prevent financial losses and sustain the cloud business services release.

Masker et al. (MÄSKER et al., 2016) evaluated a prediction approach, which introduces a Smart Grid scenario, to predict the renewable sources supply to determine the time intervals to schedule the workload. The work proposes a control center for the Cloud Computing that contains energy prediction, energy consumption, and grid exchange elements. The authors apply different energy prices to determine the most affordable workload processing strategy.

Minh and Samejima (MINH; SAMEJIMA, 2015) developed a scheduling proposal based on renewable sources costs. The costs reduction occurs by maximizing the renewable sources usage by decreasing the consumption non-renewable energy sources and scheduling the contracts according to renewables supply. The number of processors is also deployed to run the services is used as a criterion to optimize the scheduling process, along with the energy sources, avoiding contracting violations.

Ren et al. (REN; HE; XU, 2012) employ the electricity prices fluctuation for different allocations and time periods to schedule the contracted jobs. The deployment of fairness criteria, which considers the resources allocation as a priority coefficient, enables to establish the cost function for the jobs processing. The schedule decisions take into account different types of jobs, differentiated by an established geographical allocation and energy consumption

criteria, and the time intervals envisaged for the processing.

Monitoring the current power consumption of the infrastructure provides the status information for the scheduler element and enables to manage the incoming workload. The present state of the infrastructure energy consumption, considering the server capacity and CPU usage, informs the scheduler responsible for managing the workload based on the resource expenditure. For geographical distributed DC, the schedule decisions and the possible migration of jobs from one DC with a higher consumption and energy-related costs to another depend on servers status.

Hsu et al. (HSU; LIU; WU, 2011) propose a job allocation focused on reducing the extra energy for the physical servers. The authors allocate sequences of jobs to reduce the provisioned computing power of physical servers and establish three scheduling methods for the allocation of the optimized job; the authors define the distribution by jobs, according to the amount of computing power and its allocation deadline. The sum of the necessary power to run the jobs on processors establishes the optimized jobs sequence to reduce the energy consumption by the allocation process.

Wu et al. (WU; CHANG; CHAN, 2014) defines a DVFS (Dynamic Voltage and Frequency Scaling) to lower the energy consumption and schedule the resource allocation based on new servers usage. The method also establishes different weights to each incoming VM (Virtual Machine), which enables to prioritize the schedule decision.

Also considering the processors' energy expenditure, Kim et al. (KIM; CHO; SEO, 2014) evaluate an energy-aware schedule based on the estimation of energy consumption by the processors. The estimative of consumption maintains the VMs processing under contracted defined values. The estimative of power consumption by the computing resources is a strategy for the services allocation prediction and the management of the energy expenditure regarding service processing.

The referred work defines as a scheduling strategy the service description based on the energy expenditure forecasting. Setting an energy profile descriptor is mentioned, as seen before, as a management strategy for the cloud resources management. Although these works referred to an approach based on the servers status, its knowledge was required to set monitoring requirements for the energy expenditure status of a Cloud Computing and related infrastructure.

Despite the different approaches for the schedule decision taking, the works previously cited relate the capacity infrastructure and a prioritization criterion to define the most convenient way to allocate, process and monitor the cloud infrastructure. The initial establishment of SLA and infrastructure contract settings and the definition of weight criteria allow for prioritizing the service allocation; the SLA downgrade is cited as a negotiable contract item since the

service quality alters to promote the energy consumption reduction. To define the criteria for job schedule is required to set a sort criterion for resource availability, time allocation, and quality decrease acceptance.

The inclusion of internal monitoring and scheduling proposal, like the ones presented by Kim et al. (KIM; CHO; SEO, 2014), Hsu et al. (HSU; LIU; WU, 2011) and Wu et al. (WU; CHANG; CHAN, 2014), establish a profile-weighting plan as resource provisioning requirement. The definition of a profile, categorizing the contracted services demands, that enables an estimative of the required energy to process the jobs. Scheduling the Cloud Computing resource allocation taking into account the electricity usage and its availability fluctuation during determined day period enables to establish new pricing models. Due to this fact, a weight standard for the schedule proposal requires the evaluation of an energy profile.

The energy profile is defined as the accountability of the hardware components to be provisioned to fulfill the contract requirements. There are two techniques to determine the power consumption of the demanded infrastructure: CPU Power Models and VM Power Models (COLMANT et al., 2015). The first approach considers the processor's performance to measure the power consumption, and the second method uses the monitoring and machine-learning strategies to establish the power consumption by the VMs (COLMANT et al., 2015). Since the power profile requirement sets out a translation reference between the computing capacity contracted and the power consumption, the first approach is considered for the energy metrics consolidation. The profile must be able to measure the hardware usage without additional infrastructure metering (SMITH et al., 2012).

Thus, the second approach was considered to establish a pricing model during each time interval predicted by the model. A pricing estimation model that accounts the idle state and the power consumption by the servers determines the costs to allocate the services. Therefore, the power profiles enable to translate the computing infrastructure into power amount; this value is deployed to predict the power to be purchased from the electricity sector and used to maintain the infrastructure functioning.

The references for defining a power model are both directed by the direct translation of the computing infrastructure contracted into power quantity and the deployment of a power level for the processing stage. Teramoto and Huang (TERAMOTO; HUANG, 2012) developed a pricing model based on the VMs energy consumption. The price model presented considers the power consumption estimated for the VMs processing and a power model related to the resources demanded to value it. Bohra and Chaudhary (BOHRA; CHAUDHARY, 2010) evaluated a VM monitor that enables to link the energy expenditure by each cloud component - CPU, cache, memory, and storage - to the total consumption of the system. The authors set parameters for the components but do not include the network element, which is considered

by Smith et al. (SMITH et al., 2012).

The CPU consumption approach is established by Krishnan et al. (KRISHNAN et al., 2011), Jiang et al. (JIANG et al., 2013), Ma et al. (MA et al., 2013), Janacek et al. (JANACEK et al., 2012), and Ruan and Chen (RUAN; CHEN, 2015). The previously referred works establish a different relation between the idle state energy consumption and the expenditure for various VMs to define energy profiles. Besides, these works establish a relation between the VMs allocation and its energy expenditure. The relation between the energy consumption and the VMs profiles helps to set standard ways to measure the expenditure of the physical infrastructure.

Applying power profiles to account the services allocation helps to determine a balancing costs model for the Cloud Computing environment. How to price the cloud services sustainably is a complex subject for providers, moreover for the energy consumption of the structure. It ranges from metering and accounting the services processing (ANWAR et al., 2015) to dynamic pricing strategies (ZHAO et al., 2014) (MASHAYEKHY et al., 2016).

Dierent prices levels, given the exibility of the provided service or product, are the basis for dynamic pricing approach as a reaction to the market supply and demand dynamism (SCHWIND, 2007). The deployment of such criteria establishes a reactive supply method to allocate and, therefore, manage the contracted services of a Cloud Computing model.

Regarding the physical infrastructure provisioning, the accountability and prediction of the physical infrastructure are methods to compose costs (MARTENS; WALTERBUSCH; TEUTEBERG, 2012). The pay-as-you-go model for the Cloud Computing enables to charge the user for the type and amount of resources provisioned to process the service. The different quantity and profile of the resources deployed allow setting pricing strategy by the Cloud Computing provider (MASHAYEKHY et al., 2016).

The flexibility of the supply by both the Cloud Computing and the energy supplier sets conditions to establish price response according to loads flexibility. From the Cloud Computing perspective, the infrastructure provisioning set up by the service level contracted, and the quantity of infrastructure that must be deployed to process the contracts develops the condition to establish a dynamic pricing strategy.

From this perspective, to evaluate a dynamic pricing strategy for the energy-management service for a Cloud Computing environment requires a translation of the computing resources to energy-related metric. Also, the requirement to determine an energy profile for the energy management enables to evaluate weight criteria for the schedule decision taking.

Setting conditions for the users to define the processing intervals is developed by (LUCANIN; BRANDIC, 2013). The authors present the advantages of a contracting scheduling possibility to the cloud users. They evaluated the conditions for renewable sources generation and did not

develop negotiation requirements for a DR information exchange and other sources powering choices.

2.3 DEMAND-SIDE MANAGEMENT

Demand-Side Management (DSM) refers to the load management activities by taking practical measures to foster rational use of energy, saving energy, improving the energy efficiency, optimizing resources and lowering the electrical services costs (HU; HAN; WEN, 2013), (GOSWAMI; KREITH, 2015). Conceptually, the DSM is a part of the Integrated Resources Planning (IRP), which considers the supply-side and the demand-side resources to minimize the total cost of the energy service levels (HU; HAN; WEN, 2013).

DSM dates from the oil crises, as an attempt to reduce electricity demand, which implied in oil imports and negative environmental impacts (TORRITI, 2015). In the 1970s and 1980s decades, the DSM was implemented to limit the high costs to produce energy and to reduce electricity demand during energy consumption peak periods (TORRITI, 2015). At the beginning of the 2000s, along with environmental issues and concerns on the supply security, and after a break in the 1990s, DSM became an important topic for the future planning of energy.

The increase of renewable sources and new technologies, such as the Smart Grids related technologies, leads users to search for efficient energy use. The need for balancing demand and supply, in addition to environmental and infrastructure concerns, has shown the need of DSM evaluation for energy management (TORRITI, 2015).

Utilities developed and coordinated the DSM, but, initially, it did not include demand flexibility (COOKE, 2011). The financial market initiatives did not push the users' flexibility, which did not prove the effectiveness of the DSM (COOKE, 2011).

The customer flexibility creates market opportunities, for both users and energy providers. The new market situation encourages the concurrency and the launch of innovative products and services by vendors and transparent purchasing decisions and financial reward for users (COOKE, 2011). The definition of a new market based on the users' voluntary flexibility is known as demand response, demand-side participation or demand-side integration (COOKE, 2011).

Demand Response

Demand Response (DR) is a DSM program focused on the load shifting aspects for set time periods (TORRITI, 2015). The DR programs deploy commercial or financial incentives that encourage behavioral response to price or incentive techniques (COOKE, 2011;

HU; HAN; WEN, 2013). DR demands methodologies, technologies, and commercial arrangements allow consumer participation in the power system decisions, creating opportunities to renewable energy sources deployment and active management of the grid capabilities (LOSI; MANCARELLA; VICINO, 2015).

The programs are divided into passive participation, as on smart pricing services, and active participation. In passive programs, market incentives restrict the customer; on active programs, the participants can use the market strategies, services or load voluntary programs to adjust their consumption behavior (LE et al., 2016). Siano (SIANO, 2014) classifies the DR programs in Rate-Based or price programs, which implements DR through tariffs or contract signature according to electricity prices variation over time, and Incentive or Event-Based, which rewards the load reduction based on request or level of control over customer's appliances. The last program is named Demand Reduction Bids (DRB) and considers the users initiative to reduce demand offers.

The most common Rates-Based Programs are Time of Use (ToU), which defines prices for the energy according to time intervals, and Real-Time Pricing (RTP) that refers to continuous tariffs variation in response to spot markets, balancing markets, among others. Lastly, Critical Peak Pricing (CPP) establishes a higher rate triggered by pre-defined conditions, such as reliability and higher supply prices (SIANO, 2014; COOKE, 2011).

Some Incentive-Based Programs are the Direct Load Control (DLC) that allows the control of the consumer's equipment by the utility company, the Emergency Demand Response Program (EDR), which encourages the customer to reduce the load during non-reliable periods, and Capacity Market Programs (CPM) that promote customers to provide load reductions as substitutes for system capacity. Besides, the Interruptible/Curtailable program gives a discount for users that reduce loads based on requests and the Ancillary service market programs that pay the users that commit curtailing the capacity to support the power grid operation (SIANO, 2014).

The latest DR program, DRB is based on offering proposals to users when the market prices are high (SIANO, 2014). Although the DR programs encourage users to alter load usage patterns on account of cost reduction, they do not support the energy consumption reduction directly. Energy conservation programs can offer rewards for users, promoting lifestyle changes, while DR takes advantage of reliability and quality situations (TANG et al., 2014).

The reduction of energy relative costs on DC makes them probable DR programs participants due to theirs load flexibility (WIERMAN et al., 2014). The highly automated and monitored infrastructure of DCs, including ICT equipment and cooling, enables adjustments of the current load (WIERMAN et al., 2014).

Despite the cited advantages of DR programs, there is no consensus about the DC partici-

pation on these programs. Aside from the lack of maturity at most of the electricity markets nowadays, as the absence of regulation that enables the involvement of DC in the majority of market programs, there are remaining questions related to the risks, control of the infrastructure and market barriers (WIERMAN et al., 2014).

Dealing with the potential financial losses related to rate-based programs or performance decreasing, as on encourage-based programs, is a restrictive market point. The balance between the capacity usage and the grid indication of load reduction or peak-pricing strategy demands different contracts and load management terms to DCs participation in DR programs (WIERMAN et al., 2014). Along with the cited risks, the market rules must be reviewed to enable the incorporation of the bidding model on the DCs management system and prevent the market prices manipulation, since DC has the potential to participate more aggressively as major energy users (WIERMAN et al., 2014).

Siano (SIANO, 2014) mentions that some technical requirements should be provided to guarantee the DR availability. Demand reduction strategies are related to the energy price and emergency events, communication-enabler metering, energy tools that allow full knowledge of load data, load controllers, and energy management systems, and on-site generation equipment for emergencies or facilities first power requirements.

Demand Response programs for Data Centers

For DCs and, consequently, the Cloud Computing environment, the cited technological requirements can be evaluated as management systems capable of communicating and interpreting the DR requests of electric utilities and managing the ICT infrastructure. Load reduction and prices should be understood considering the management conditions of the operators, i. e., the resources handling is a responsibility of DC owners and not of the utilities.

Banerjee et al. (BANERJEE et al., 2009) evaluate the concept of sustainable DC based on the management of supply and demand. The authors determine the improvement of the physical infrastructure and the usage of local resources to reduce the impacts of the energy transmission and distribution. Alongside, the physical modules, pervasive sensing, resource discovery, and autonomous control enable the supply of the infrastructure and guarantee the SLA agreed with the user (BANERJEE et al., 2009).

The work evaluates the management of the infrastructure of a DC, including cooling, and IT types of equipment to achieve energy savings. The authors determine different management techniques, based on the users' requirements, costs and contracted SLAs, as a demand-side management of the DC resources. Also, the referred work establishes the balance of the supply and demand to evaluates the concept of energy efficient DC and unites the two areas

to achieve the best energy deployment (BANERJEE et al., 2009).

Berl et al. (BERL et al., 2013) mention the integration of DC to DR programs as a suitable alternative. The work sets a supply and demand agreement to encourage the energy cost reduction. The authors define green agreements between DC and energy supplier and establish SLA terms for the energy cost reduction. The green arrangements determine contract strategies to guarantee the security of the grid and also the financial gains for the energy supplier. The targets are established based on greenhouse gases emissions and to encourage the consumption of renewable energy sources.

Similarly, different works evaluate methodologies to enable the DC participation in DR programs. (WANG et al., 2013), (LE et al., 2016), (SUN et al., 2015), (LIU et al., 2014) and (LIU et al., 2013) determine different frameworks focusing on the DR programs for the IT infrastructure. Ranging from the energy costs reduction until the energy consumption or the possibility to raise renewable energy sources to power the infrastructures, as cited in the previous works, different authors mention prices and resources management as alternatives to the DR entry.

Despite the costs and consumption decrease presented by the suggested approaches, for Cloud Computing environments, the users contracted terms, and quality service levels should be considered for the infrastructure management. Quality contracted terms, such as SLA and QoS, undressed by the previously cited works, DC capacity usage, time and geographical fluctuation of the power generation enable to establish a scheduling arrangement for the Cloud Computing infrastructure.

As for the DR variation costs and availability through epochs of the energy offering, the power-aware management of the Cloud infrastructure is proposed. The power aware Cloud Computing environment modulates the power consumption of the equipment based on the processing tasks (MARKOVIC et al., 2013), and it can be implemented based on the Cloud instances provisioning, workload assignment, time-scheduling, and metering and monitoring of the infrastructure (NARAYAN; RAO, 2014).

The management of the virtual environment and efficient servers deployment are not considered, to limit the energy-management scope. Focusing on the costs decrease, based on the temporal and geographical flexibility of the electricity prices and power generation, the cited work examines the workload execution and scheduling of the capacity of the DC (KONG; LIU, 2015), including IT equipment, such as servers, storage devices, and network equipment, and cooling.

2.4 POWER-AWARE WORKLOAD MANAGEMENT

Power-aware management deals with the workload assignment under energy consumption status. It concerns for optimizing the resources usage taking into account energy costs budget and energy efficiency terms (PEDRAM; RABAEY, 2007). For the Cloud Computing, it relates to the best deployment of the infrastructure, including servers, heating, and cooling, to improve its energy expenditure performance.

The implementation of such concept for the Cloud Computing ranges from a geographical balance of the infrastructure, power-aware metering, and optimizing servers and related infrastructure usage (ADDIS et al., 2014), (NARAYAN; RAO, 2014).

Distributed DCs enables to deploy not only a temporal scheduling of the workload but an energy-aware spatial assignment. The Geographical Load Balancing (GLB) considers geographical distributed DC to assign the workload (KONG; LIU, 2015). Rao et al. (RAO et al., 2012) evaluate a management method based on the energy costs for the defined DC. Based on energy market real-time pricing strategies, the work assigns the computing load for distributed DC based on the prices and energy demanded.

The authors evaluated a resource allocation optimization to allocate the required load for different DCs, based on a two-stage technique. The first stage considers the announced prices by the utilities, focusing on Smart Grid scenarios, and the power consumption by each predicted DC. Thus the work considers both the power usage and the price focusing on the service levels contracted; the authors discuss neither the renewable sources deployment nor the evaluation of price differentiation allocation method.

Similar approach is considered by (YU; JIANG; CAO, 2015), (LIU et al., 2015), (DONG et al., 2013) and (CHEN; HE; TANG, 2012). The first referred work considers the energy management of DC as smart microgrids scenario. The authors consider a front-end that controls the services allocation for each predicted microgrid region, responsible for powering the DC. The front-end is responsible for dealing with the costs of the DC, distributing and scheduling the services and handling the transactions between the microgrid and the grid. Based on the workload knowledge, the model considers the microgrid status to allocate the services.

The last cited works determine the workload assignment based on renewable energy generation variation on time and spatial perspectives. Aware of the renewable sources provisioning, the authors developed methods to manage the workload to minimize the non-renewable deployment. The works evaluated different management method to deal with the variation of the renewable power generation in various geographic locations. Despite the fact that their primary purpose is to reduce the consumption of non-renewable sources, the authors cite the

dynamic pricing strategy to increase the efficiency of the approaches.

Lucanin and Brandic (LUčANIN; BRANDIC, 2016) developed a controller for the Cloud Computing that distributes the VMs from the renewables peak load perspective. Through a geo-temporal knowledge, the method forecasts the peak energy loads to manage the load, avoiding non-compliance of QoS and excessive migrations. The resources are optimized to prevent performance decrease through the maximum exploitation of renewables supply.

The availability and prices of the energy resources, especially the renewable ones, enables to establish the management of the contracted workload from the energy prices and availability perspective. Instead of considering a consumption approach, usually applied by efficiency-techniques, the services assignment takes advantage of the energy flexibility terms. This method enables to view the performance evaluation of the infrastructure from a different perspective.

2.5 CHAPTER CONSIDERATIONS

The concept of the energy as a manageable indirect resource of the Cloud Computing environment enables to establish strategies to cope with the infrastructure provisioning and resource allocation. Although many of the works reference the energy deployment as an energy efficiency strategy, the power-aware management of the workload regarding electricity provisioning has evaluated as well.

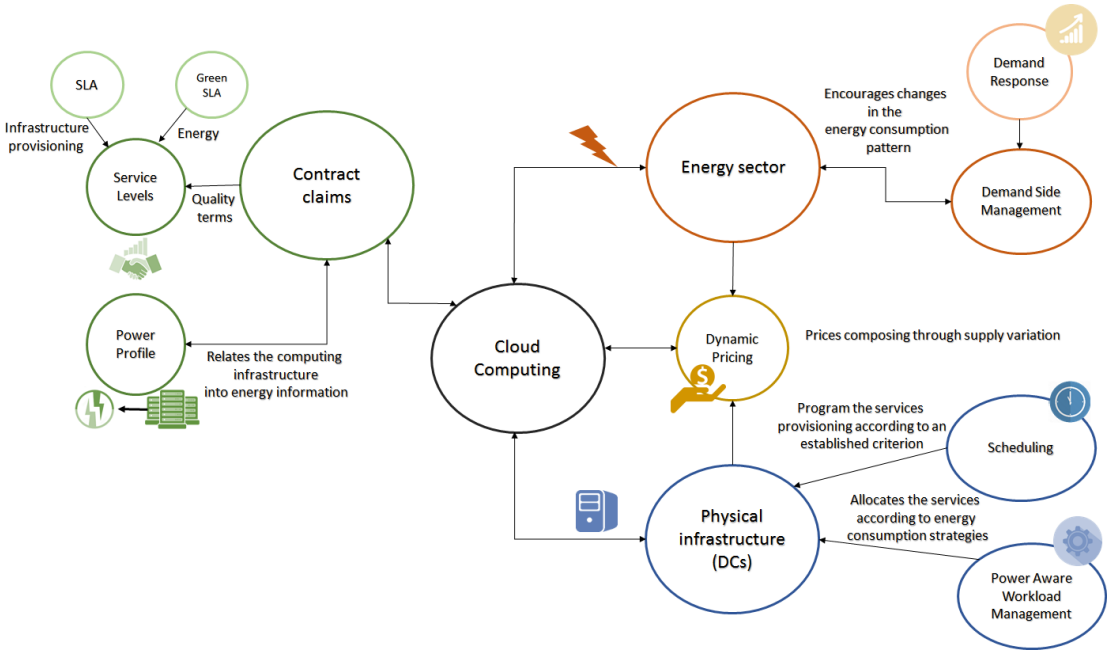
The demand management concept, previously established for the electricity sector, enables to state strategies to cope with provisioning the contracted workload. Ranging from renewable energy sources increase until a most efficient usage of the infrastructure, the provisioned demand may be manageable to achieve most favorable results for the service provider. The cited sources developed techniques to control from server allocation until methods to increase the renewable energy sources consumption.

Deploying management strategies from an energy consumption perspective enable to deal with the range of prices and generation fluctuation. Using these terms to control the contracted workload allows setting dynamic pricing strategy during the allocation. The dynamic pricing weights the services assignment and establishes a management approach from the costs context.

Figure 1 resumes how the different concepts connect to the Cloud Computing. It does not describe the implementation of the method or its flow, but bonds how the contracting stage, the concepts related to energy utilities and the management of the infrastructure.

SLA establishment and power profile concerns to the contracts claims and the translation of the computational infrastructure into power amount and quality terms. The energy sector

Figure 1 - Flow of the concepts studied and their relation to the Cloud Computing



relates to the behavioral programs that encourage changes in the resource consumption. The dynamic pricing, an economics concept, relates the supply variation and price composing.

The figure resumes the different concepts detailed in this chapter. The next chapter, 3, describes the problem and the assumptions assumed for this work development. Also, the concepts and the refereed works related in this chapter were applied to state functional requirements for an energy-management method development.

3 ENERGY-MANAGEMENT STRATEGY SYSTEM REQUIREMENTS

This chapter presents the motivation to develop an energy management approach for the Cloud Computing environment. First, section 3.1 describes the problem encountered in the existing literature. The section presents the problems referenced for a DR development for the ICT sector. It also encompasses the scheduling strategies developed by studied works, as referred in chapter 2.

Section 3.2 presents the assumptions taken for the development of a management approach, from an ICT energy consumption scenario and the evaluation of behavioral programs by the energy sector. The functional requirements for such approach are referred in section 3.3.

3.1 PROBLEM DESCRIPTION

The energy prices and supply flexibility through time periods are applied to encourage consumers to change its usage pattern. To take advantage of the given energy terms, a method able to cope with the time fluctuation of the resource is required. For ICT sector to make use of the energy provisioning variation, an approach capable of dealing with the load terms concurrently with the energy information is demanded.

Siano (SIANO, 2014) establishes that technical requirements should be provided to guarantee the availability of the ICT services alongside the sector participation in DR programs. The author claims that it must deal with energy price and emergency events, including full knowledge of load data, load controllers, and energy management systems. Therefore, to enable the ICT participation in DR programs, the method must be able to understand the infrastructure deployment and the energy supply.

Mainly for the Cloud Computing services provisioning, the solution should cope with the contracted quality constraints. The approach deals with the contract claims and enforces a strategy to provide the services in the most beneficial way for the owner of the ICT infrastructure.

Thus, a management approach for the Cloud Computing environment must deal with such information and ensure the services provisioning according to energy supply terms. A scheduling technique is deployed to manage the infrastructure assignment following energy variation and contracted constraints.

The specificities related to an energy-focused deployment range from contracts terms knowledge to the resource supply range. Table 1 shows the considered requirements for the management proposed. Its requirements were obtained from the scheduling references detailed in section 2.2. Taking into account these conditions, an energy-aware management method

was developed, selecting the best strategy to provide the services and fulfill the quality terms.

Table 1 - Scheduling methods proposed by other authors

Work	Contract Claims	SLA Terms	Infrastructure Monitoring	Renewable Energy Source	Grid Status	Allocation Costs
Hsu et al.		X	X			
Ren et al.						X
Wu et al.	X		X			
Kim et al.						
Lucanin & Brandic	X			X	X	
Minh & Samejima				X	X	X
Goiri et al.				X	X	
Masker et al.					X	

3.2 ADOPTED ASSUMPTIONS

Implementing energy efficiency strategies is an approach applied by the DC administrators to reduce the infrastructure powering impact on the business costs. Concurrently, the electricity sector evaluates itself to provide energy availability and information to the users. The transparency in the data information enables the customers to change the patterns of energy usage.

The load and time flexibility related to the computing services provisioning allows the development of Demand Management strategy regarding the Cloud Computing environment energy consumption. The associated elasticity of provisioned Cloud Computing infrastructure resources is managed to take advantage of prices and availability variation of electrical energy.

Establishing a management focused on the demanded resources enables to set negotiation terms regarding the powering maintenance of the infrastructure. The settlement of energy-related service terms allows taking advantage of the energy fluctuation and the computing resources to state bargaining claims with the Cloud Computing users.

The evaluation of energy terms allows setting the electricity deployment as a flexible element of the Cloud Computing model, setting favorable conditions for the resource allocation in different situations. The development of the electricity as a compliant resource of the Cloud

Computing environment sets conditions to provide the capability as a contractible item for the user and its purchasing negotiable within the energy market.

3.3 FUNCTIONAL REQUIREMENTS

For the development of an energy-aware management for the Cloud Computing environment, the following Functional Requirements were defined. In agreement with the referred works on section 2 and the problem and motivation previous described, the Functional Requirements outline the functions that the system can accomplish.

- FR1: Establish an interface between the Cloud Computing services provider and the electricity sector.
- FR2: Enable services allocation according to the energy supply availability.
- FR3: Establish a technical approach that enables to include DCs in DR programs deployed by the energy utilities.
- FR4: Set an interface with a DCs infrastructure orchestrator, responsible for the services physical distribution.
- FR5: Set an assignment strategy according to the physical consumption of the computing resources, from the monitoring of the infrastructure perspective.
- FR6: Provide contract requirements that allow the provisioning of the energy as a service for the Cloud Computing user.
- FR7: Support SLA and GreenSLA contracted claims that refer to the energy deployment for the infrastructure and the demanded quality of the services provisioning.
- FR8: Assign the services contracted in geographically allocated DC.
- FR9: Guarantee the computing infrastructure provisioning according to the contracted service levels.
- FR10: Establish a dynamic pricing strategy for the contracts assignment.
- FR11: Include the energy efficiency as a contractible term for the Cloud Computing user.
- FR12: Add the type of energy source as a contractible term for the Cloud Computing user.

- FR13: Establish a scheduling strategy following the energy supply, prices and efficiency requirements.
- FR14: Set the services schedule depending on time requirements, predecessor activities and computing resources allocation.
- FR15: Establish a power-aware assignment of the provisioned infrastructure, according to energy status.
- FR16: Set a power-aware management to take advantage of geographical allocation of the physical infrastructure.
- FR17: Guarantee the reliability of the data exchanged with the Cloud Computing user.
- FR18: Replicability of the model, regardless of the energy market regulation directives.

4 METHOD DESCRIPTION

This chapter describes the energy management approach proposed. Taking as the basis the addressed requirements in chapter 3, the developed method is detailed.

Section 4.1 presents the concept of the energy as a manageable resource. The section details the detachment of the energy and the related computing infrastructure of the Cloud Computing. Section 4.2 presents the management approach for the environment, and its contracting, organization and scheduling steps. Last, section 4.3 defines the workflow presented as a possible implementation of the management approach.

4.1 ENERGY AS A MANAGEABLE RESOURCE

The work uses the energy prices and supply variation to define strategies for the contracts allocation. The variety of the resource supply, including its amount of self-generated energy and geographical location, and its prices establish the conditions for the demanded infrastructure provisioning. Additionally, the energy expenditure of the DCs is continuously monitored, enabling the operator to have full knowledge of the infrastructure deployment.

Demanded infrastructure is defined as the computing capacity to be provisioned, according to contract terms. Additionally, the power required to process the contracts is described as a demanded infrastructure in this work. Therefore, the resources deployed to maintain the services provisioning according to the requested quality terms are defined as demanded infrastructure.

The cited energy terms allow the operator to set the services powering. This strategy encompasses a costs reduction approach, better use of generation peaks, and efficiency targets for the contracts assignment. How to power the infrastructure, concerning the most favorable deployment of the energy terms, sets an energy-aware management strategy.

To establish the energy-aware management, the energy is considered as a separated infrastructure resource of the Cloud Computing, and the distribution of the computing resources follows the energy availability. This condition sets detached management strategies for the computing and the energy resources. At this point, the computing infrastructure concerns to the usage of the servers, but includes cooling and other resources required for the service provisioning.

Although the developed approach defines the energy as a detached resource, the information regarding the computing resources provisioning sets the conditions to power the infrastructure. The demanded computing infrastructure is applied to adjust the quantity of the energy to guarantee the functioning of the environment. Also, the monitoring of the computing

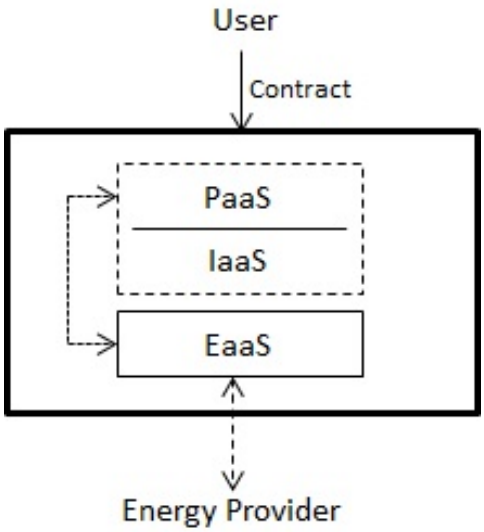
infrastructure reports the status of the services powering.

Along with the full knowledge of the computing capacity demanded, the energy management requires the Cloud Computing to be aware of the energy prices, supply and grid status. This work does not specify a regulation for the energy market. Thus, the related information concerning the energy status, as referred, is referred as energy sector and it updates the supply and prices data. The information regarding the energy deployment and supply status must be done bilaterally. These data enable to set time and quality requirements for the services supply, helping to establish powering strategies.

The Cloud Computing environment must, therefore, be capable of translating the energy information into terms that enable to set management strategies. A new service layer is proposed to cope with the management established. The new service layer exchanges information with the energy sector translates the computing workload into powering terms and knows about the infrastructure deployment.

The new service layer, named Energy as a Service (EaaS), receives the different information concerning the contracts provisioning and sets the most beneficial method to allocate them. The EaaS, as shown in figure 2, receives information of the DCs and energy sector to be aware of the resource status.

Figure 2 - Service Layers for a Cloud Computing environment.



To know the demanded computing resources, the EaaS communicates with the existing services layers. Although the energy layer focuses in contracts assignment independently of the other layers, they exchange information regarding the required computing resources. This positioning of the EaaS, displayed in figure 2, sets the layer as the one that manages the contracts but uses the information of the other layers to predict resources provisioning.

Defining the energy as a manageable resource allows selling the energy provisioning as services for the user. As a result, the user sets constraints to establish the quality of the service assignment, including how to power the contracts allocation. Claims such energy efficiency, type of energy source, and prioritization of the services are stated as contracted constraints; these claims are defined as energy service terms concerning the allocation.

Service levels regarding the resources assignment may be offered to the user of the Cloud Computing. The service levels that guarantee the quality of the service provisioning may be negotiable to set the performance tradeoffs (such as longer delays, limited computational resources, among others) acceptable in function of the energy consumption. The energy efficiency and availability become negotiable along with the allocation and migration of the services to more affordable DCs, in return of a different SLA or cost-cutting.

4.2 ENERGY MANAGEMENT PROPOSAL

The previously presented concept of the energy as a manageable resource of the Cloud Computing environment enables to set strategies to cope with the workload from the energy consumption strategy. Section 4.2 details how the management proposed from the energy deployment perspective was developed.

Services hiring, including the stated contracting terms and its translation into energy constraints, are described in subsection 4.2.1. The contracts organization and prioritization during the allocation process are presented in subsection 4.2.2. The decision-making related to the services schedule is shown in subsection 4.2.3.

4.2.1 Services hiring

This work applies the demand-side management concept, from the Cloud Computing services provider perspective, to negotiate the services allocation. Although the concept predicts the demand management by utilities or government sectors, the energy load management and the incentives enforcement are deployed by the Cloud Computing provider. The proposed management does not interfere with the cited sectors and their responsibilities.

An incentive-based program is intended to change the load pattern for the Cloud Computing environment; the incentive is deployed through contract terms that determine quality and negotiation availability for the resource provisioning. These claims include, at this stage, type of energy source, willingness to energy efficiency techniques and availability to negotiate the time and geographical services allocation.

To comprise the quality term acceptable by the user, processing plans are offered during the

hiring stage. The processing plans enable the service provider to state the demand flexibility to the energy and capacity variation through time. Named **reserved**,**flexible** and **on-demand**, the plans define the resources provisioning to the user, both computing and electrical energy resources, and the pricing strategy for the allocation.

To what extent the knowledge, no Cloud Computing service provider sells plans according to the availability of the energy resources. Currently, Amazon Web Services (AWS) sells infrastructure provisioning deploying on-demand, spot, and reserved instances (AWS, 2017), (AWS, 2017). Instances correspond to the infrastructure required: the reserved instances enable the client to buy in advance for the infrastructure, assuring the most affordable resources provisioning. The spot instances offer the possibility to pay for unused instances; this infrastructure is contracted through bids. The **on-demand** type of plan provides the resources provisioning through immediate contracting, without the previous provisioning according.

Although these plans correspond to the resources provisioning, they do not relate the powering of the infrastructure during the services allocation. Lucanin and Brandic (LUCANIN; BRANDIC, 2013) establish a *green plan* that allocates services during time intervals with higher renewables availability. The former work develops a scheduling strategy for the contracts based on the availability of cleaner energy sources.

Despite the definition of different plans that deal with the energy supply, the cited **on-demand**, and reserved plans, as defined by Amazon, do not set a contract constraint that enables the user of the Cloud Computing to choose how to power the services allocation. The variation of the electricity and business requirements demands contracting terms to cope with the services distribution according to the electrical energy fluctuation.

The present work states three different contracting constraints that enable the user to set time flexibility and management option regarding the energy availability, through prices, supply and sources, and the computing capacity. These plans determine the services processing flexibility according to the electricity deployment; thus, they do not set a prioritization for physical allocation or a fastest physical processing. The plans establish the priority concerning the lower processing costs and time requirements from the energy expenditure and its purchasing perspective.

- The **reserved plan** is contracted in advance and setting time restraints to process the contracts. This plan takes advantage of a previously defined contract to negotiate the assignment of the service; the benefit for the Cloud Computing services provider is to manage how to allocate these contracts regarding the demanded computing infrastructure and the electrical energy generation and prices prediction. Despite the negotiation capacity for these plans, they are not the most favorable for the user. Due to the processing time defined by the user, the contracting costs may endure the variation of the

electrical energy availability and costs during the period.

- The **on-demand plan** requires the immediate allocation of the services, independently of infrastructure usage status and energy resources costs and supply availability. Due to the lack of previous resources provisioning account, this type of plan is the most expensive one independently of the time interval required. The **on-demand plan** is the most critical one since its allocation can be demanded on intervals with an extensive use of the infrastructure, causing competition for the resources with the others contracts or overpassing a security limit for the resources deployment.

- The **flexible plan** enables that the allocation during the periods with the lowest electrical energy prices and availability, and the highest amount of available computing infrastructure. During the contracting stage, a maximum time interval defines the allocation, but the parameter is not open for the user. This plan has the more resilient quality and time restraints terms, since its distribution is dependent on the energy management approach of the infrastructure.

The plans encourage the Cloud Computing services users to adapt their processing demand to the hours with most available computing capacity or intervals with more affordable energy, establishing a dynamic accountability for the infrastructure deployment. They also enable to account the convenience to determine hour and requirements for the contracts assignment, by an electrical energy deployment perspective. The costs of the selected plan vary according to the electricity information and expenditure; this cost variation is deployed through service levels.

Due to this fact, energy efficiency, type of energy source, time constraints, quality of the service and the estimated budget are some of the aspects considered during the contracting step. As the commitment of greatest Cloud Computing providers in powering their infrastructure with renewable energy sources tends to grow (AWS, 2017), (MICROSOFT, 2017), inserting the energy source as a contractible constraint may introduce a new business perspective. The proposal offers the renewable energy source, non-renewable source, and a third option if the user is indifferent to the energy source.

Time requirements define the initial and final hour for the allocation period; also, the maximum interval agreed for the processing, named latency, is established. The current approach applies time restraints according to Time of Use (ToU) DR programs, i.e., 24-time intervals of 1 hour of duration. These 1-hour intervals are deployed to estimate the energy to be provisioned and to schedule the jobs. The time interval for the infrastructure monitoring and the trade of data regarding the energy sector may be defined by other periods.

The latency is considered to avoid the violation of quality terms, especially if the contracts accept rescheduling. The time conditions depend on the strategy established by the services provider; during this work evaluation, the one-hour interval was chosen, but this period is not a limitation for the management implementation.

To determine the incentives, a cost composing item, named Service Level (S_L) and a variable, according to time-requirement terms, that charges the users according to the capacity usage and energy consumption terms. S_L and compounds the cost and is the constraint that balances the type of contracted terms to the resources supply, and forms the value of the contract assignment.

S_L links the type of plan contracted to the price of the services allocation for the user. The value of S_L changes during the day, in line with the energy supply and prices. Also, this variable allows composing the cost regarding the supply variation through the contracted period and available resources capacity. This value refers to pricing the services assignment for the user, from the perspective of the energy deployment at the predicted time interval.

Therefore, S_L is a weight term that enables to negotiate the quality and assignment terms with the Cloud Computing users and to encourage the user to search for the most affordable processing hours, as determined by the energy market. The **flexible** processing plan is the most favored by the S_L variation, and S_L is the shortest for this type. On the contrary, the **on-demand** plan has the highest S_L cost, considering the risks and priority demanded by this one. The cost for the **reserved plans** varies through the day based on the S_L range.

After the user set the contract restraints, the service provider translates the information into energy-related terms. The contracts turned into an energy-provisioning term is named job. The job describes the Cloud Computing contracts into electrical constraints, including the computing resources predicted and the quality terms.

During the contract settlement, the Cloud Computing services provider translates the computing resources into an amount of energy required to maintain the services. The computing infrastructure demanded is converted into electricity metrics by power models. The power model establishes a relation between the amount of computing resource - memory, storage, network and CPU capacity - and the quantity of power demanded to execute the contracted services.

The resources supply, both computational and energy, distinguishes the processing deadline and availability terms for each contracted job. The jobs description stage interprets the quality and resources supply concerning the electricity availability and costs prediction. The amount of power predicted, and the computing infrastructure demanded, informed by the jobs, determine the allocation costs.

The capacity contracted, predicted energy and S_L are applied to determine the distribution

costs for the user. The JobCost (1) allows setting the value regarding the required energy service terms. JobCost determines the cost for each contracted service through the energy prices, the energy consumed, the capacity demanded.

The JobCost provides information that enables to negotiate and search for the most affordable, from the electrical energy perspective, allocation of the services. The assignment cost, according to the contracted plan, and time requirements are represented by the S_L index. The S_L is variable along the time contracted, and it changes the price of the energy resource, according to energy prices *p(t)* information.

Two different states reflect the demanded infrastructure to run the services. The amount of energy required to run the job (*JobEnergy*), defined during the job description stage, determines the cost focused on the energy deployment. Along with the amount of energy, a correspondent price related to the resources deployed to run the job compounds the costs for the user. The cost to free the infrastructure for the service processing resumes both the computing and powering deployment. The variable Service Cost (S_C) represents the cost of the provisioned infrastructure to run the jobs, and it is dependable of the physical resources deployed.

The JobCost for the user is determined for each time interval contracted for the user, and its result is the sum of the cost, from the electrical energy perspective, for the whole epoch established.

$$JobCost(t) = \sum_{t=ti}^{t=tf} SL(t) * SC + p(t) * JobEnergy \tag{1}$$

- t: time period stated by contract (h)
- ti: initial interval of the allocation (h)
- tf: final interval of the services allocation (h)
- S_L: Service Level
- S_C: Service Cost (monetary unit $)
- p: energy price (monetary unit/Wh)
- JobEnergy: estimated energy to run the service (Wh)

At the end of the contracting stage, the information compounds jobs. The jobs contain data that enables to classify the contracts into costs and energy deployment terms. The next step sets a strategy to deal with the allocation of the jobs and manage the quality demanded.

4.2.2 Contracts prioritization

After the contracting stage, the contracts are described as jobs. From this step ahead, the contracts are only named after jobs and managed, until the end of the allocation stage, at this term. The jobs are handled, from an electrical energy perspective, related to the predicted consumption and costs.

At this stage, the initial information regarding the jobs concerns the type of contracted demand. Although the processing plan does not establish a physical capacity prioritization between the jobs, it points out the requested time or processing urgency, and the willingness to reduce the processing costs by the assignment during the most affordable energy consumption time intervals. Thus, the first detachment among the jobs is by plans; the form chose to deal with the jobs at this step is queuing.

The queues, one for each established plan, enables to organize the time requirements of the processing stage and control the demanded infrastructure. The proposed management handles with the queues according to the energy consumption perspective, including the current costs and type of source, and the computing resources predicted to process it.

Managing jobs into a queue strategy allow to control the remaining energy and to predict some resources demanded the future allocation. This format enables establishing priority specificities for each defined plan, including the time required and quality contracted terms. The queues manage the remaining energy load as a method to deal with the remaining jobs.

The demanded energy composes the workload for the queues; for each included job on the queue, the predicted energy resource is summed to the current workload. The workload is managed according to the energy deployment and purchasing stated, and usage level of the physical infrastructure of the Cloud Computing environment.

The comparison between the energy supply, the quality-required levels, and the workload establish how to cope with energy efficiency and consumption. During exceptional situations, the jobs can be renegotiated following energy supply and consumption levels. According to the contracted plan and other quality constraints, the energy management approach may alter the processing time or migrate the jobs to a more affordable DC. The renegotiation of the jobs, for the cited situation, is dependable of the status and organization of the queues.

In specific cases, such as overloaded physical infrastructure, an excessive amount of **on-demand** plans, spare energy supply, shortage of electrical energy or prices variation, the jobs can be re-managed. The referred situations imply in active management of the Cloud Computing provider; from the user perspective, since the rearrangement of the jobs shall mean the quality decrease, the adjustment must be bilaterally agreed. In such cases, the queue and jobs are reorganized to deal with the energy provisioning requirements.

The management of queues must be constant to deal with the different time periods contracted, quality terms and usage status. Therefore, there are two separate management stages for queues. The first management is required during the inclusion of new jobs on the queues, to guarantee that the jobs are going to be allocated accordingly to the time and priority requested by the user. The second one concerns to control the remaining energy and to renegotiate the jobs status.

Flexible plans are the most willing to be altered during the allocation stage; they are the first ones verified for renegotiation. Since it is open to search for the most favorable conditions for the distribution, the job can be paused, has its processing time re-scheduled or migrated.

The **reserved plan** can also be re-scheduled. Its acceptance of new scheduling is lower in comparison to the **flexible** plans, but it is possible to search for new physical allocation and most favorable costs to reassignment the jobs. The only acceptable renegotiation for the **on-demand** processing plan is the physical migration of the job, due to the time limits and immediate allocation.

For each queue, an internal sort process is required to establish a prioritization for the distribution and renegotiation stage. Both steps demand the knowledge of which are the most appropriate jobs, from a quality and time view. Therefore, for each queue, a criterion is established to determine their control.

The internal standards place a dynamic pricing strategy to prioritize the jobs. The variation of the energy supply and prices and the computing capacity establish conditions to price the services assignment. Since these terms form the costs for the resources delivery, the variation of these terms enables to set prices for the users. The variation of the costs assignment provides the information related to the priority of the allocation or the renegotiation of the contracts.

The priority criteria allow organizing the jobs for the physical distribution stage. However, the full knowledge of the infrastructure consumption status, the electrical energy generation levels and purchasing information, and the contracted jobs allows searching for the most favorable allocation for the services. The next section details how the distribution stage was evaluated for the energy management of the Cloud Computing environment.

4.2.3 Contracts Scheduling

The demand-management applied until the organization and renegotiation stage enables to cope with the contracted terms regarding the most favorable energy deployment perspective. The contracts provide the information related to the infrastructure prediction to process the demand and energy-related terms. The quality terms and time requirements are managed to

cope with the different energy-related information.

To establish the beneficial mean to allocate the services, the Cloud Computing services provider maintains the contracted demand. From the previous steps, energy service terms set the quality and time requirements for the services provider to handle and foreseen the capacity. To set processing plans allows the manager of the environment to change electricity load by the fluctuation of electrical energy during the period. To learn the energy consumption, the service provider must exchange information with the suppliers of the resource and with the physical infrastructure.

The demand-side management forecasts infrastructure requirements and processing cost to establish the most beneficial jobs distribution. The interaction with the demand-side, represented by the queued jobs, sets the contracted workload for the environment and provides the acceptable energy service terms to ensure the processing. The supply-side gives the information regarding the electricity supply, both self-generated and the purchased amount, and prices variation.

Although the proposed management interacts with the electricity sector, how the market broadcasts the information regarding the energy and the purchasing process is not detailed in this work. Due to restrictions and several different business models for the sector worldwide, specifying the market is not an objective of this proposal. From the energy management perspective, the established prices during stated time periods and the available electrical energy are adequate for the proposed approach.

A scheduling strategy manages the demand according to energy terms declared. The contracts schedule uses the energy fluctuation as the primary criterion to allocate the jobs. Target intervals establish how to deal with the jobs, computing infrastructure, and the electricity supply and prices. Despite the fact that one target period is stated, demand, supply, and computing capacity are continuously monitored to detect free resources opportunities or situations that may jeopardize the services provisioning.

The physical infrastructure consistently provides information about the energy consumption and energy efficiency opportunities. The energy consumption includes both the current deployment of the electricity and cases that the usage of the resources is under the predicted level; the energy efficiency notices the best usage of the infrastructure and warns about the capacity and negotiation chances. To deploy energy-efficiency techniques or the physical allocate the jobs are not a target for this work. From the electrical energy deployment, the jobs schedule is a suggestion regarding a strategy to manage the energy of the Cloud Computing environment, and do not concern about the best consumption of the infrastructure.

Since the expectation is the infrastructure growth, there is the concern to deploy the scheduling of jobs for various DCs. The scheduling strategy considers an amount of multiple DCs

under the same manageable environment. Also, the scenario of a significant number of DCs defines the possibility to migrate the jobs according to the energy deployment. The information exchanged between the DCs and the scheduling interface allow to monitor the infrastructure deployment.

Different variables compose the costs to maintain the functioning of a DC, such as maintenance of computing resources, cooling infrastructure, human resources, etc.. How to deal with such operational costs concerns to the business model of the service provider, and it is not a focus of this work. Thus, this proposal deals only with the energy provisioning costs, including its use to sustain the infrastructure functioning.

The approach constantly searches for most affordable assignment opportunities, according to the current costs. The approach sets two different stages for the contracts assignment, named *prediction* and *renegotiation*. The two stages compare the current costs for the services provider as a weight to select the cheapest infrastructure deployment. The first stage, the *prediction stage* deploys historical energy supply information, to set which DC is the most affordable to process the contracts.

The second stage, the *prediction*, continually compares the current costs of the infrastructure deployment, supply and energy prices to reschedule the jobs. The last stage applies the quality terms contracts to reschedule the jobs and not to violate the latency. The action to reschedule the jobs is taken based on the comparison of the JobCost and the cost of each infrastructure of the Cloud Computing environment.

The Energy Cost (EC) for each DC determines the current value to allocate one job. The cost is the average amount to maintain the DC powered, at the measurement moment; the cost splits the total cost of the energy expenditure among all jobs allocated. The powering cost includes the power expenditure by servers in the idle state (P_{idle}) summed to the total energy consumed to assign the jobs (JobEnergy).

Equation 2 calculates the cost for each DC (*EC*), deployed as comparison criteria for jobs allocation, rescheduling or migration. EC changes depending on the number of jobs allocated on the DC (JobEnergy and Job) and the price paid for the energy supply (pe). Although the proposed managed does not deal with the internal management of the servers, the amount of power consumed by the physical infrastructure is applied to determine the price of the DC (Pidle and JobEnergy). The measured energy enables the inclusion of the capacity factor for the cost composing.

$$EC(t) = \sum_{t=ti}^{t=tf} \frac{\sum_{s=1}^{S} Pidle * pe(t) + JobEnergy(t) * pe(t)}{Job} \tag{2}$$

- EC:Energy Cost for the infrastructure (monetary currency)

- ti: initial measurement time (h)
- tf: final measurement time (h)
- s:quantity of servers used
- P_{idle}idle: amount of power consumed by the server during the idle state (W)
- pe: price of the energy at the monitoring moment (monetary currency/Wh)
- JobEnergy: power used to run the current processing services (Wh)
- Job:quantity of jobs processed during the monitoring

The DC status is informed to the system in established intervals; the present work proposes intervals of 15 minutes. This period is chosen due to the standard energy measurement interval, based on the electrical measurement devices. During this interval, the management approach knows the price paid for energy and quantity of jobs addressed to each DC. Thus, the sum of power deployed for each job and to maintain the servers functioning – even in idle state -, and the value paid for the energy used divided among all the running jobs provides the amount of the energy costs for the DC.

The energy costs for each DC and the allocation cost of the job are applied as a comparison to avoid the physical distribution in an infrastructure more expensive than the cost. For the physical distribution, the type of power source is enforced as well: the jobs that demand the allocation powering by renewable sources must be ensured. For the jobs that do not specify the type of source, the criterion is only the cost.

The responsible for coping with the demand, supply, and the infrastructure is the service layer proposed, named EaaS (Energy as a Service). The service layer receives the status of the DCs, the information of the supply and the jobs demanded and set the conditions regarding the supply and the infrastructure energy consumption levels. The EaaS is, therefore, responsible for coping with the different energy terms and points out the most favorable manner to deal with the demanded resources by managing the jobs.

Several steps, from the request receipt and translation into energy-related terms, until the balance of the supply and infrastructure, compose the proposed solution to manage the energy in a Cloud Computing environment. The proposed stage comprises a solution for a demand management focusing on a more efficient deployment of the energy. A solution that includes the different steps composing a workflow is detailed in the next section.

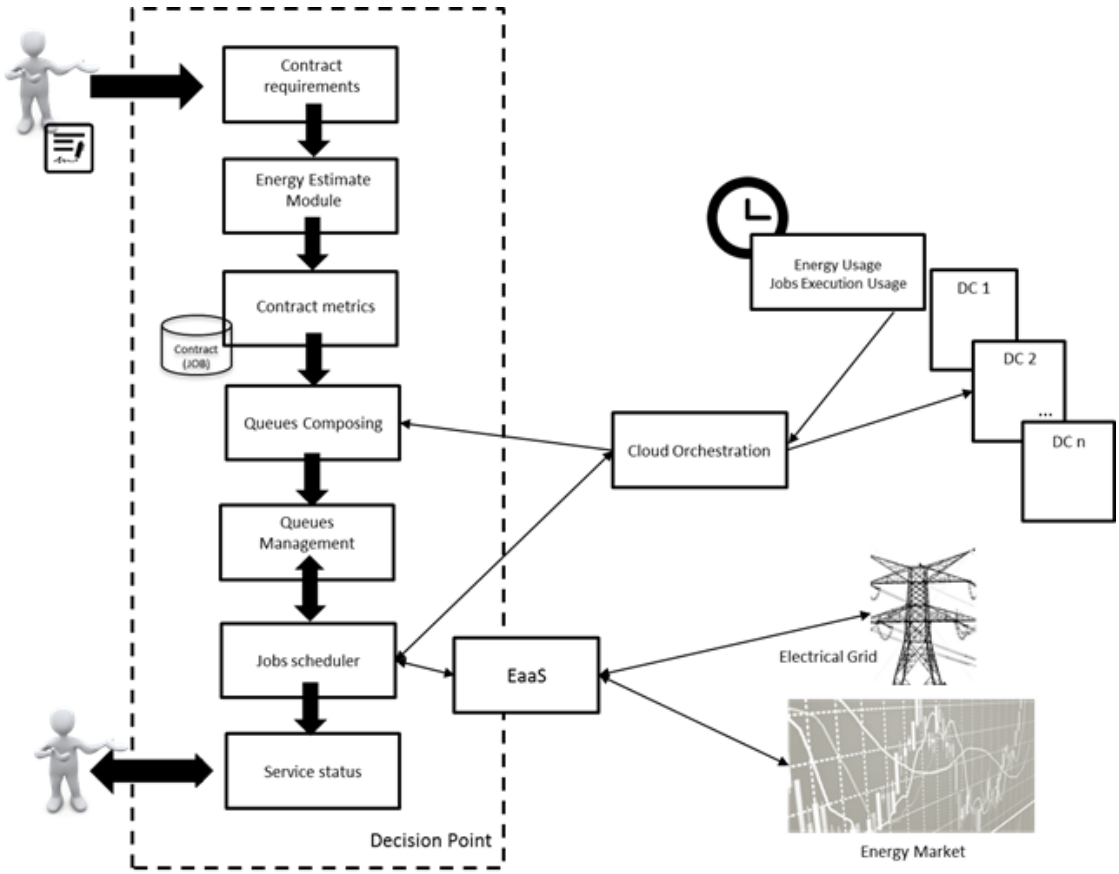
4.3 ENERGY MANAGEMENT WORKFLOW

The electrical energy management proposed applies the demanded energy service terms, information regarding the availability and prices, and the monitored status of the infrastructure to process services in a Cloud Computing environment. The proposal consists of three-stage management, which enables the Cloud Computing services provider to deal with the different information in the most favorable way. A service layer, named EaaS, deals with the stated roles interested in the management proposed.

A scheduling workflow, which enables the implementation of the proposed solution, is defined. The solution contains modules that deal with the contracts, the electricity sector, and the physical infrastructure. The workflow is named Energy-Efficient Cloud (E2C) and manages the specificities proposed to take decisions about the energy deployment. The E2C composes the Cloud Computing structure, focusing only on manage the resources regarding the energy consumption, according to management and expenditure strategies.

Figure 3 shows the workflow for the E2C. The Decision Point composes the E2C, that deals with the demand-side information and the EaaS. The service layer deploys the energy-related information, negotiates with the energy supply and knows about the DCs status.

Figure 3 - Decision Point Workflow



The Decision Point contains modules responsible for receiving the contracted terms, describe the items into jobs, organize the information and control the workload. Also, the structure negotiates with the user the possibility of processing changes and warns about the processing status. The first module is responsible for receiving the contracts defined by the users of the cloud. The *Contract Requirements* module is the interface with the user, identified for the infrastructure and energy needs of the user.

The first stage receives the contracts defined by the user – PaaS, IaaS, and EaaS requirements. The following module (*Energy Estimate*) is responsible for the translation of the contract requirements into metrics. The module estimates the amount of energy to be deployed to run the service based on the computational resources pledged.

The *Energy Estimate* module applies a chosen power model to establish the future purchased energy. The module estimates the quantity of power to maintain the contracted service functioning. The computing resources demanded, including memory, hard disk, networks resources and CPU (Central Processing Unit) capacity, required by contract, enables to predict a quantity of electricity to maintain the services allocation.

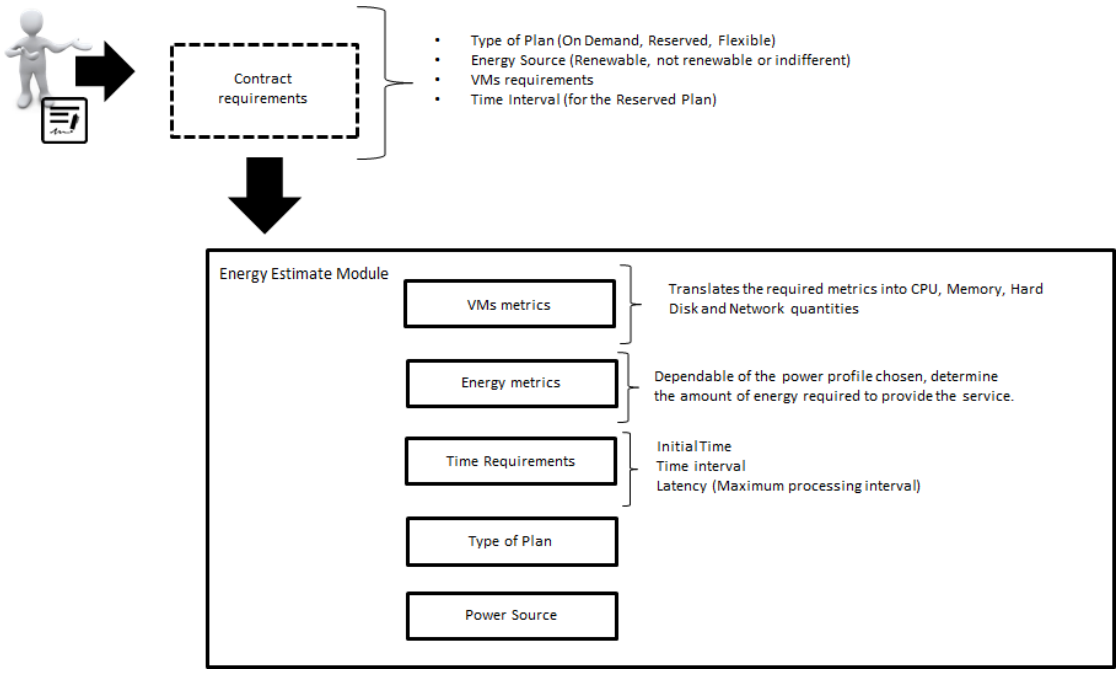
As a power model that deploys the physical infrastructure referenced to estimate the required energy expenditure, the one developed by (SMITH et al., 2012) was used to improve the present work. Because this model applies a direct relation between the energy expenditure of the computing resources and its employed amount, this model allows translating the contracted resources into a quantity of power. Since predicting the amount of energy required to maintain the services functioning is not the focus of this work, a power model that provides the power in agreement with the computing resources was appropriate.

The computational resources, the energy, the time constraints, the type of processing plan and the type of energy source are metrics contracted and deployed to determine the amount of power to maintain the services allocation process. Figure 4 shows the *Contracting Requirement* and *Energy Estimate* modules. At the end of the contracting terms translation, the energy service terms compose the jobs.

The next module, named *Contract Metrics* receives the predicted electrical energy and service terms and describes the jobs. Processing plans, initial and final time processing period, latency, availability, and quality of service conditions, type of energy source and cost are the terms that describe the jobs. Also, the contracts estimate a budget value; the budget helps to evaluate a dynamic pricing strategy for the allocation. The budget is compared with the costs and contributes to compose the queues and to search for the most affordable energy costs.

After the jobs description, the *Jobs Organization* module organizes the queues and sorts each queue for the quality, time terms, and costs/budget. The module also controls the amount of remaining workload and the jobs allocation flow. In the case of pausing and renegotiation of

Figure 4 - Contracting requirements and energy service terms translation



assignment, this module controls the restraints to reallocate and re-schedule the jobs.

The *Queues Management* module deals with all the queued jobs, including the **on-demand** ones. This module exchanges information with the scheduler module and controls which jobs are more suitable to be processed. The management module is the one responsible for comparing the costs and budget data, monitor the time demanded and latency. The module does not control the workload, although it is the one in charge of the scheduling direction strategy.

Jobs Scheduler module receives jobs selected for the assignment stage. The module is the responsible for managing the demand based on the electrical energy availability and prices, and energy consumption and infrastructure deployment. The EaaS is the module that balances the request information, provided by the *Jobs Scheduler* module, the infrastructure usage level, and the electricity sector.

Since the quantity of DCs is not a restriction for the energy management deployment, the control of the resources consumption is represented by the Cloud Orchestration. The Cloud Orchestration and the EaaS modules exchange information regarding the physical infrastructure usage levels and energy efficiency requirements for each DC. The communication between the two modules is suitable to inform which jobs are assigned; due to this fact, the EaaS knows the infrastructure but does not act on the internal allocation of the resources.

The EaaS communicates with the electricity sector, represented by the Electrical Grid and Energy Market. The communication enables the EaaS to know about the prices, considering that the DCs may be situated in different regions, the self-generated electrical energy, and the

grid situation. The exchanged information opens a purchasing and negotiation channel with the sector since the EaaS has the knowledge of the energy predicted consumption.

The *Jobs Scheduler* module is responsible for detecting renegotiation opportunities. Due to this fact, the information regarding the jobs processing is bilateral with both the user and the *Queues Management* module. In the case of pausing or re-scheduling the jobs, the module warns the Cloud Orchestrator, and it reassigns the workload to the *Queues Composing*. The new allocation follows the workflow again, according to the quality terms negotiated.

Figure 5 - Data translation during the assignment process

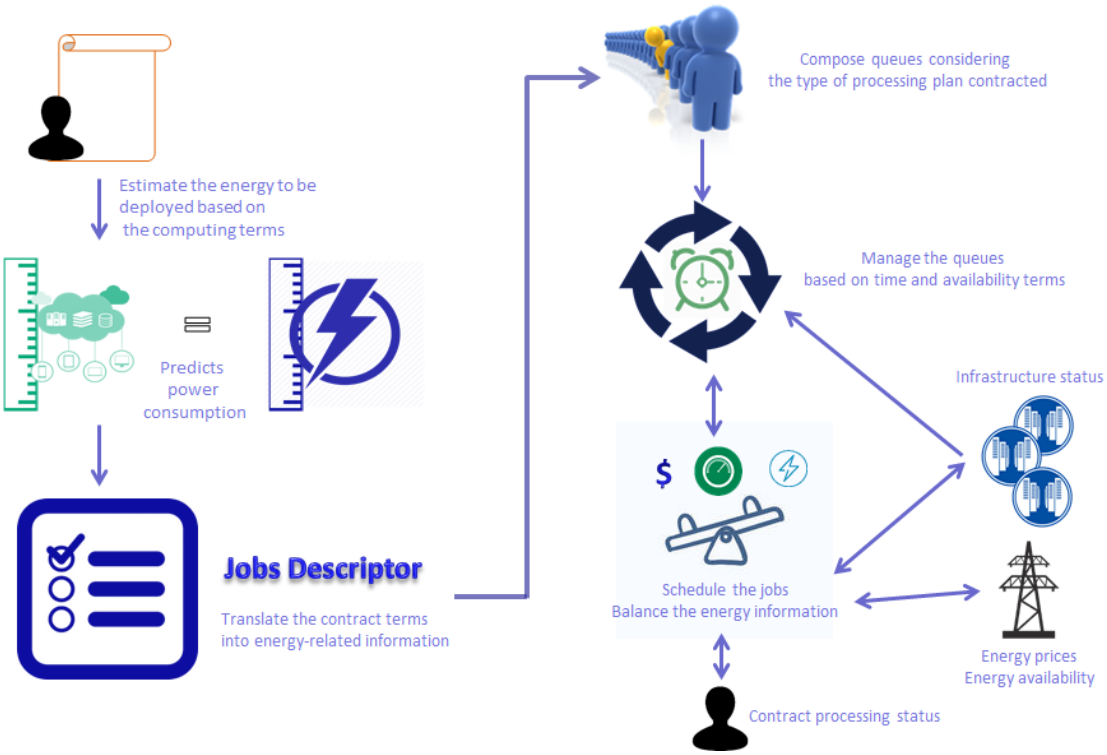


Figure 5 shows the contracts assignment process from the exchanged information perspective. The first section of the figure represents the contracting of the services, as specified by the *Contract requirements module*. The information relates to the contractible energy-service terms required to negotiate and manage the allocation during the stated period. The *Energy Prediction Module* calculates the amount of energy needed to run the contracts, as the power model determined. The contracting stage ends with the description of the contracts as jobs, which defines the energy service terms deployed to set the allocation. The module *Contract Metrics* represents the jobs composing step.

The next share of the figure exhibits the management of the contracts into queues, according to the modules *Queues Composing* and *Queues Management* specificities. As determined, the jobs are organized by processing plan, and the information regarding the electrical energy availability, prices, and the resources usage enables the administration of the jobs by time

periods and quality restraints.

The *Jobs Scheduler* and *EaaS* modules balance the infrastructure usage levels and energy-related information published to manage the demanded services. The former schedules the jobs and the latter assigns them to the physical allocation according to the energy consumption strategy defined by the Cloud Computing service provider. The communication exchange between the infrastructure and the queues management and the scheduler modules represents the renegotiation of the assignment.

The workflow enables to implement the provisioning and management of the electrical energy demanded by a Cloud Computing environment, according to the proposal and the EaaS strategies proposed. The stated modules set the conditions and prioritize the processing constraints according to the demand management more adequate for the environment. Also, it opens a communication exchange that enables to know the availability and energy usage levels.

4.4 CHAPTER CONSIDERATIONS

The chapter details the concept of the energy as a manageable infrastructure of the Cloud Computing environment. Considering the relative lower flexibility of the energy generation, in comparison to the computing resources contracting and provisioning, a management strategy from the electricity expenditure and purchasing brings a new perspective to deal with the workload.

From the energy consumption and supply fluctuation during the settled period, the Cloud Computing deals with the contracted demand to assign the services. New contract constraints allow the allocation of the contracts, through a demand-side management approach. Due to this fact, three different processing plans set costs levels and time requirements for the processing.

Although the energy efficiency is a relevant matter to Cloud Computing providers, there is no acknowledgment of providers that provide such contracting terms. Along with other energy service terms, the services provider may schedule the services according to energy requirements.

Therefore, an interface to understand the demand of energy, the market information and the expenditure of the capacity is required to develop the electricity as a compliant resource of the Cloud Computing. EaaS, introduced as a new service layer for the environment, deals with the different information and metrics; the service layer has the full knowledge of the electricity demanded to maintain the services functioning.

The described methodology to cope with information related to the energy consumption is present at this chapter, along with a workflow that enables to develop the resources mana-

gement. The set of workflow modules and the management approach proposed to establish an energy-aware Cloud Computing environment, named E2C. The E2C copes with the electrical energy deployment to determine the most beneficial strategy to deal with contracts and infrastructure provisioning.

5 ENERGY MANAGEMENT MODEL APPLICATION AND RESULTS

Chapter 5 presents the model designed to prove the management proposed. Section 5.1 details the workflow implementation and the Use Case scenario defined. The achieved results are shown in section 5.2. The section details the results for two different scenarios: an energy-sufficient and a scenario that demands energy purchase. Six different Test Cases were considered for each scenario; this section presents the graphs of the first Test Case, and the remaining are shown in appendixes A and B.

5.1 GREATER SÃO PAULO USE CASE

The workflow implementation takes into account DCs distributed in the Greater São Paulo area. The scenario Use Case considers a fifth generation (5G) scenario, with smaller size DCs spread through the metropolitan region. Smaller DCs are chosen due to the lack of space availability in greatest urban centers and related problems to maintain larger DCs (GELENBE, 2012).

The smaller sized DCs allows distributing the capacity, both computing and energy resources, over a large number of reduced sites. Also, the distributed DCs enables to operate by the fluctuation of self-generated energy supply or different energy prices. A management system controls the energy provisioning, rather than guaranteeing a high level of power availability all the time. The smaller DCs are deployed for a 5G scenario reducing the distance between users and infrastructure, and decreasing the access speed.

The model is a combination of CORD (Central Office Re-architected as a Datacenter) architecture and computes power installed in a central office location and geographical location data for main central-offices in Europe, from the FP7 COMBO (COnvergence of fixed and Mobile BrOadband access/aggregation networks) project (CORD, 2015). The geographic characteristics from Europe are replicated to São Paulo Metropolitan Area due to its industrial plant, and population density variation throughout the area.

The numbers of DCs are defined based on population density for each city of area (ESTATíSTICA, 2016). The density classifies each city onto one of the categories "Ultra Dense-Urban," "Urban," "Suburban" and "Rural." The number of DCs is calculated by dividing the surface area of each city (METROPOLITANO, 2017) with the corresponding "Main CO area size" from the respective geo-data table (Table 20) of the CORD document (CORD, 2015).

The Use Case counts 392 DCs (n=392) distributed at the Greater São Paulo area. The computing capacity of the DCs does not vary, as well as the energy capacity. The DCs are managed by an orchestrator, which distributes the contracts internally.

To establish different self-generated levels of energy supply, the metro area is divided into three sub-regions, according to the geographical position of the cities. Tables 2, 3 and 4 shows the number of DCs defined for each city of the sub-region. The amount of self-generated energy, per sub-region, and the range of energy prices are informed to the management system. Both the referred information represents the energy sector in the Use Case.

Figure 6 exemplifies the São Paulo proposed use case. The Decision Point represents an energy management node for the Cloud Computing environment. The referred node deals with the assignment of the contracts according to established strategies and informs the DCs orchestrator of the most favorable assignment concerning the energy range. The node is aware of the quantity of processing jobs and amount of required power by each DC.

The primary objective is to prove the model capacity to assign the contracts based on the energy supply and prices through a 24-hour period. The designed model must be able to distribute and renegotiate the contracts assignment focusing the greater usage of the self-generated energy. At this point, the work does not concern on reducing the operational costs for the energy deployment.

A randomly generated table simulates the contracts. These data refer to the energy service terms contracting, respectively the *Contract Requirements* module. The energy service restraints refer to the type of energy source (renewable, non-renewable or indifferent), availability and quality of service provisioning, budget, time constraints, and type of computing resources to be provisioned. The time constraints refer to the initial and final processing hours, for both

Figure 6 - São Paulo Metropolitan Area Use Case

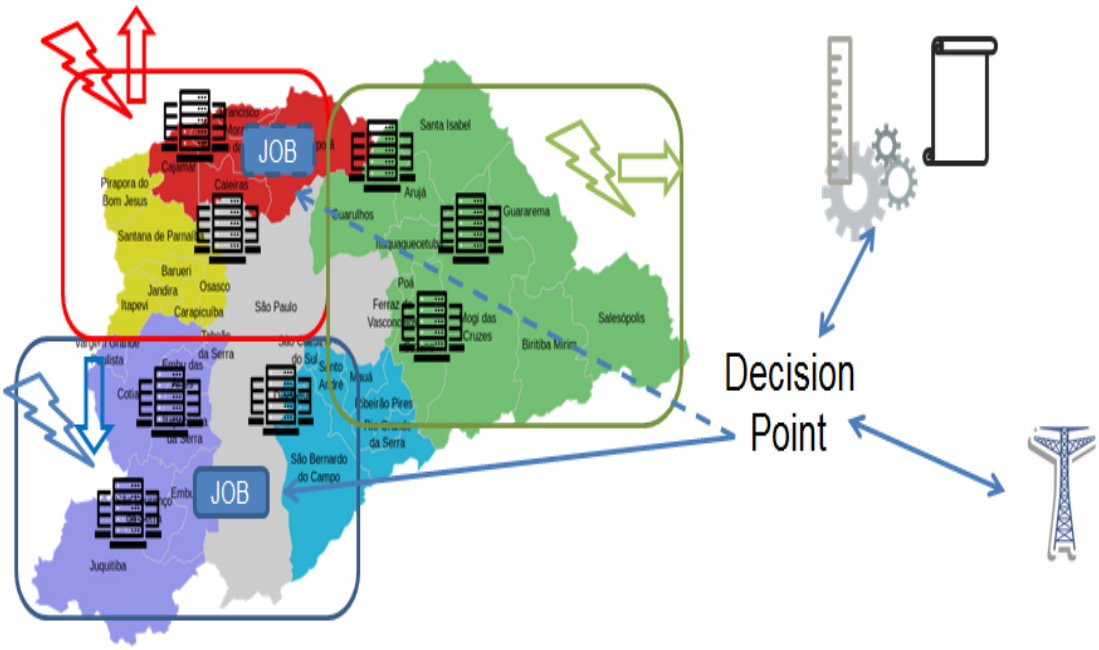


Table 2 - First sub-region defined for the Use Case

Sub-region - North, West and Southwest of São Paulo Metropolitan Area	
City	Number of DCs
Mairiporã	11
Francisco Morato	3
Franco da Rocha	6
Caieras	4
Cajamar	5
Pirapora do Bom Jesus	4
Santana de Parnaíba	7
Barueri	3
Jandira	1
Osasco	33
Carapicuíba	18
Itapevi	4
Vargem Grande Paulista	2
Cotia	12
Taboão da Serra	11
Embu das Artes	4
Itapecerica da Serra	7
São Lourenço da Serra	7
Embu-Guaçu	6
Juquitiba	17

on-demand and **reserved** plans, and latency, for three types of processing plans.

Four classifications for the computing resources are available to the user. These categories establish a range, from one to four, that refers to computing resources to be provisioned; the computing resources levels are named small, medium, large and extralarge. The level refers to the quantity of CPU usage, memory, storage, and network demanded the service functioning.

For each computing resources classification, there is a corresponding amount of energy defined. The correspondent power value replaces the power model stated for the *Energy Prediction* module. Based on the energy consumption of a high-performance server (SUPERMICRO, 2016), four values for the energy expenditure are estimated. The baseline of the energy consumption is the server in the idle state, which consumes P_{idle}=300 W (SUPERMICRO, 2016). The small level corresponds to P_{idle}+30%*P_{idle}, and the medium P_{idle}+75%*P_{idle},

Table 3 - Second sub-region defined for the Use Case

Sub-region - Southeast and city of São Paulo	
City	Number of DCs
São Paulo	67
São Caetano do Sul	8
Santo André	8
Mauá	3
Ribeirão Pires	4
Rio Grande da Serra	2
Diadema	16
São Bernardo do Campo	15

Table 4 - Third sub-region defined for the Use Case

Sub-region East	
City	Number of DCs
Guarulhos	14
Arujá	4
Itaquaquecetuba	4
Santa Isabel	12
Guararema	9
Poá	1
Ferraz de Vasconcelos	2
Suzano	8
Mogi das Cruzes	25
Biritiba Mirim	11
Salesópolis	14

large P_{idle}+150%*P_{idle} and extralarge accounts the maximum load, P_{max}=1000W, as the energy expenditure, respectively.

The management model sets energy service terms for each type of processing plan and organizes the referred information into queues; this stage corresponds to the *Energy Prediction* and *Queues Composing* modules of the management model. At the end of both steps, the contracted information describes the jobs and structures them into queues. The queues are organized according to the processing plan and time terms.

Due to restrictions, the established results evaluation deals with **on-demand** and **reserved**

plans similarly. These plans are differentiated along the determination of the allocation costs values, but the management deals with their defined queues setting and organizations equally. **Flexible** plans set the latency; the time interval helps to establish the costs during the settled stage.

After the queues organization, a budget prioritization sorts the queues. The method applies the budget as a strategy management item to take advantage of the most affordable energy available, considering a self-generation and different sources availability scenarios. Also, the deployment of the budget enables to establish a costs-allowance strategy, according to the dynamic pricing strategy proposed. These criteria are deployed as management strategies by the *Queues management* module of the workflow.

A fine tuning scenario based on the costs to process the jobs establishes the conditions for sequencing the queues assignment. Still, referring to the *Queues organization* module, the flexible queue is the first one allocated, followed by the **reserved** and **on-demand** ones. This management strategy enables to exploit to the maximum the self-generated energy, which is deployed as the cheapest energy available. Reserved queue is the next one managed for assignment, followed by the on-demand queue.

During the appointment process, the *Queues management* module values each job, according to the JobCost equation (1). This cost predicts the cost for the job assignment according to the energy price, availability and source, and the Service Level (S_L) of the contracted period. The energy availability defines the energy price for the allocation prediction; the model prioritizes the lowest budget contracts distribution by the self-generated electrical energy. In the case of the preferred energy ends, the method searches for the most affordable prices offered by the electricity sector.

The module varies the S_L according to the type of energy source and time interval. For the model development, S_L ranges from 0.1 to 1.0. To foster the **flexible** plans contracting, S_L has the lowest value possible for this plan; therefore, S_L for **on-demand** plans values 1.0. For **reserved plans**, the S_L ranges according to the minimum price predicted for the electrical energy, by source type, and the current price of the resource. The model calculates the JobCost value of the respective job.

After the costs calculus for each job, the contracts assignment is the further step. The jobs are reorganized by the budget and predicted costs, and the scheduling step determines which is the most favorable DC. Regarding the *Jobs Scheduler* module, the **flexible**, **reserved** and **on-demand** plans, orderly manner, are allocated.

The module deploys the Energy Cost (2) to determine the cost for each DC of the environment. At this stage, the module adopts the prices provided by the electricity sector to set the value for each DC, and it assumes that all the energy deployed is purchased. Also,

the model does not set a group of DCs to buy only renewable or non-renewable source; the module assures the purchasing of the electricity amount to power the service, satisfying the energy service terms contracted. Therefore, all the DCs are eligible for the allocation.

While there is extra self-generated energy, the jobs take advantage of the lowest related costs and are scheduled to be powered by this type of resource. The scheduler module searches for the lowest budget jobs to assign during this stated situation; despite the fact that the lowest budget jobs are prioritized during this stage, the module searches the DC with the lowest cost to assign the job. The concurrency between the DCs does not consider the internal computing or electrical energy capacity.

After the module sets the costs for each DC, it assigns the job and updates the queue and DCs information regarding jobs allocation. The amount of electrical energy to be provided to guarantee the job processing is summed for each DC. If the required energy must be purchased from the energy market, the corresponding amount is pointed by the model. The amount of energy at the DC and the indication of the acquired resource to maintain the infrastructure functioning are equivalent to the *Cloud Orchestrator* module, i.e., it emulates the physical appointment of the jobs.

At this stage, the model does not establish the *EaaS* as a single stage. The electricity sector is represented by the estimated renewable and non-renewable energy sources prices, stated for each hour of a 24-hour period, and the predicted amount of self-generated energy. Since the physical capacity of each DC is not deployed at this stage of the model development, by simplicity, the information exchanged between the DCs and the *EaaS* is represented by the assessment of the jobs.

At the end of the allocation step, the expected result is that the Decision Point allocates every job to a physical distribution, by the deployment of a cost-comparison strategy. As a second stage of the assignment stage, a renegotiation of the jobs is a further phase.

For the renegotiation, the Energy Management module applies a fluctuation of the self-generated electrical energy and electricity market broadcast prices to search for a different time interval or most favorable DC to process the jobs. The renegotiation stage regards to the *Queues Management*,*Jobs Scheduler* and*Eaas* modules of the E2C.

Initially, the module sorts all the queues by the availability and budget service terms. The model searches for each job nominated to be re-scheduled or migrated to a most affordable DC to allocate it. The Decision Point manages the queues by each previously assigned DC, i. e., the management module searches for DCs located in regions with more energy available. Another option is rescheduling if it is a beneficial decision for both user and services provider. If the job is willing to alter its service terms regarding the best allocation, from an energy deployment perspective, the module determines the JobCost again and compares with the

previous one. Also, the module determines the new cost to re-schedule the job or to migrate it, according to the contracted availability term.

If the difference between the previously assigned cost and the current one decreases, the job is re-scheduled. The modifications concern the current value for each DC, provided by the Energy Cost (2) cost. The new energy is summed for the respective DC and update the energy purchasing data.

This step simulates the migration and the renegotiation of the jobs for the Energy Management Model. The expected result of this stage is the reallocation of jobs by the electrical energy variation from one region to another. The final costs to the service provider may not be reduced, as in cases that must of the jobs requires processing during costly intervals and does not accept any reallocation proposed. Due to this fact, the primary purpose of the Use Case is to demonstrate the management of contracts regarding electrical energy available and related costs fluctuation.

Two different electricity management scenarios were developed to show this purpose. The Use Case defines 3000 randomly generated contracts to simulate the contracts requirement step. For both the scenarios, an allocation stage considers a past generation prediction to determine the costs and search for the most favorable assignment. Also, the stages deploy energy prices fluctuation to state the most beneficial allocation of the jobs.

Test Cases simulate the contracts input, distribution of the jobs and their renegotiation during the current day. Six different test cases, for each scenario, were considered. The contracts for each test case vary, thus, there is not an equal amount of processing plans. The information taken for each test case is jobs status, according to the processing plans, energy, and jobs distribution, and the amount of energy migrated from one sub-region to another. The information obtained for allocation and renegotiation steps.

The first stage, the allocation one, renegotiates the contracts distribution considering a current energy supply data. The historical prediction data refers to a wind power in Europe, during a random day in December 2016, according to the Nord Pool Power Market (MARKET, 2017). The electrical energy for the renegotiation stage was generated randomly, considering power ranges from 2000 to 7000 W. Sub-regions 1 and 3 have the lowest generation range, while sub-region 2 has the greatest amount of energy generated.

The prices for both stages do not alter. Based on a Time of Use (ToU) program, the Use Case model considers day-ahead announced prices to take decisions regarding the services allocation. The Nord Pool Market (MARKET, 2017) publicizes the prices for the energy purchasing for the next day. However, the prices do not establish the type of power source; due to this fact, the prices are composed of two different ranges, considering a renewable energy source and non-renewable energy source, without type specifying.

The first scenario takes into account that the Cloud Computing environment is self-sufficient in generating the energy to maintain the DCs functioning. The autonomous scenario, for both stages, deploys the generated energy to assign the jobs, and the energy purchasing is not required. The second scenario considers the amount of self-generated energy to assign the jobs, but an amount of energy acquisition is demanded to power the infrastructure.

5.2 RESULTS

The results obtained from the designed model shows the contracts allocation from the energy supply available during a 24-hour period (T=24). The quantity of self-generated energy provides the information necessary to distribute the jobs and to set the amount of resource to be purchased from the energy market.

The Use Case assigns 3000 contracts (C=3000) in 392 DCs(DC=392). The contracts are distributed hourly (t=0,1,2,...,24) for the stated period, and the accountability is done by the period. The terms of the contract terms are randomly generated, which enables to show the management of different contract profiles for the established use case.

The value of the self-generated energy is fixed during the allocation stage but is randomly generated during the re-scheduling stage. Different price values for the non-renewable and renewable energy sources are stated for each interval; the prices do not alter from one stage to another. The jobs assignment accounts the quantity of jobs allocated (Job) and the amount of electrical energy (JobEnergy) required for the interval for each DC of the use case.

The first scenario considers that the predicted self-generated energy is enough to power the infrastructure spread on the three sub-regions. The contracts renegotiation stage searches for the sub-regions or time periods with the greater amount of electricity available.

The second scenario manages the energy from a non-sufficient amount of self-generated energy and predicts the volume of electricity to be purchased from the market. During the renegotiation stage, the stated model takes into account the current self-generated energy to reallocate the jobs, from the resources availability perspective. The re-scheduling of the contracts does not imply that the required energy is not purchased anymore.

The next subsections show the results obtained for both the proposed scenarios. The findings demonstrate the distribution according to the processing plans, and the *JobEnergy* and jobs allocated for each region. For the step that demands to purchase of the resource, the amount of energy obtained is shown as well.

5.2.1 Powered by self-generated energy

The initial analysis of the management considered that the self-generated energy was sufficient to power the set of DCs. Six different test cases were applied at this stage; the test cases were generated from the C=3000 contracts randomly generated. Table 5 shows the number of processing plans accounted for each test case created.

At this stage, the cost of the electrical energy does not change during the period. Value to acquire the equipment to produce the energy supply and to maintain its functioning is applied as its purchasing price; the price does not vary during the stated 24-hour. During the whole allocation period, the cost to allocate the job, determined by the *JobCost* (1) equation, values the energy price ($p(t)$) as a constant.

Therefore, the *JobCost* and the budget values are used to set a prioritization condition of the jobs assigned. Both values are deployed at the renegotiation stage to test if the rescheduling or migration proposed presents an advantage for the user and service provider. The infrastructure cost (EC), determined by equation 2, is deployed to differentiate the DCs stipulated for each region.

At the end of the prediction stage, the assignment of the **flexible** processing plans must follow the availability of the energy source foreseen. In the case of electrical energy purchasing requirement, this value must be pointed by the management model. It is assumed that the energy and jobs distribution must follow the generation peak during the day.

The renegotiation stage must search for most favorable allocation. At the end of this step, it is expected that the jobs reassignment according to energy supply peaks.

Once the number of stipulated jobs does not differ significantly from one test case to another, and the number of jobs does not vary from one processing plan to another, the predicted energy and jobs distribution graphs are very similar. Also, the fact that the foreseen self-generated energy is the same for all the Test Cases justifies the significant similarity between the graphs obtained for the prediction stage.

Figure 7 shows the energy expenditure predicted for each time interval of the period. The generation growing starts at 7 am, and its peak occurs during 9 am to 11 am, the management allocates the flexible plans during this period. Accompanying the energy prediction for the 24-hour period, the number of jobs for each hour, shown in figure 8, reaches the peak at 10 am.

The energy management set an approach to reallocate the jobs according to current values of the self-generated energy as a second stage. Table 6 contains the amount of migrated job, according to the contracted processing plan. The referred migration numbers show contracts that accept any renegotiation, considering the processing rescheduling and the physical

Table 5 - Quantity of randomly generated jobs for the self-generated energy scenario

Test X Type of processing plan	Reserved jobs	On-demand jobs	Flexible jobs
Test case 1	1005	982	1013
Test case 2	989	1037	974
Test case 3	967	1015	1018
Test case 4	1050	994	956
Test case 5	1023	947	1030
Test case 6	986	998	1016

infrastructure migration.

As the majority of the generated contracts allows any reallocation predicted, the number of

Figure 7 - 24-hour energy distribution for the prediction stage - Test Case 1

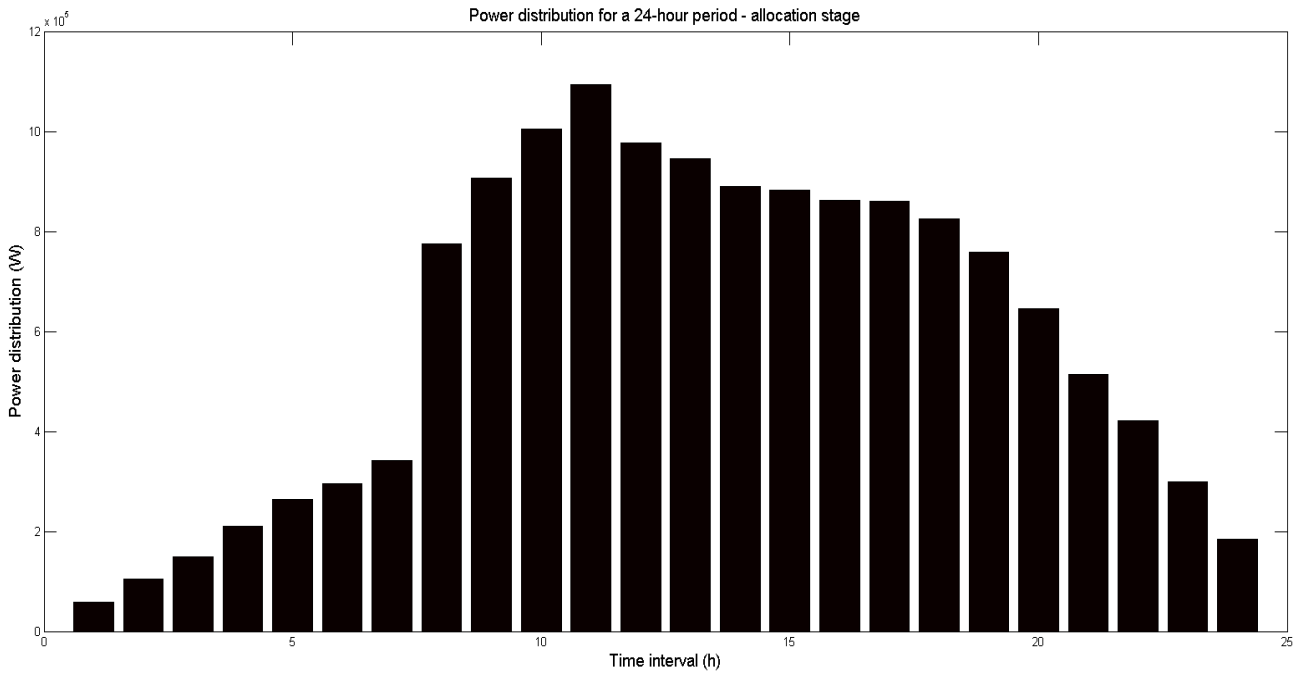
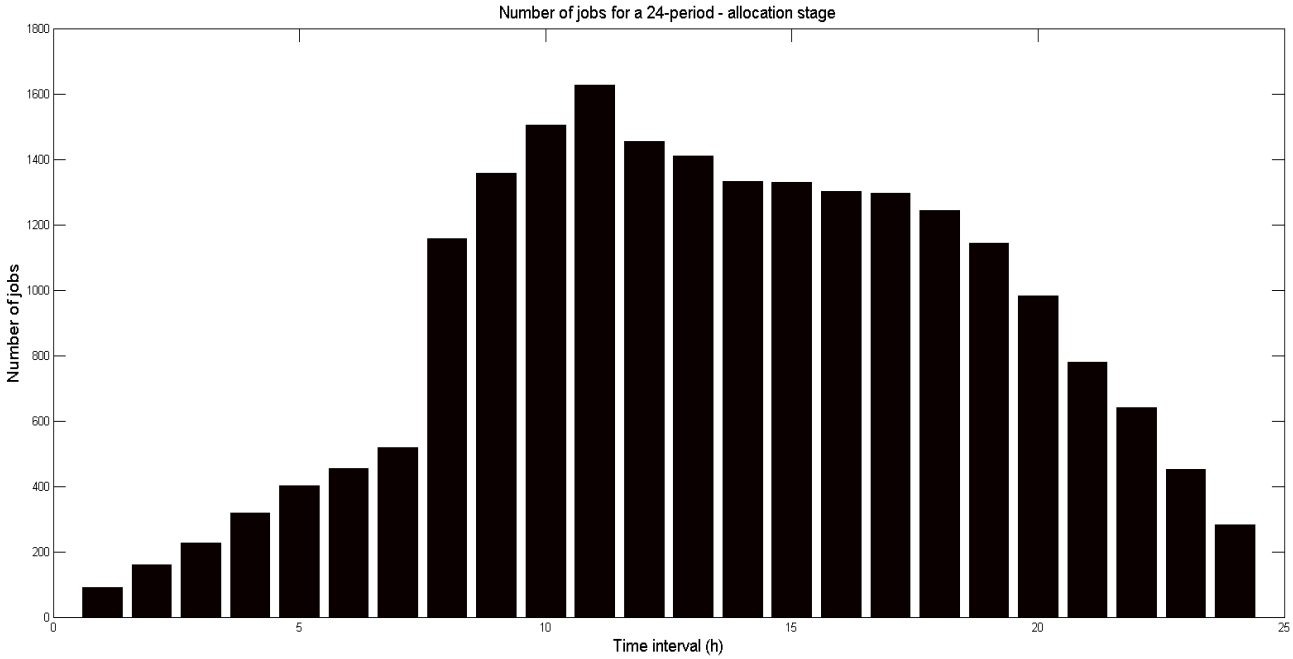


Figure 8 - 24-hour jobs assignment for the prediction stage - Test Case 1



jobs that are rescheduled or migrates from one region to another is significant. The fact that the amount of self-generated energy at this stage is sufficient to power the infrastructure, the management approach searches for areas with most energy availability to allocate the jobs.

Table 6 - Quantity of migrated jobs for each proposed processing plan

Reserved migrated jobs	On-demand migrated jobs	Flexible migrated jobs	Sum of the migrated jobs
761	982	1012	2755
727	1037	847	2611
729	1015	1018	2762
777	994	932	2703
748	947	1012	2707
728	998	987	2713

In comparison to the energy expenditure predicted, the current energy distribution has the consumption peak altered. Also, the amount of electrical consumption varies from one test case to another, as exhibited in figure 9. The difference between the figures is because the values of the current self-generated electrical energy change for each test case, which enables that each case exhibits a different reallocation scenario.

The electrical energy consumption expected for each region depends on the amount of

Figure 9 - 24-hour energy distribution for the negotiation stage - Test Case 1

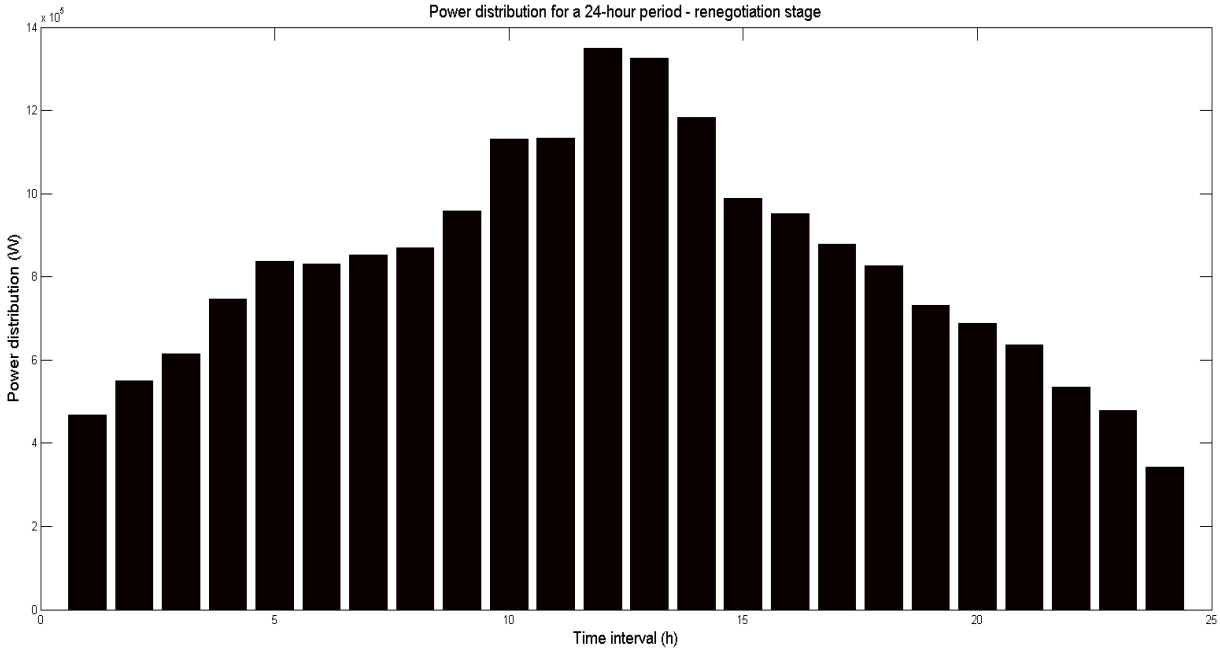
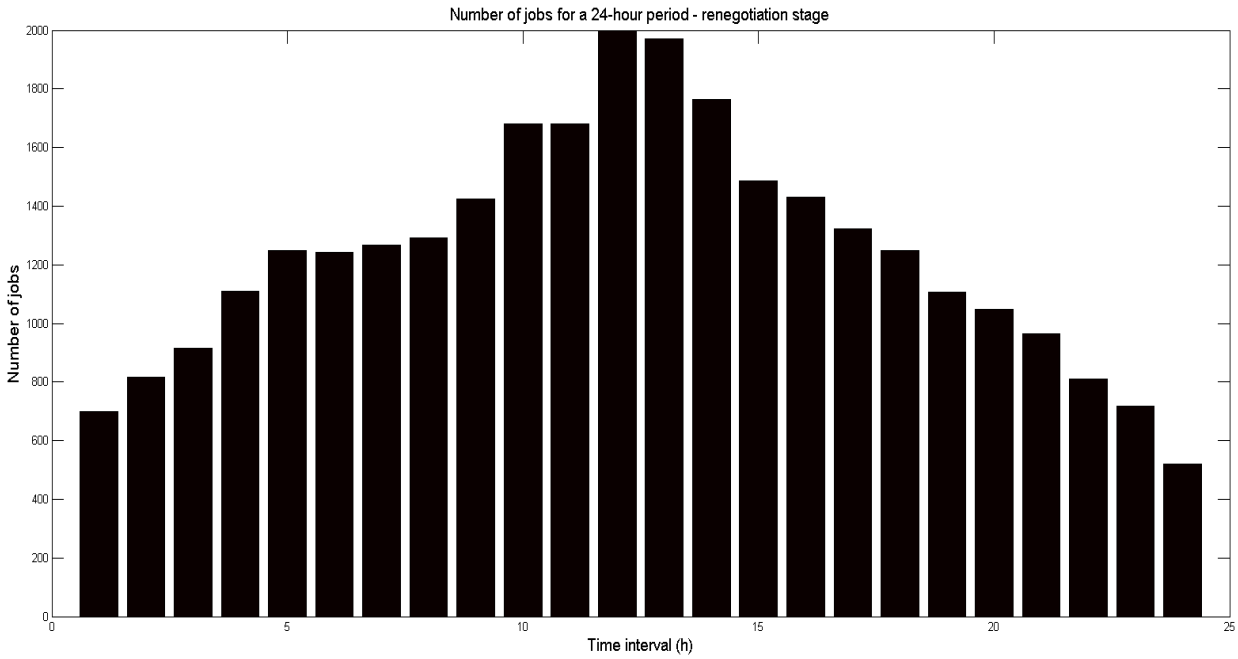


Figure 10 - 24-hour jobs assignment for the negotiation stage - Test Case 1

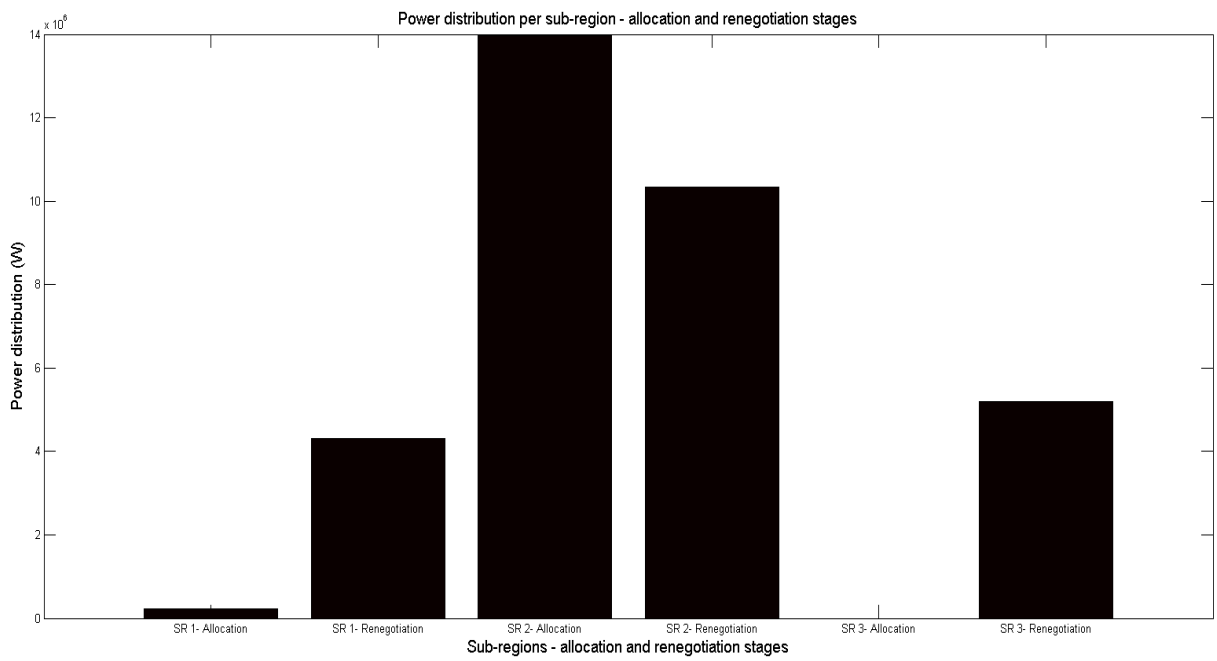


resource generated during the period. Region 2, the one that generates most of the energy supply, allocates most of the jobs, as seen in figure 12. The difference of the jobs allocated by region enables to see which region provides more energy to power the DCs but not shows

the area with the most computing capacity available.

The difference between the values of the predicted and the currently generated energy justifies the difference of the expenditure number presented by the graphs that account the consumption per region. Figure 11 enable to verify the fluctuation of the foreseen and the current value of the generated electricity.

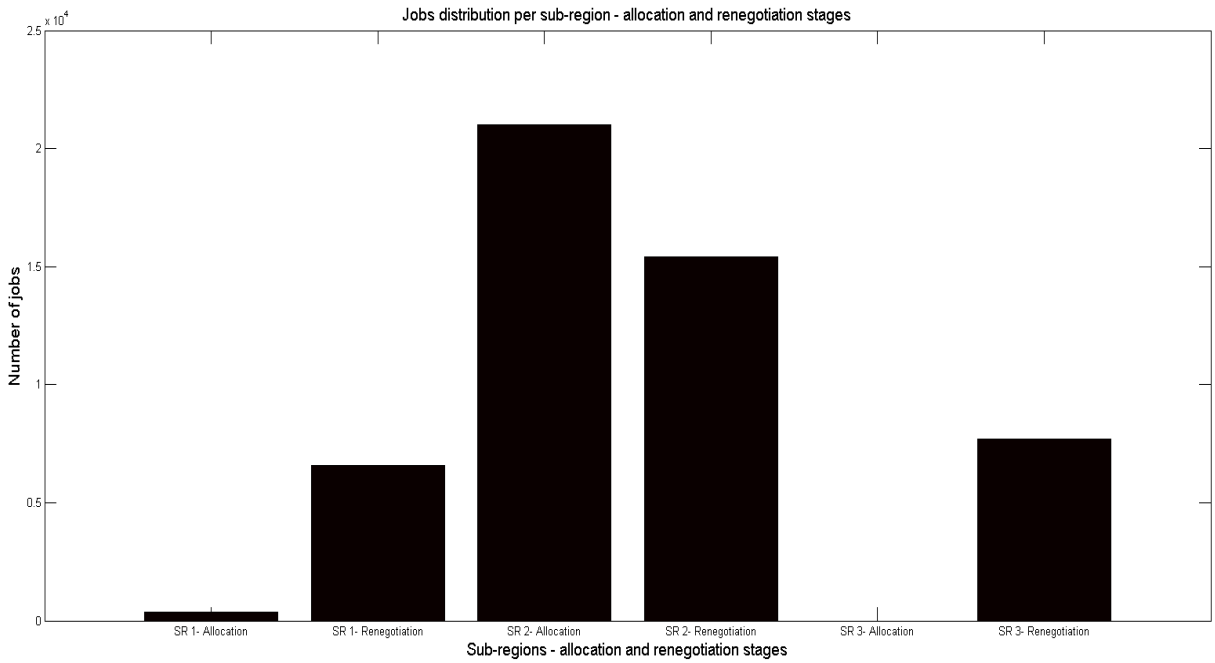
Figure 11 - Energy expenditure estimated for each sub-region, for prediction and negotiation stages - Test Case 1



For all the test cases, there was no need to purchase energy. The self-generated energy allowed to power the contracted services independently. Despite this fact, the energy management approach searches for opportunities to cope with the jobs regarding the major availability of the electricity, from one region to another. Also, the proposed model renegotiates the services favoring the costs decrease, for both user and service provider.

The method dynamically recalculates the new allocation, regarding the energy consumption of each DC. The current generated electrical energy is applied to search for better distribution opportunities, from the availability perspective. At this point, the calculated *JobCost* for each contract enables to differentiate the jobs and to set costs conditions for the renegotiation stage.

Figure 12 - Jobs assignment for each sub-region, for prediction and negotiation stages - Test Case 1



5.2.2 Powered by self-generated and purchased energy

This stage takes into account an amount of self-generated energy by the infrastructure. The quantity of the generated electricity, however, is not enough to power the services assignment, which implies that the service provider must purchase energy from the market to maintain the infrastructure functioning.

For both the prediction and the renegotiation stages, this scenario exists. Since the script demands that more electricity must be bought to sustain the business operation, the stated *JobCost* determines the cost for each job and the most favorable difference between this value and the contracted budget sets which job is allocated with the self-generated energy and which ones are powered by the purchased energy.

The *JobCost* is deployed as a prioritization weight by each processing plan. The price *p(t* of the self-generated energy does not fluctuate during the 24-hour period; however, when the purchasing of the electricity is required, the prices published by the market are deployed to calculate the *JobCost*.

The value of the self-generated energy during the prediction stage is the same for all the test cases stated. It is expected that the energy management approach allocates the flexible plans according to the peak of the self-generated energy, reducing the costs for the Cloud Computing

services provider. For the other plans, at this stage, the contracted time constraints are not altered.

For the renegotiation stage, the energy management approach renegotiates the allocation regarding the energy availability. The primary purpose is to reduce the to-be-purchased energy by the deployment of the quality terms contracted. In cases that the self-generated electrical energy is not enough to power the services during the stated interval, the model searches for opportunities to turn the allocation cheaper.

For both the stages, a range of prices for the electricity simulates the role of the energy market. The prices are established for each hour of the period and sets two different levels, one for purchasing renewable energy source and other for a non-renewable source. The range of the prices is the same for the prediction and renegotiation stages. The number of jobs for each test case proposed is detailed at Table 7.

Table 7 - Quantity of randomly generated jobs for the self-generated energy scenario

Test X Type of processing plan	Reserved jobs	On-demand jobs	Flexible jobs
Test case 1	1003	979	1018
Test case 2	1019	988	993
Test case 3	1016	925	1059
Test case 4	995	1013	992
Test case 5	987	990	1023
Test case 6	1015	989	996

The values of the energy expenditure foreseen for each time interval of the stated period do not vary from one test case to another. As the peak of the self-generated energy occurs during the 9 am, and 10 am intervals, the majority of the flexible plans are allocated during this period. However, the predicted consumption peak does not occur at this time range.

For the prediction stage, the peak occurs at the range of the 10 am to 1 pm. The peak happens because the self-generated energy ends and the energy management approach allocate the flexible processing plans during the hour with the most affordable price of the electricity.

The consumption predicted for the stated period is shown in figure 13. The difference of the summed energy for each hour is not visible in these graphs.

A number of jobs distributed at the 24-hour period, as seen in figure 14, follows the energy consumption. Since the amount of self-generated energy is low, in comparison to the number of contracted jobs, the foreseen purchased energy values are similar to the predicted energy consumption numbers.

As shown in figure 15, the difference between the values of the purchasing energy and its consumption are similar. Therefore, at the end of the prediction stage, the energy model proposed enables the Cloud Computing services provider to estimate the amount of energy to be purchased and its relative costs.

Figure 13 - 24-hour energy distribution for the prediction stage - Test Case 1

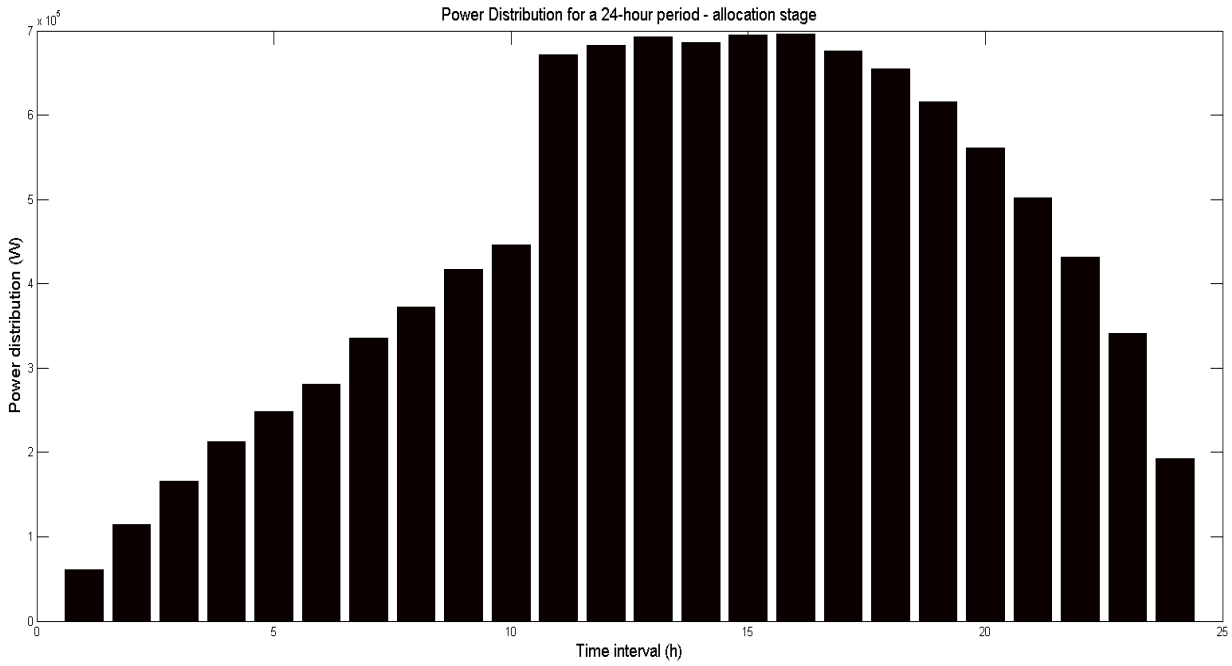


Table 8 shows the number of jobs that are re-scheduled or migrate to a different physical infrastructure, under the quality contracted terms and current self-generated electrical energy. The range of prices that simulate the energy market does not change from the prediction stage to the renegotiation stage. Since the availability of the self-generated energy is reduced, the negotiation for a more affordable allocation is restricted. Due to this fact, the quantity of migrated jobs, when compared to the previous scenario, is lower.

Figure 16 displays the electrical energy consumption for the period after the migration of the jobs. Despite the fact that the values for the self-generated energy change from one test case to another, the amount of the energy expenditure during the day presents few variations.

Figure 14 - 24-hour jobs assignment for the prediction stage - Test Case 1

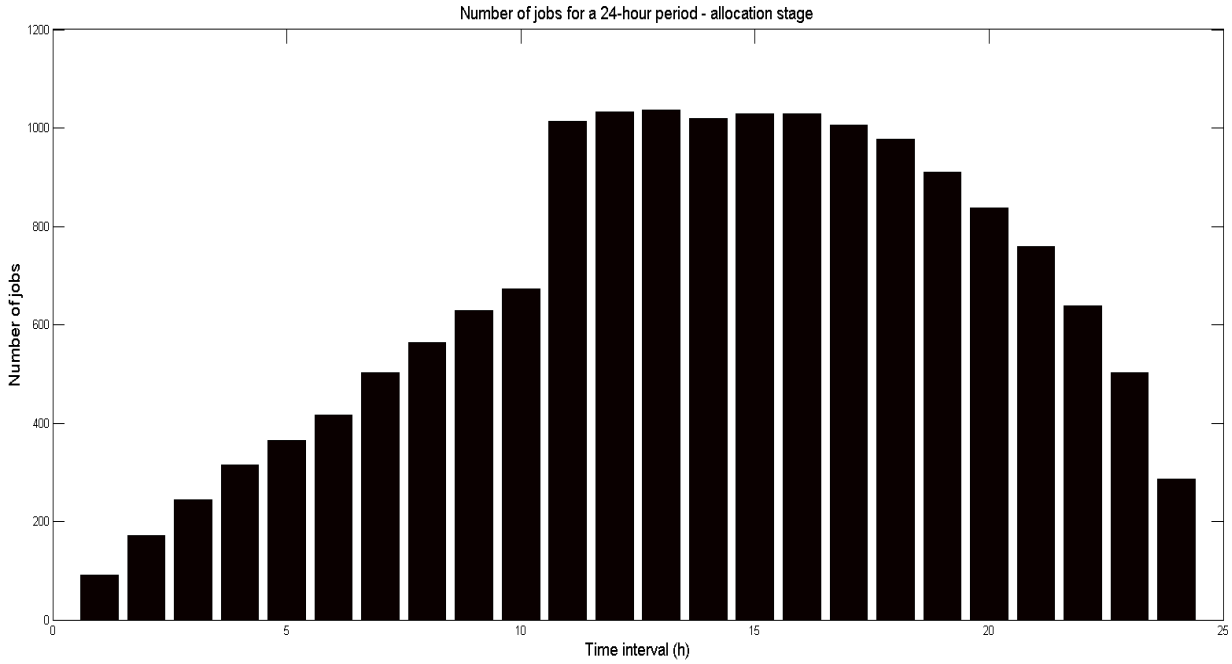
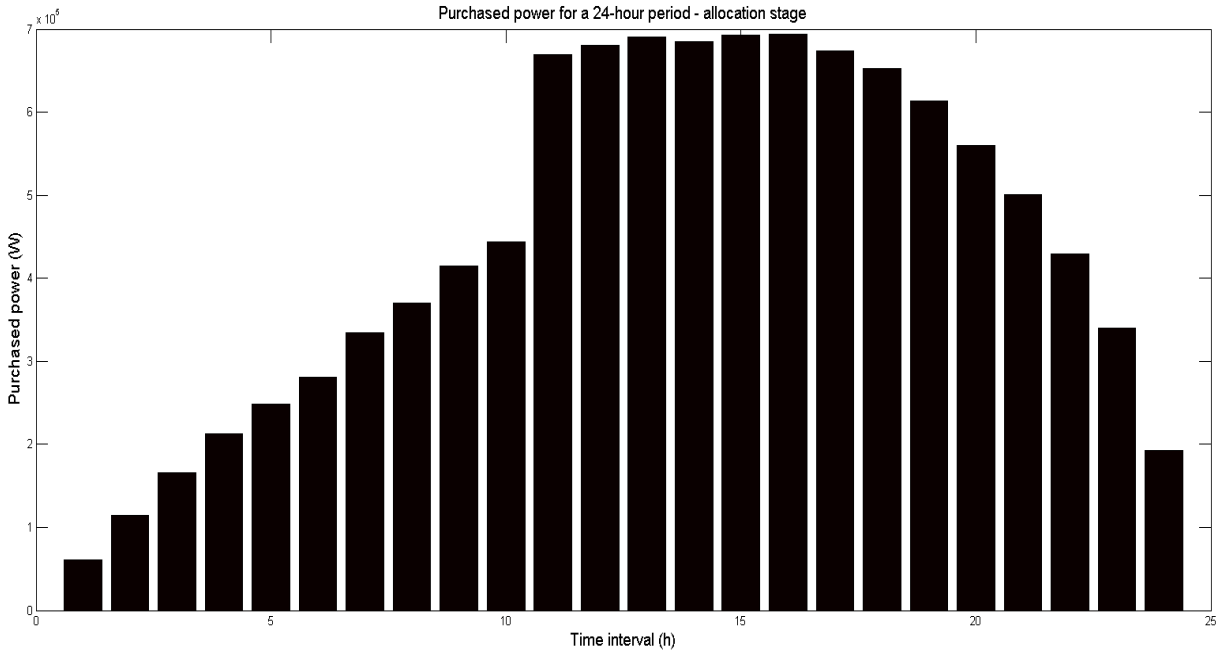


Figure 15 - Expected acquired energy for a 24-hour period, prediction stage - Test Case 1



The peak of the energy expenditure occurs during the initial hours of the day. As the most affordable hour for the allocation of the jobs occurs during the early hours, according to the appointed prices, during the renegotiation stage, the method assigns the jobs during this time.

Table 8 - Quantity of migrated jobs for each proposed processing plan

Reserved jobs	On-demand jobs	Flexible jobs	Total migrated jobs
15	34	932	981
120	29	947	1096
36	27	974	1037
142	12	960	1114
36	34	951	1021
36	33	914	983

The jobs distribution, for a 24-hour period, is shown in figure 17.

Figure 16 - 24-hour energy distribution for the negotiation stage - Test Case 1

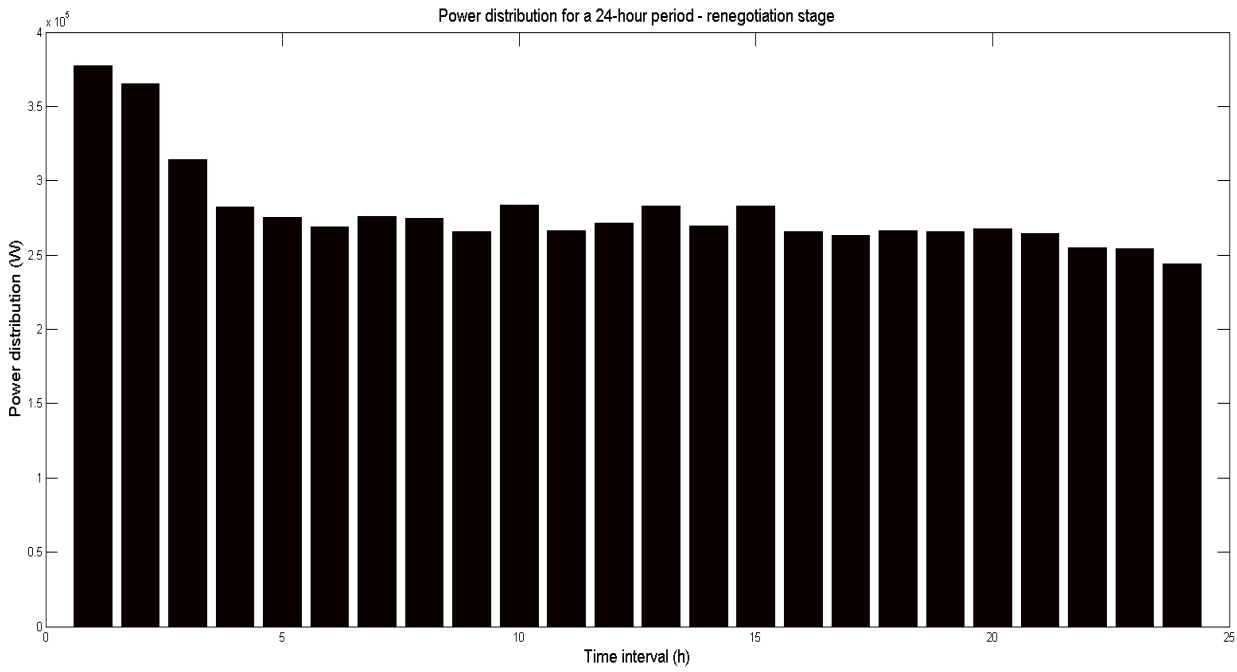
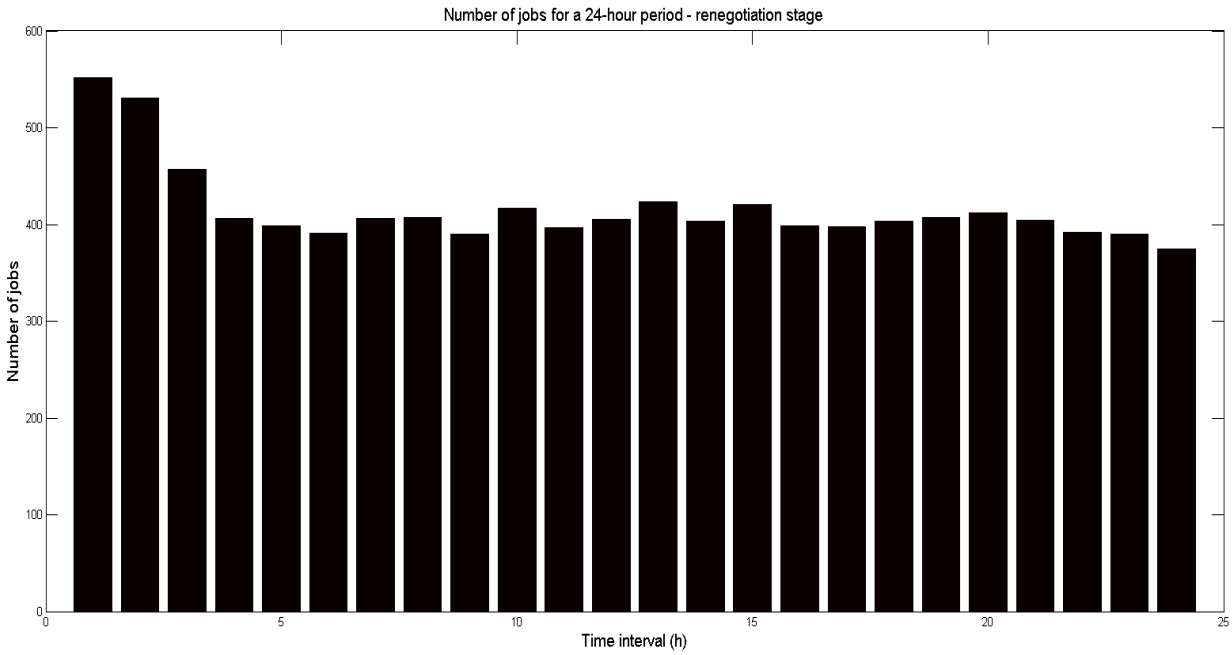


Figure 18 shows the quantity of purchased energy for the period. For all the cases, the peak of the purchasing energy occurs during the initial hours, accompanying the energy distribution shown previously.

After the renegotiation step, the required energy purchased is lower. Since the self-generation supply is higher than at the previous stage, the amount of purchased power reduces. However, after the modification of the processing period or area, the greater quantity of power is obtained during the first time interval of the day. The acquiring interval occurs since this time interval is the cheapest determined by the electricity prices; therefore, the flexible and reserved

Figure 17 - 24-hour jobs assignment for the negotiation stage - Test Case 1



plans renegotiated are re-scheduled for this period.

The migration incentive occurs due to self-generated energy range, the number of jobs migrating is reduced, in comparison to the first scenario proposed. After the self-generated energy ends, the contracted jobs search for the most low-cost allocations. As this new allocation also depends on the availability term contracted, the energy distribution for each region is lower. Figures 19 and 20 show the distribution for each stated sub-region.

Figure 18 - Expected acquired energy for a 24-hour period, prediction stage - Test Case 1

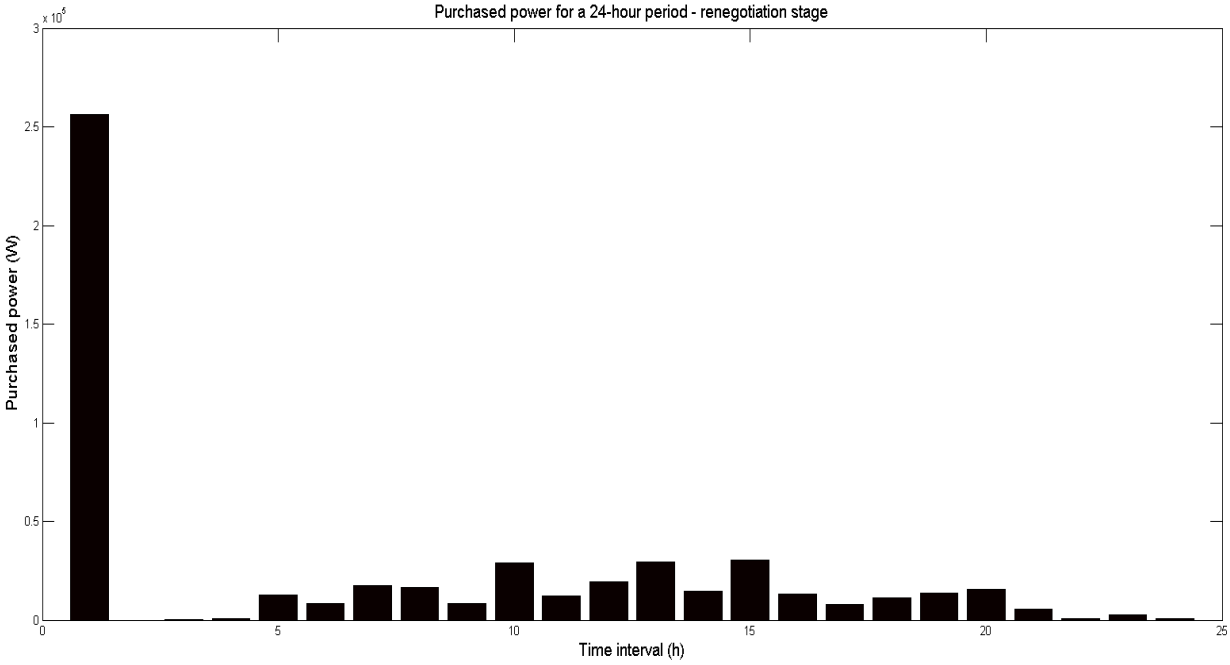


Figure 19 - Energy expenditure estimated for each sub-region, prediction and negotiation stages - Test Case 1

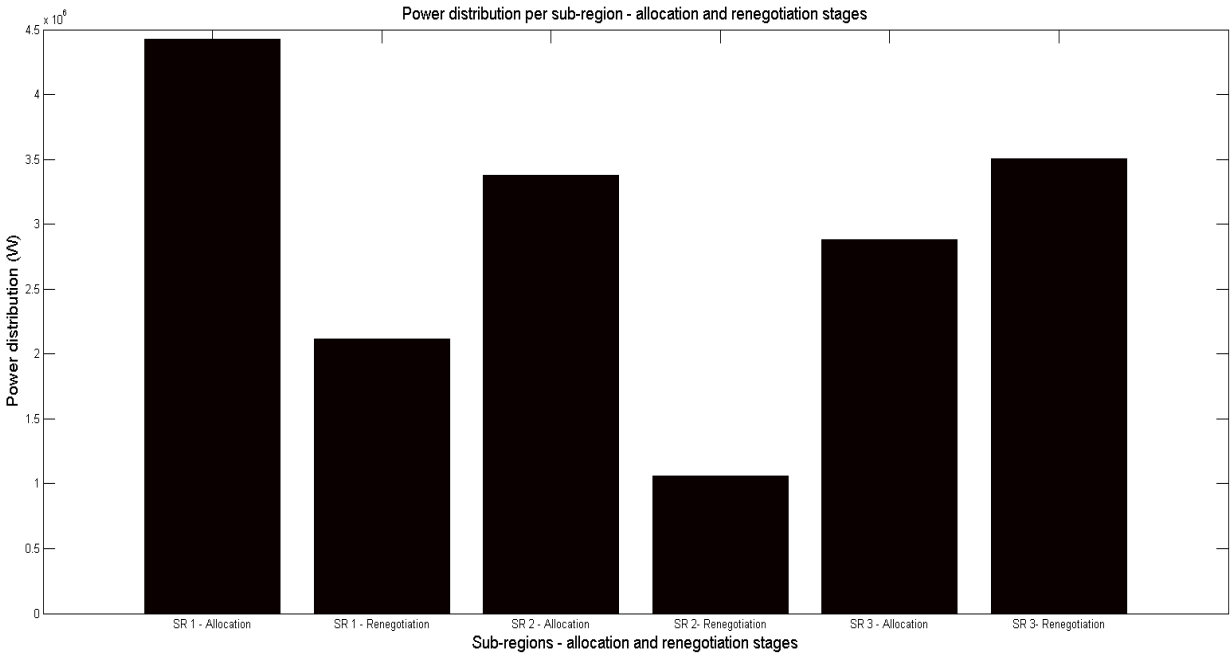
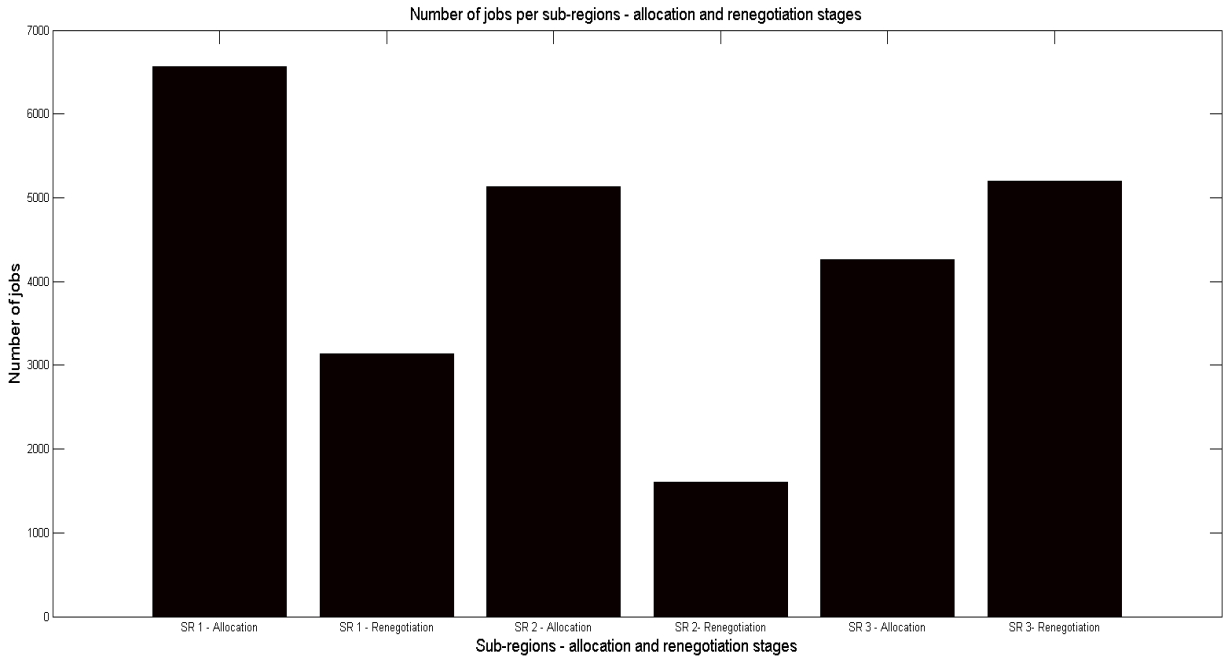


Figure 20 - Jobs assignment for each sub-region, prediction and negotiation stages - Test Case 1



5.3 CHAPTER CONSIDERATIONS

The developed Use Case simulates an infrastructure allocation closer to a real scenario. The SP Use Case establishes an apportionment of the infrastructure according to the population range perspective, differentiating the capacity and setting the quantity of DCs. Although the Use Case scenarios do not take into account the capacity factor, the greater amount of DCs enables to set electricity generation and prices scenarios.

The settlement of the region into three sub-regions allows developing different quantities of self-generated electrical energy, which helped to show the proposed management from various scenarios. The self-generated supply enabled to set conditions to allocate and renegotiate the contracts assignment, including this type of energy as a negotiable term to the processing plans. Therefore, the plans were not just allocated from the prices perspective, but the energy supply was deployed as a weight factor.

The results acquired for the method application demonstrates the management of the workload for two different scenarios. For both scenarios, the obtained graphs show that the proposed method searches for the most favorable assignment, deploying an infrastructure costs weight and energy availability terms. Also, the method adjusts itself the distribution, comparing two different generation ranges for the electricity.

6 FINAL CONSIDERATIONS

The objective of this work is to develop an energy management approach for Cloud Computing environment that enables to combine the computing infrastructure usage, energy prices and availability to set negotiation terms in the energy capacity allocation contracts.

The energy prices and availability through time periods allow setting a management strategy for the Cloud Computing environment focused on the resource deployment. Information regarding the infrastructure and the referred energy terms are balanced to establish a strategy to cope with the infrastructure provisioning. Along with these conditions, the contracted terms must be fulfilled.

The developed strategy implements a demand management, from the service provider perspective, taking into account the computing infrastructure and energy terms. This approach considers:

- Energy contractible constraints.
- Scheduling strategy for the services regarding the most favorable energy deployment perspective.
- Implementation regardless of specifying the energy market.

To show the contributions of this management proposal, this chapter reports the functional requirements, an analysis of the results, contributions and future works. Section 6.1 presents the defined requirements collected. The section compares the requirements fulfillment by the referred scheduling works and the present proposal. The requirements not attended are detailed as future works in section 6.4.

Section 6.2 details the results obtained in comparison to the intended objectives. The section 6.3 compares the targets and the contributions of the proposed approach.

6.1 REQUIREMENTS FULFILMENT

Table 9 shows the functional requirement listed in section 3.3 and those fulfilled by the proposed management strategy and the referenced scheduling methods. The energy management approach (Nascimento) met the requirements concerning prices and supply as incentives for the allocation (FR2, FR3). Therefore, the scheduling technique regarding the two terms, FR12, is accomplished as well.

To achieve the results of such a scheduling approach, the interface between the Cloud Computing and the energy sector (FR1) was achieved by this strategy. The interface between

Table 9 - Requirements Fulfilment by referred scheduling works and the present proposal

WorkX FR	Hsu et al.	Ren et al.	Wu et al.	Kim et al	Lucanin & Brandic	Minh & Samejima	Goiri et al.	Masker et al.	Nascimento
FR1					X				X
FR2	X					X	X	X	X
FR3									X
FR4									X
FR5	X	X	X	X		X			
FR6									X
FR7			X	X					
FR8						X			
FR9		X	X	X		X			X
FR10								X	
FR11									X
FR12									X
FR13	X		X			X	X	X	X
FR14	X	X		X		X			X
FR15	X	X		X		X			
FR16						X		X	X
FR17									
FR18									X

the energy sector and the service provider sets an information flow, but do not restrict the type of market that it is applied. Therefore, it is possible to define the requirement FR18 as fulfilled.

The knowledge of the energy sector and the infrastructure status (FR4) enabled the scheduling method to know the current deployment of the capacity, the allocation time requirements and the previously scheduled services. Due to this fact, the requirement FR 14 is considered as fulfilled by this work.

Besides, requirements relating to renewable supply (FR11), and setting terms that enable the user to contract the energy as services were fulfilled too (FR11, FR12). The readiness of the method to cope with the renewables variation through time intervals, according to the contracted claims, allows the scheduling of the services following the resources availability.

As contract claim sets the SLA and GreenSLA terms, both service levels guarantee are targeted during the services assignment. Both conditions are required by FR7 and FR9.

They relate to energy service conditions, and the maintenance of the quality of the services provisioning is achieved through a bilateral negotiation between users and services provider and established energy terms. To consider the energy efficiency as a factor in the scheduling decision, concerning to GreenSLA terms, requirements FR11 and FR13 are stated.

In a view to guaranteeing the SLA and GreenSLA tradeoffs, conditions related to the energy provisioning are offered as contract constraints for the user. Defining such terms as contractible items for the user fulfills the requirement FR6. Therefore, energy efficiency, energy supply and the quality of the services regarding power consumption are treated as contractible constraints and decision taking terms for the distribution. The negotiation of these terms establishes a dynamic pricing strategy, cited in FR10.

To increase the gains concerning the availability of the energy and its prices, the geographical allocation of the DCs is referred as functional requirements (FR8 and FR16). The full knowledge of the energy status, provided by an interface between Cloud Computing services provider and the energy sector (FR1), and the interface between the DCs and the provider (FR4) enables the fulfilling of other requirements.

The ability of the approach to encourage the user to change the demanded allocation regarding the energy status sets a technical solution for the DCs entrance on DRs programs. Along with the cited incentive, the definition of an interface that enables the Cloud Computing environment to understand the energy sector status allows the participation of the DCs on such programs. As established on FR3, the strategy allows the DR involvement of the ICT sector.

Requirements that address the services on the physical infrastructure, i.e., concerning the computing resources status are not accomplished. Requirements FR5 and FR15 that refers to the VMs management on the DCs were left for future works. The present strategy acts as an energy management node inside the whole Cloud Computing environment but, at this point, without interfering on the computing resources directly. Also, the requirement related to the reliability of the data exchanged between users and the services provider (FR17) is not accomplished. The system achieves the receipt and negotiation of contract terms, but do not specify security and reliability conditions.

6.2 RESULT ANALYSIS

The developed model for the energy management approach can distribute the contracts according to the energy availability and price variation. First, the model applies an energy generation prediction and the publicized prices to assign the jobs; after, a negotiation model considers the current generation of the resource to distribute the services. In this case, it

simulates a present situation to reschedule the services, taking into account the demanded quality of the service.

To demonstrate the fulfilling of the proposed requirements, two different scenarios were defined. For both, the model managed the jobs according to the significant quantity of the self-generated energy or rescheduled the greatest amount of the contracts.

For the scenario with sufficient energy to provide the services, the results demonstrate the capacity of the strategy to search for the largest quantity of available energy. For the allocation and the negotiation stages, the model distributes the contracted demand according to the processing plan and costs. Graphs that show the distribution of the jobs by region demonstrate the capacity of arranging the contracts per generated energy amount.

For the non-sufficient scenario, which requires an energy negotiation with the sector and the users, the results show the management of workload though time intervals and the three sub-regions. The achievements represent the distribution according to the greatest level of resource, but, also, it illustrates the rescheduling of contracts for the time intervals with the most affordable energy prices. Either way, the ability of management was demonstrated, showing its behavior through different periods and energy availability scenarios.

6.3 CONTRIBUTIONS

The main contribution of this work is the definition of an energy management approach for the Cloud Computing environment that negotiates the contracts allocation according to computing infrastructure provisioning and energy prices and supply conditions.

This strategy is sensitive to prices and energy load variation, and it enables to include the DCs in demand-management programs, from the electricity utility perspective. As an additional result of the integration of the Cloud Computing infrastructure to such programs, the strategy searches for the most affordable way for the infrastructure provisioning. Also, the management of the workload concerning the energy supply allows a most beneficial deployment of renewable energy sources.

For the strategy evolution, concepts associated with energy, ICT, and economy sectors were deployed. Although previous works studied in chapter 2 apply the presented concepts, there is no reference to a method that encompasses such different areas to manage the infrastructure. Also, combining concepts of different areas on behalf of the resource management of a Cloud Computing is a contribution of this work.

The integration of such concepts also enables to set the energy as a manageable resource of the environment. Thus, this enables to sell the energy as a contractible service for the user and direct provisioned resource of the Cloud Computing.

The main contributions of this work are:

1. Setting a demand management strategy according to energy terms, from the Cloud Computing services provider perspective.
2. Propose a schedule schema for contracts according to the energy deployment, infrastructure usage, and quality conditions.
3. Establishing energy service-level terms that allow the assignment of the contracted demand in energy provisioning terms.
4. Definition of a service layer able to manage the infrastructure under provided energy information.
5. Setting a relation between concepts of ICT, Cloud Computing, energy sector and economy.
6. Enable a technical solution for the DCs entry in DR programs.
7. Guarantee SLA and GreenSLA terms, concerning the quality of the service provisioning and efficiency of the energy deployment.
8. Allow a power-aware workload for geographical allocated DCs.
9. The independence of energy market definitions for the strategy implementation.

6.4 FUTURE WORKS

As future works, the internal capacity of the DCs must be considered as manageable information for the contracts assignment. As detailed in section 6.1, the requirements concerning this measurement and this information are not, at this point, considered for the management of the environment. Although the proposal takes into account the physical infrastructure to define scheduling strategies, during the model evaluation of the work, this information was not developed as well.

The internal capacity enables to set energy efficiency as possible contract constraints, as its level is an important fact to be considered during the allocation. The resources deployment level also enables to establish a more robust management strategy and to sell the efficiency and financial savings as contractible items for the user.

The evaluation of energy service terms related to the efficiency and cost savings may increase economic gains for the service provider, along with a more sustainable deployment of the infrastructure expenditure.

REFERENCES

ADDIS, B. et al. Energy-aware joint management of networks and cloud infrastructures. **Computer Networks**, Elsevier, v. 70, p. 75–95, 2014.

ANWAR, A. et al. Cost-aware cloud metering with scalable service management infrastructure. **2015 IEEE 8th International Conference on Cloud Computing**. p. 285–292, 2015.

ARROBA, P.; BUYYA, R. et al. Dvfs-aware consolidation for energy-efficient clouds. **2015 International Conference on Parallel Architecture and Compilation (PACT)**. p. 494–495, 2015.

BANERJEE, P. et al. Sustainable data centers: enabled by supply and demand side management. **Proceedings of the 46th Annual Design Automation Conference**. p. 884–887, 2009.

BERL, A. et al. Integrating data centres into demand-response management: a local case study. **Industrial Electronics Society, IECON 2013-39th Annual Conference of the IEEE**. p. 4762–4767, 2013.

BOHRA, A. E. H.; CHAUDHARY, V. Vmeter: power modelling for virtualized clouds. **Parallel & Distributed Processing, Workshops and Phd Forum (IPDPSW), 2010 IEEE International Symposium on**. p. 1–8, 2010.

BUNSE, C.; KLINGERT, S.; SCHULZE, T. Greenslas: supporting energy-efficiency through contracts. In: **Energy Efficient Data Centers**. [S.l.]: Springer, 2012. p. 54–68.

CHEN, C.; HE, B.; TANG, X. Green-aware workload scheduling in geographically distributed data centers. **Cloud Computing Technology and Science (CloudCom), 2012 IEEE 4th International Conference on**. p. 82–89, 2012.

COLMANT, M. et al. Process-level power estimation in vm-based systems. **Proceedings of the Tenth European Conference on Computer Systems**. p. 14, 2015.

COOK, G. et al. **Clicking Clean: How Companies are Creating the Green Internet**. [S.l.: s.n.], 2014.

COOKE, D. **Empowering Customer Choice in Electricity Markets**. [S.l.: s.n.], 2011.

CORD. **Analysis of transport network architectures for structural convergence**. [S.l.: s.n.], 2015.

DONG, C. et al. Green power analysis for geographical load balancing based datacenters. **Green Computing Conference (IGCC), 2013 International**. p. 1–8, June 2013.

FURHT, B.; ESCALANTE, A. **Handbook of Cloud Computing**. [S.l.]: Springer US, 2010. (Computer science).

GELENBE, E. Energy packet networks: adaptive energy management for the cloud. **Proceedings of the 2nd International Workshop on Cloud Computing Platforms**. p. 1, 2012.

GOIRI, Í. et al. Greenslot: scheduling energy consumption in green datacenters. **Proceedings of 2011 International Conference for High Performance Computing, Networking, Storage and Analysis**. p. 20, 2011.

GOSWAMI, D.; KREITH, F. **Energy Efficiency and Renewable Energy Handbook, Second Edition**. [S.l.]: CRC Press, 2015. (Mechanical and Aerospace Engineering Series).

HAQUE, M. E. et al. Providing green slas in high performance computing clouds. **Green Computing Conference (IGCC), 2013 International**. p. 1–11, 2013.

HASAN, S. et al. Exploiting renewable sources: when green sla becomes a possible reality in cloud computing. IEEE, 2015.

HILES, A. **The Complete Guide to IT Service Level Agreements: Aligning IT Services to Business Needs**. [S.l.]: Rothstein Catalog on Service Level Books, Rothstein Associates Incorporated, 2002. (Service Level Management Series).

HSU, Y.-C.; LIU, P.; WU, J.-J. Job sequence scheduling for cloud computing. **Cloud and Service Computing (CSC), 2011 International Conference on**. p. 212–219, 2011.

HU, Z.; HAN, X.; WEN, Q. **Integrated Resource Strategic Planning and Power Demand-Side Management**. [S.l.]: Springer Berlin Heidelberg, 2013. (Power Systems).

JANACEK, S. et al. Modeling and approaching a cost transparent, specific data center power consumption. **2012 International Conference on Energy Aware Computing**. p. 1–6, 2012.

JENNINGS, B.; STADLER, R. Resource management in clouds: survey and research challenges. **Journal of Network and Systems Management**, Springer, v. 23, n. 3, p. 567–619, 2015.

JIANG, Z. et al. Vpower: metering power consumption of vm. **Software Engineering and Service Science (ICSESS), 2013 4th IEEE International Conference on**. p. 483–486, 2013.

KIM, N.; CHO, J.; SEO, E. Energy-credit scheduler: an energy-aware virtual machine scheduler for cloud systems. **Future Generation Computer Systems**, Elsevier, v. 32, p. 128–137, 2014.

KONG, F.; LIU, X. A survey on green-energy-aware power management for datacenters. **ACM Computing Surveys (CSUR)**, ACM, v. 47, n. 2, p. 30, 2015.

KRISHNAN, B. et al. Vm power metering: feasibility and challenges. **ACM SIGMETRICS Performance Evaluation Review**, ACM, v. 38, n. 3, p. 56–60, 2011.

LE, T. N. et al. Joint capacity planning and operational management for sustainable data centers and demand response. **Proceedings of the Seventh International Conference on Future Energy Systems**. p. 16, 2016.

LIU, Z. et al. Greening geographical load balancing. **IEEE/ACM Transactions on Networking (TON)**, IEEE Press, v. 23, n. 2, p. 657–671, 2015.

LIU, Z. et al. Pricing data center demand response. **ACM SIGMETRICS Performance Evaluation Review**, ACM, v. 42, n. 1, p. 111–123, 2014.

LIU, Z. et al. Data center demand response: avoiding the coincident peak via workload shifting and local generation. **Performance Evaluation**, Elsevier, v. 70, n. 10, p. 770–791, 2013.

LONGO, A.; ZAPPATORE, M.; BOCHICCHIO, M. A. Service and contract composition: a model and a tool. **Integrated Network Management (IM), 2015 IFIP/IEEE International Symposium on**. p. 1434–1440, 2015.

LONGO, A.; ZAPPATORE, M.; BOCHICCHIO, M. A. Service level aware-contract management. **Services Computing (SCC), 2015 IEEE International Conference on**. p. 499–506, 2015.

LOSI, A.; MANCARELLA, P.; VICINO, A. **Integration of Demand Response Into the Electricity Chain: Challenges, Opportunities and Smart Grid Solutions**. [S.l.]: Wiley, 2015. (Electrical Engineering).

LUČANIN, D.; BRANDIC, I. Pervasive cloud controller for geotemporal inputs. **IEEE Transactions on Cloud Computing**, v. 4, n. 2, p. 180–195, April 2016.

LUCANIN, D.; BRANDIC, I. Take a break: cloud scheduling optimized for real-time electricity pricing. **Cloud and Green Computing (CGC), 2013 Third International Conference on**. p. 113–120, 2013.

MA, C. et al. Virtual machine power metering and its applications. **Global High Tech Congress on Electronics (GHTCE), 2013 IEEE**. p. 153–156, 2013.

MANVI, S. S.; SHYAM, G. K. Resource management for infrastructure as a service (iaas) in cloud computing: a survey. **Journal of Network and Computer Applications**, Elsevier, v. 41, p. 424–440, 2014.

MARKOVIC, D. S. et al. Smart power grid and cloud computing. **Renewable and Sustainable Energy Reviews**, Elsevier, v. 24, p. 566–577, 2013.

MARTENS, B.; WALTERBUSCH, M.; TEUTEBERG, F. Costing of cloud computing services: a total cost of ownership approach. **System Science (HICSS), 2012 45th Hawaii International Conference on**. p. 1563–1572, 2012.

MASHAYEKHY, L. et al. An online mechanism for resource allocation and pricing in clouds. **IEEE Transactions on Computers**, IEEE, v. 65, n. 4, p. 1172–1184, 2016.

MÄSKER, M. et al. Smart grid-aware scheduling in data centres. **Computer Communications**, Elsevier, 2016.

MELL, P.; GRANCE, T. **The NIST Definition of Cloud Computing**. [S.l.: s.n.], 2011.

MINH, H. T.; SAMEJIMA, M. An evaluation of job scheduling based on distributed energy generation in decentralized data centers. **Systems, Man, and Cybernetics (SMC), 2015 IEEE International Conference on**. p. 1172–1177, 2015.

NARAYAN, A.; RAO, S. Power-aware cloud metering. **IEEE Transactions on Services Computing**, IEEE, v. 7, n. 3, p. 440–451, 2014.

PEDRAM, M.; RABAEY, J. **Power Aware Design Methodologies**. [S.l.]: Springer US, 2007.

POMERANTZ G. COOK, K. R. B. J. **Clicking Clean: A Guide to Building the Green Internet**. [S.l.: s.n.], 2015.

RAO, L. et al. Coordinated energy cost management of distributed internet data centers in smart grid. **IEEE Transactions on Smart Grid**, IEEE, v. 3, n. 1, p. 50–58, 2012.

REN, S.; HE, Y.; XU, F. Provably-efficient job scheduling for energy and fairness in geographically distributed data centers. **Distributed Computing Systems (ICDCS), 2012 IEEE 32nd International Conference on**. p. 22–31, 2012.

RUAN, X.; CHEN, H. Performance-to-power ratio aware virtual machine (vm) allocation in energy-efficient clouds. **2015 IEEE International Conference on Cluster Computing**. p. 264–273, 2015.

SCHRÖDER, K.; NEBEL, W. Behavioral model for cloud aware load and power management. **Proceedings of the 2013 international workshop on Hot topics in cloud services**. p. 19–26, 2013.

SCHWIND, M. **Dynamic Pricing and Automated Resource Allocation for Complex Information Services: Reinforcement Learning and Combinatorial Auctions**. [S.l.]: Springer Berlin Heidelberg, 2007. (Lecture Notes in Economics and Mathematical Systems).

SIANO, P. Demand response and smart grids—a survey. **Renewable and Sustainable Energy Reviews**, Elsevier, v. 30, p. 461–478, 2014.

SMITH, J. W. et al. Cloudmonitor: profiling power usage. **Cloud Computing (CLOUD), 2012 IEEE 5th International Conference on**. p. 947–948, June 2012. DOI: 10.1109/CLOUD.2012.112.

SUN, Q. et al. Fair rewarding in colocation data centers: truthful mechanism for emergency demand response. **2015 IEEE 23rd International Symposium on Quality of Service (IWQoS)**. p. 359–368, 2015.

SUPERMICRO. **SuperServer 8028B-TR4F**. [S.l.: s.n.], June 2016.

TANG, C.-J. et al. A load control method for small data centers participating in demand response programs. **Future Generation Computer Systems**, Elsevier, v. 32, p. 232–245, 2014.

TERAMOTO, K.; HUANG, H. H. Poster: pay as you go in the cloud: one watt at a time. **High Performance Computing, Networking, Storage and Analysis (SCC), 2012 SC Companion:**. p. 1548–1548, 2012.

TORRITI, J. **Peak Energy Demand and Demand Side Response**. [S.l.]: Taylor & Francis, 2015. (Routledge Explorations in Environmental Studies).

WANG, R. et al. Datacenters as controllable load resources in the electricity market. **Distributed Computing Systems (ICDCS), 2013 IEEE 33rd International Conference on**. p. 176–185, 2013.

WAZLAWICK, R. **Metodologia de Pesquisa para Ciência da Computação, 2ª Edição**. [S.l.]: Elsevier Brasil, 2014. v. 2.

WIERMAN, A. et al. Opportunities and challenges for data center demand response. **Green Computing Conference (IGCC), 2014 International**. p. 1–10, 2014.

WITHNEY, J.; DELFORGE, P. Data center efficiency assessment. 08/2014 2014.

WU, C.-M.; CHANG, R.-S.; CHAN, H.-Y. A green energy-efficient scheduling algorithm using the {dvfs} technique for cloud datacenters. **Future Generation Computer Systems**, v. 37, p. 141 – 147, 2014.

YU, L.; JIANG, T.; CAO, Y. Energy cost minimization for distributed internet data centers in smart microgrids considering power outages. **IEEE Transactions on Parallel and Distributed Systems**, IEEE, v. 26, n. 1, p. 120–130, 2015.

ZHANG, Z. et al. Policy-based adaptive service level agreement management for cloud services. **Software Engineering and Service Science (ICSESS), 2014 5th IEEE International Conference on**. p. 496–499, 2014.

ZHAO, J. et al. Dynamic pricing and profit maximization for the cloud with geo-distributed data centers. **IEEE INFOCOM 2014-IEEE Conference on Computer Communications**. p. 118–126, 2014.

Appendix A - Results obtained for the First SP Use Case Scenario

Appendix A presents the remaining results for the self-sufficient Cloud Computing environment. The appendix contains the graphs from Test Case 2 to Test Case 6. The Use Cases are described in chapter 5; two different scenarios are proposed, according to the self-generated energy supply. Section 5.2 presents the obtained graphs for the Test Case 1, although it details the results obtained for all the Test Cases proposed.

The difference between each Test Case obtained is the quantity of processing plans contracted, as referenced in Tables 5 and 7.

Test Case 2 - 989 reserved plans, 1037 on-demand plans and 974 flexible plans

Figure 21 - 24-hour energy distribution for the prediction stage - Test Case 2

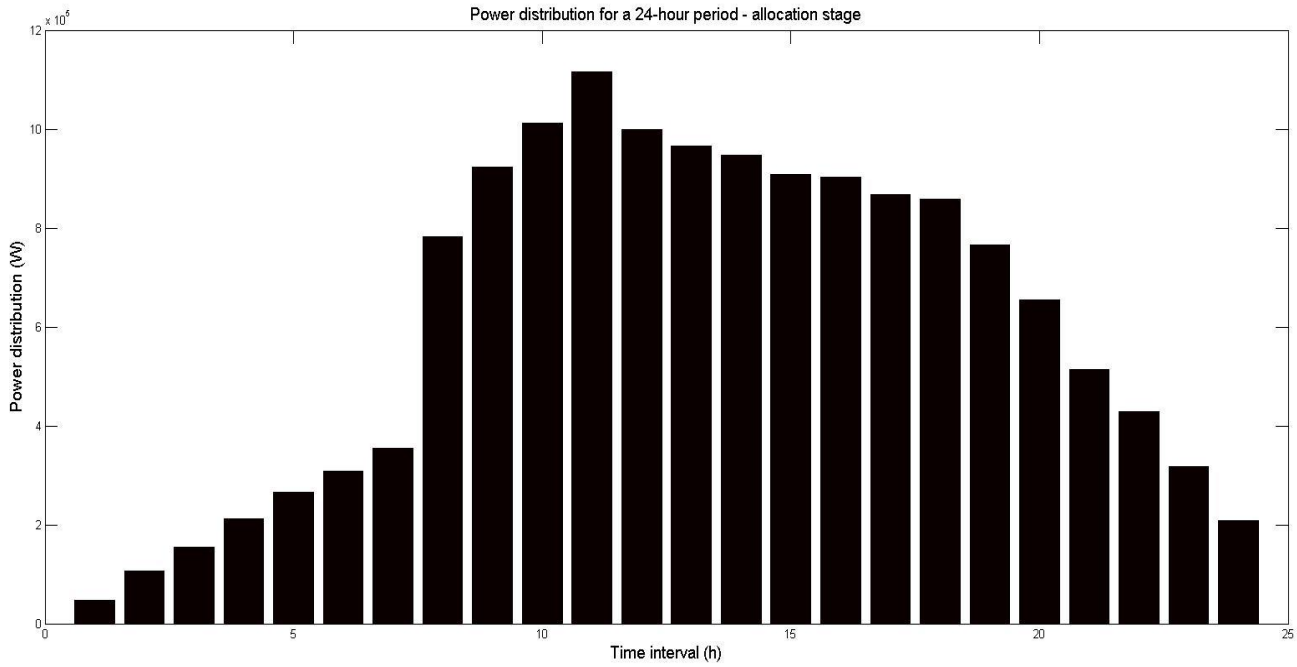


Figure 22 - 24-hour jobs assignment for the prediction stage - Test Case 2

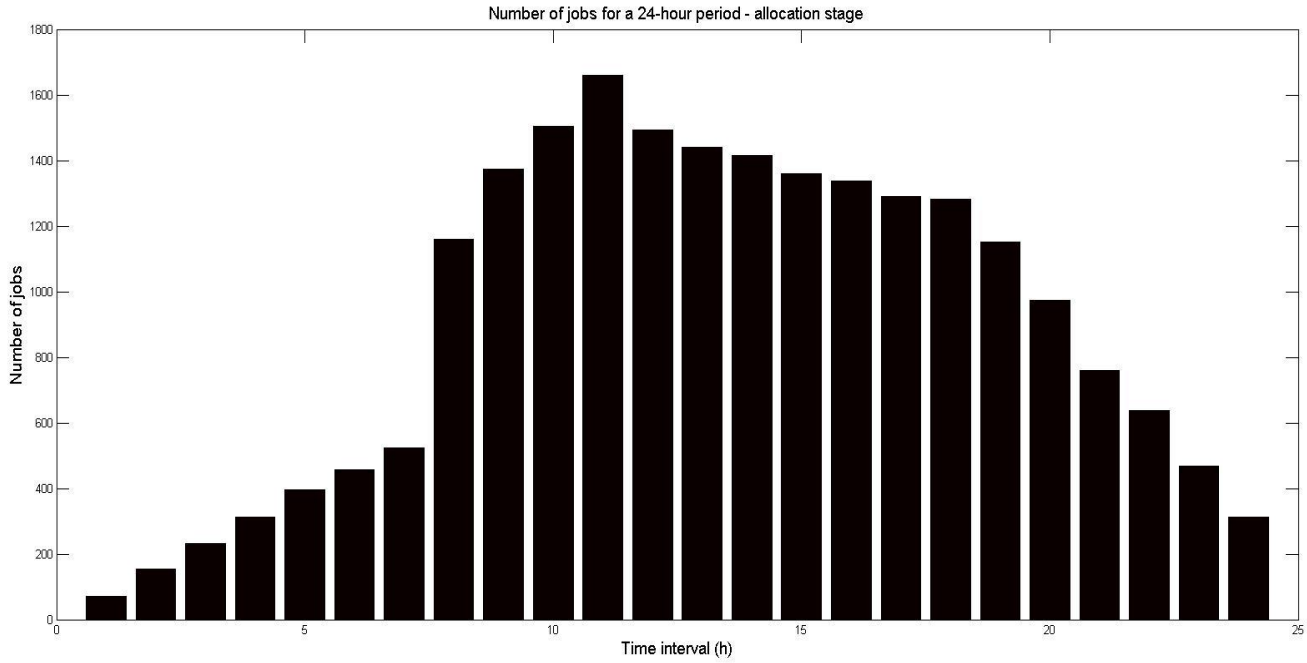


Figure 23 - 24-hour jobs assignment for the negotiation stage - Test Case 2

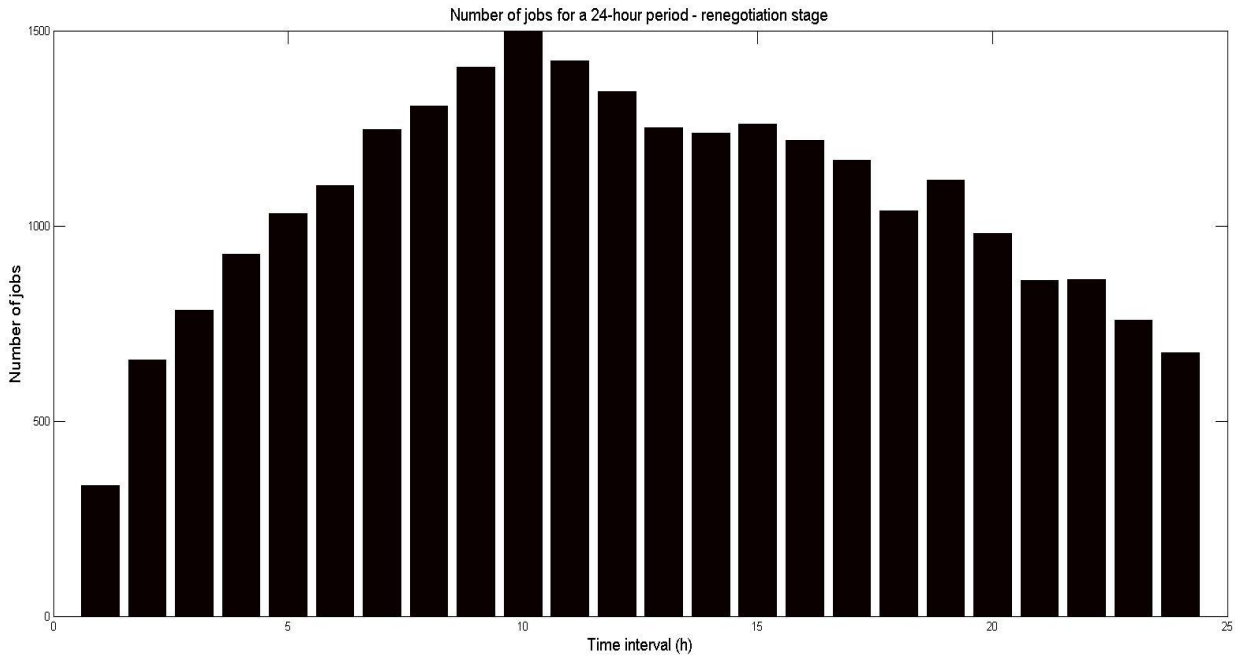


Figure 24 - Energy expenditure estimated for each sub-region, for prediction and negotiation stages - Test Case 2

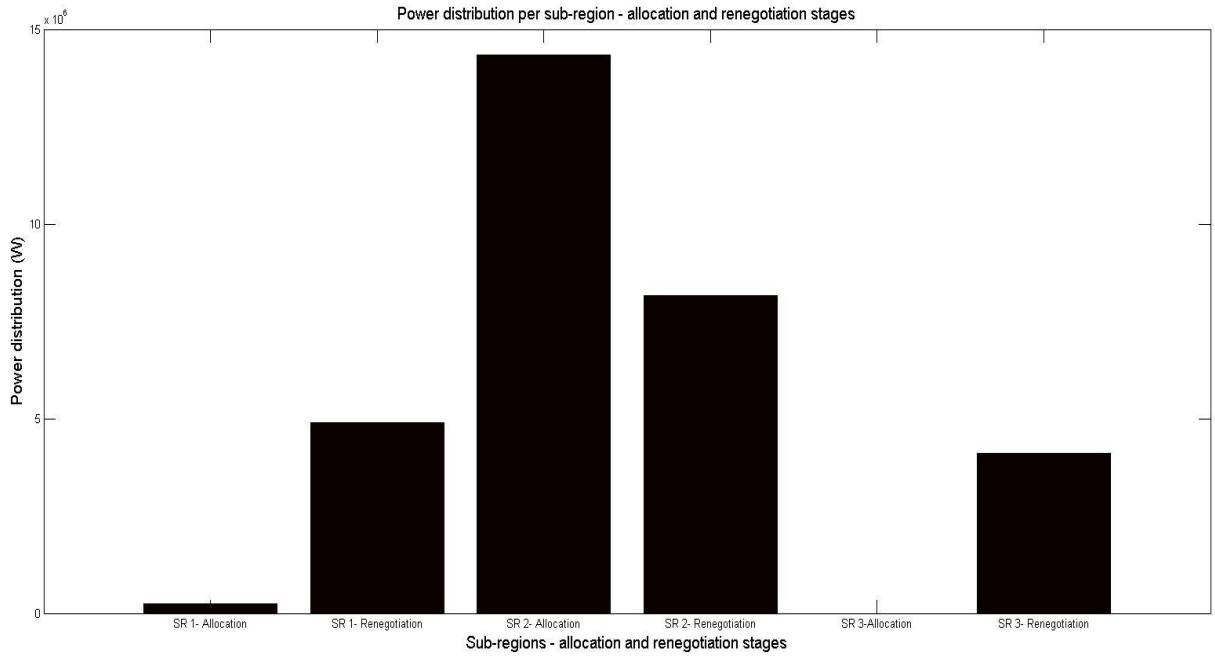
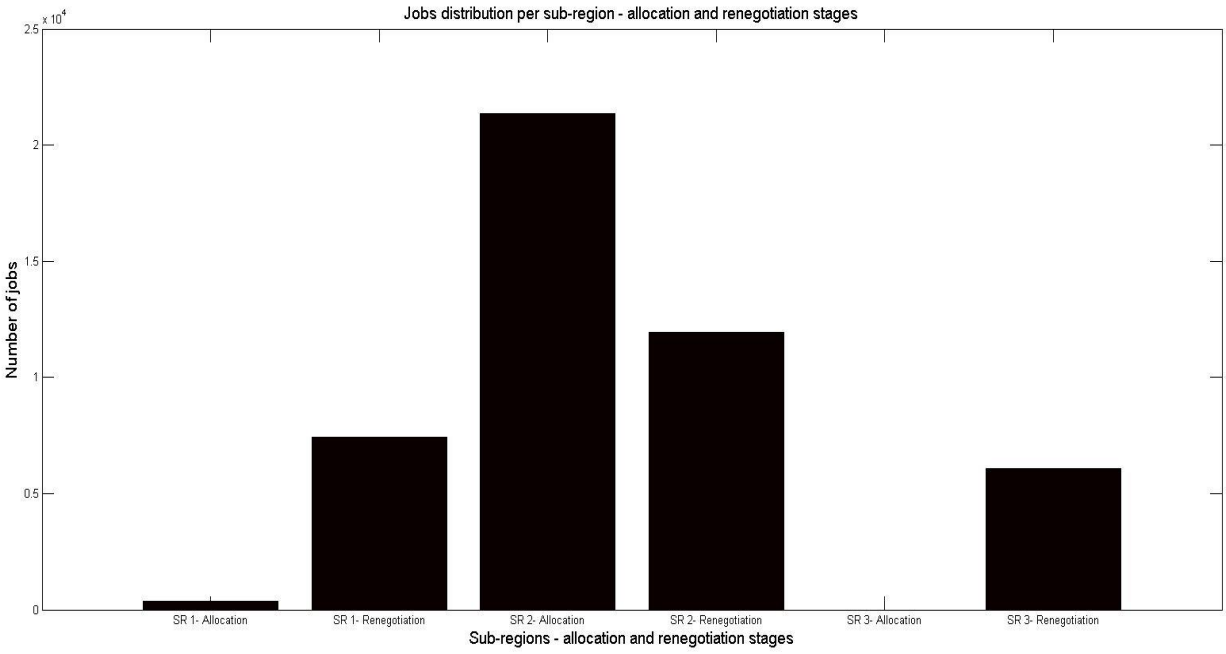


Figure 25 - Jobs assignment for each sub-region, for prediction and negotiation stages - Test Case 2



Test Case 3 - 967 reserved plans, 1015 on-demand plans and 1018 flexible plans

Figure 26 - 24-hour energy distribution for the prediction stage - Test Case 3

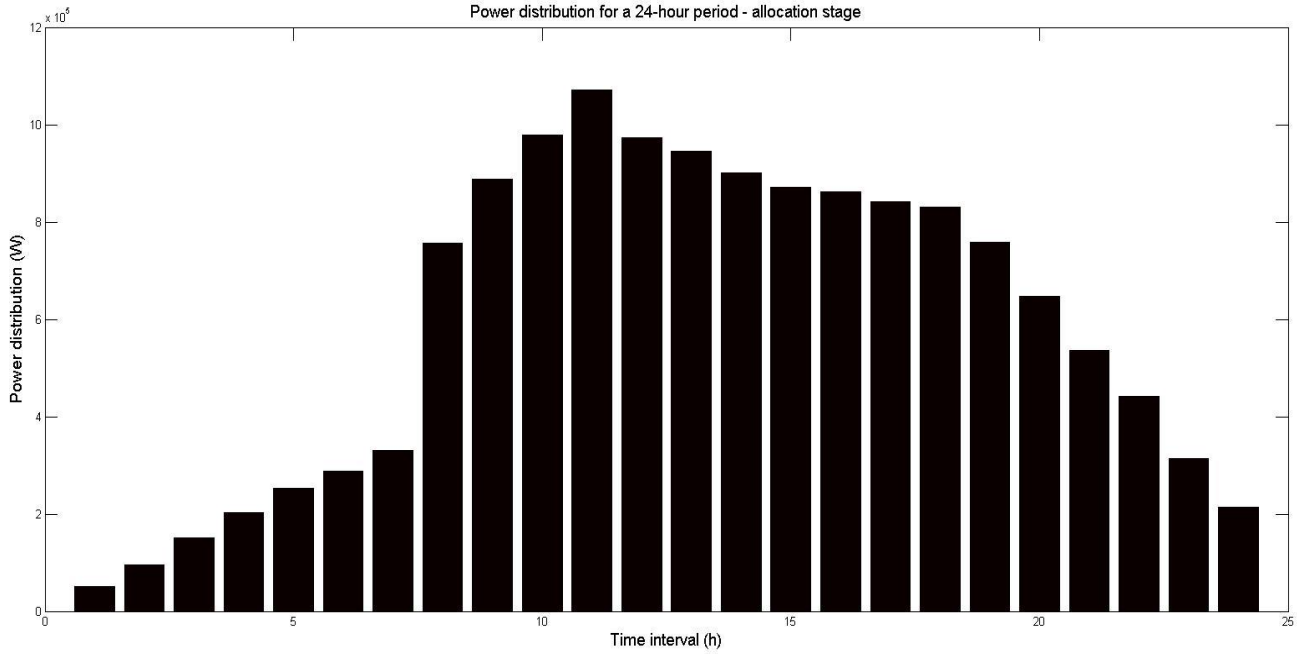


Figure 27 - 24-hour jobs assignment for the prediction stage - Test Case 3

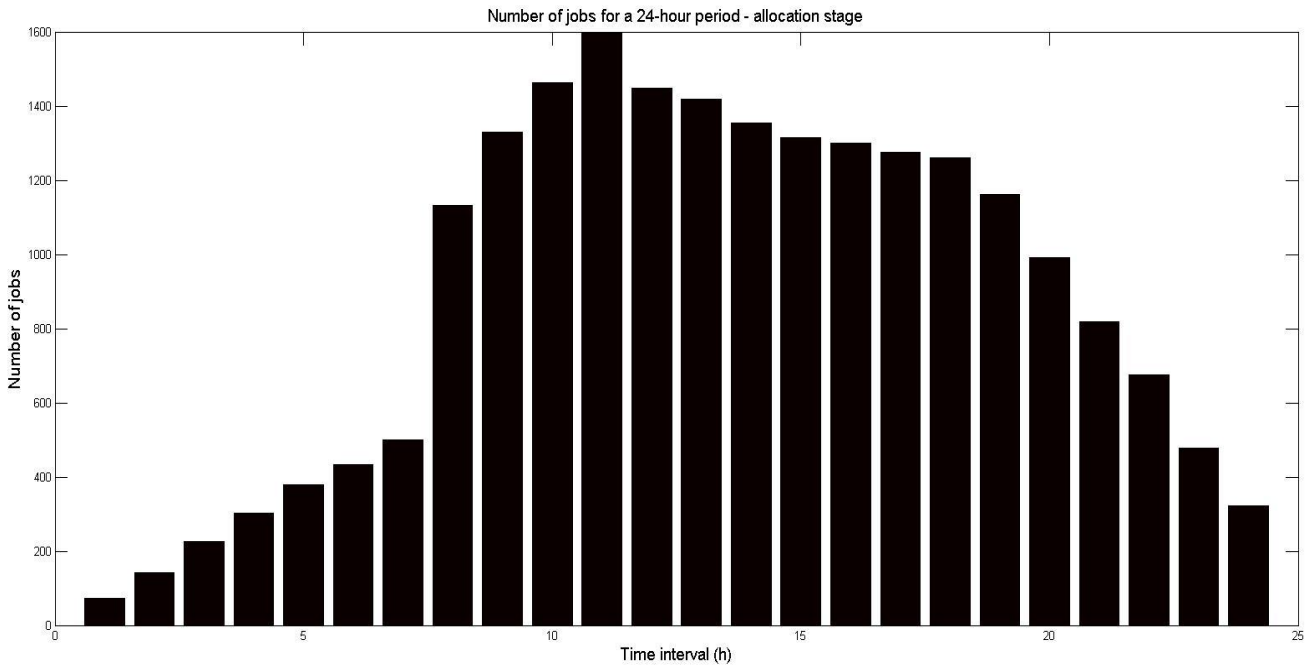


Figure 28 - 24-hour jobs assignment for the negotiation stage - Test Case 3

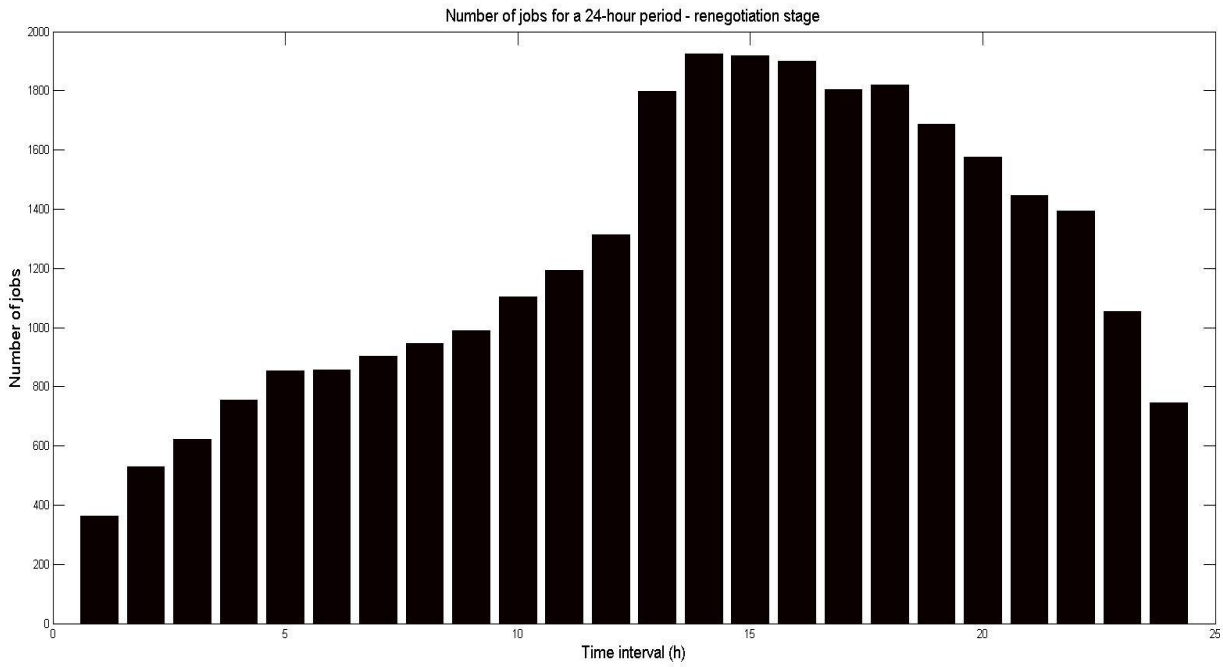


Figure 29 - Energy expenditure estimated for each sub-region, for prediction and negotiation stages - Test Case 3

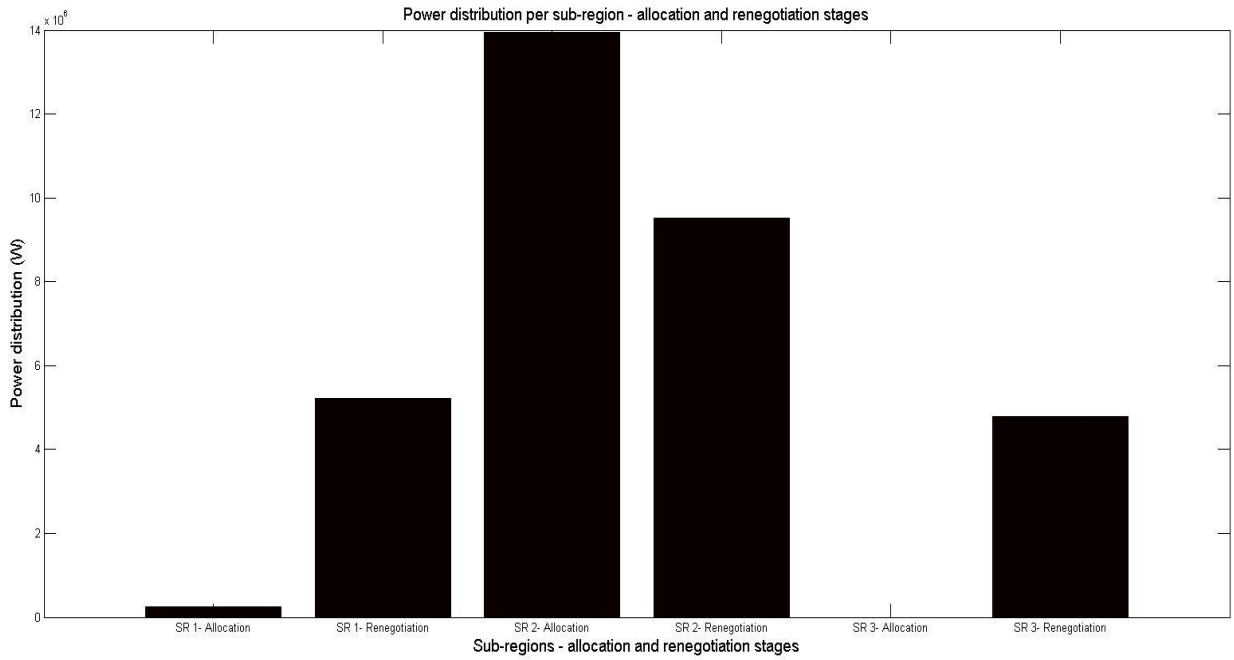
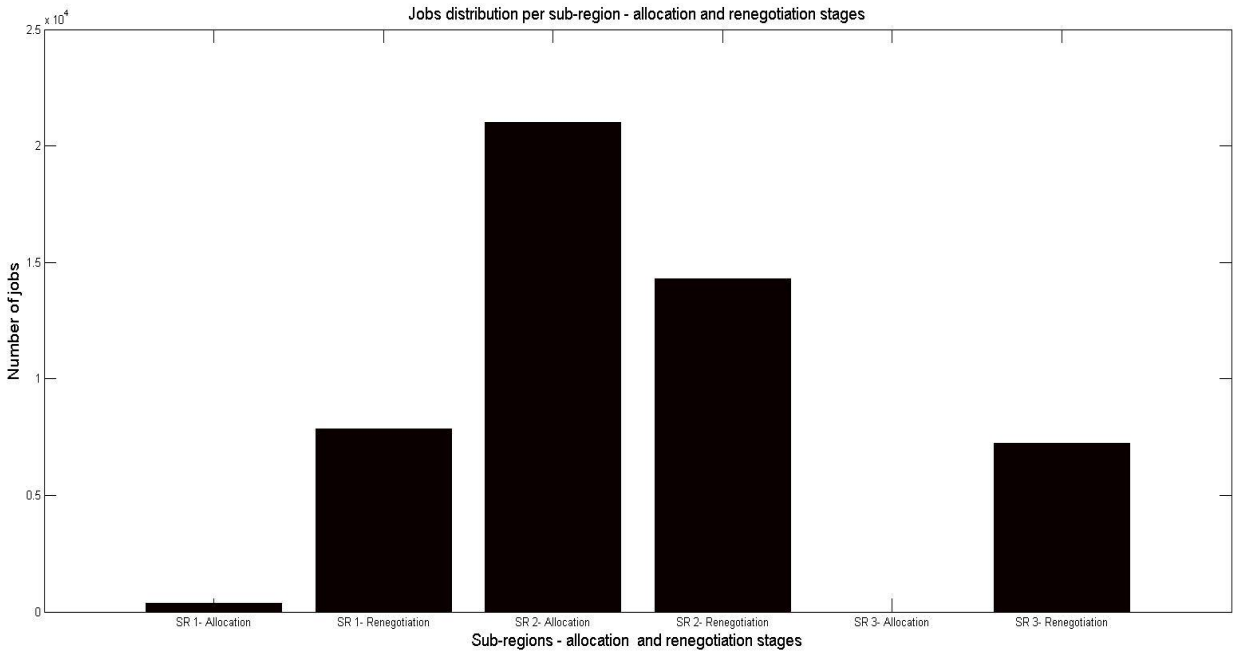


Figure 30 - Jobs assignment for each sub-region, for prediction and negotiation stages - Test Case 3



Test Case 4 - 1050 reserved plans, 994 on-demand plans and 956 flexible plans

Figure 31 - 24-hour energy distribution for the prediction stage - Test Case 4

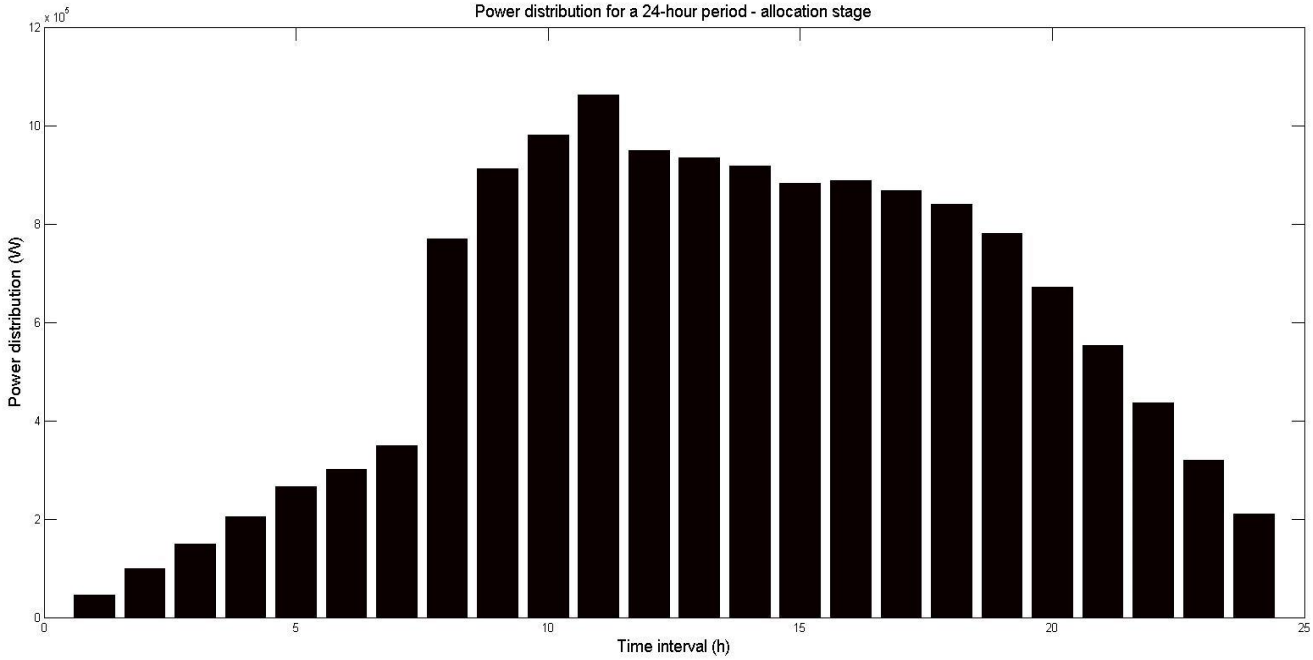


Figure 32 - 24-hour jobs assignment for the prediction stage - Test Case 4

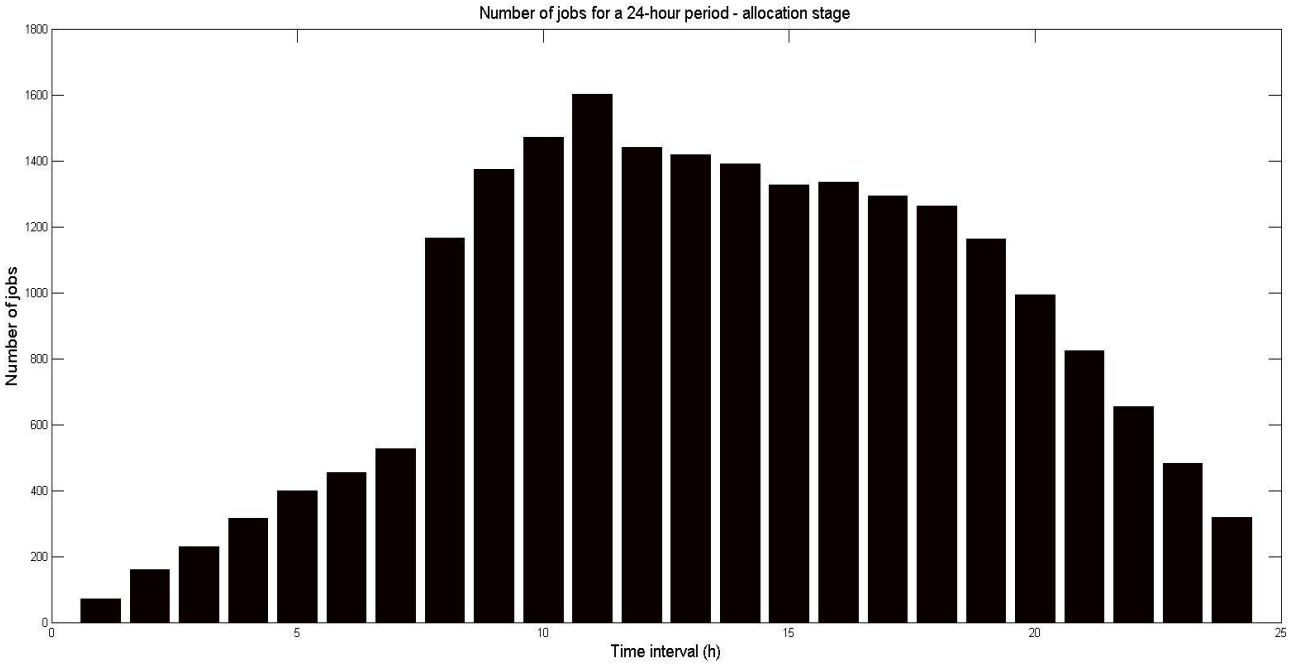


Figure 33 - 24-hour jobs assignment for the negotiation stage - Test Case 4

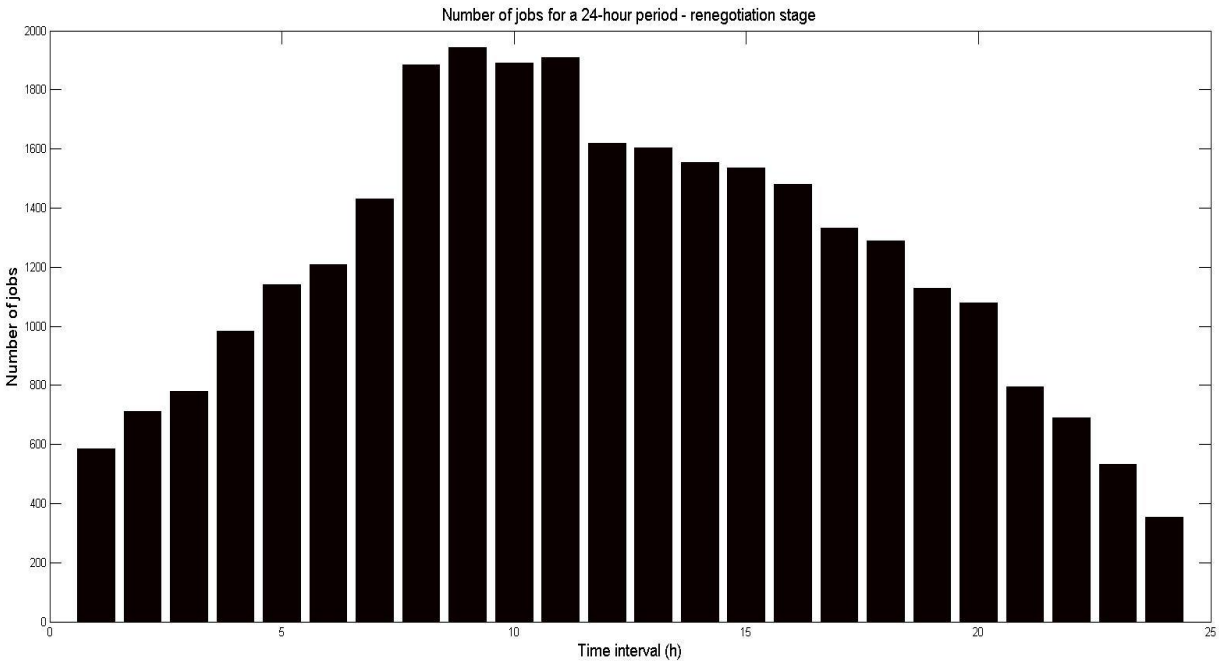


Figure 34 - Energy expenditure estimated for each sub-region, for prediction and negotiation stages - Test Case 4

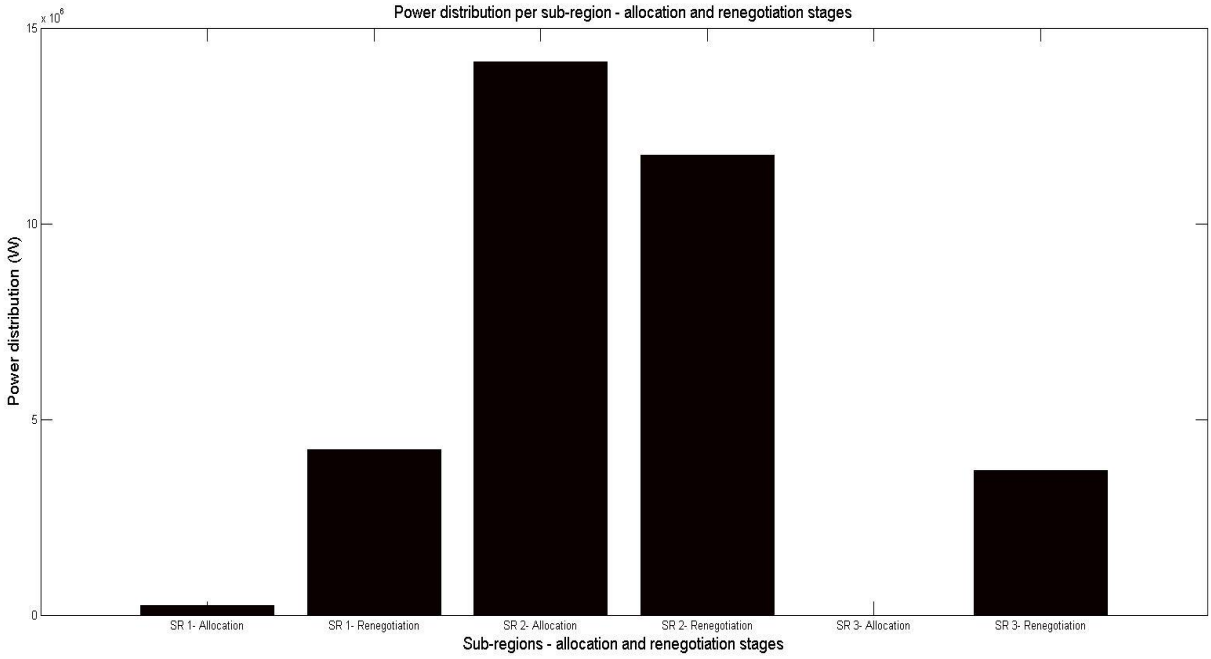
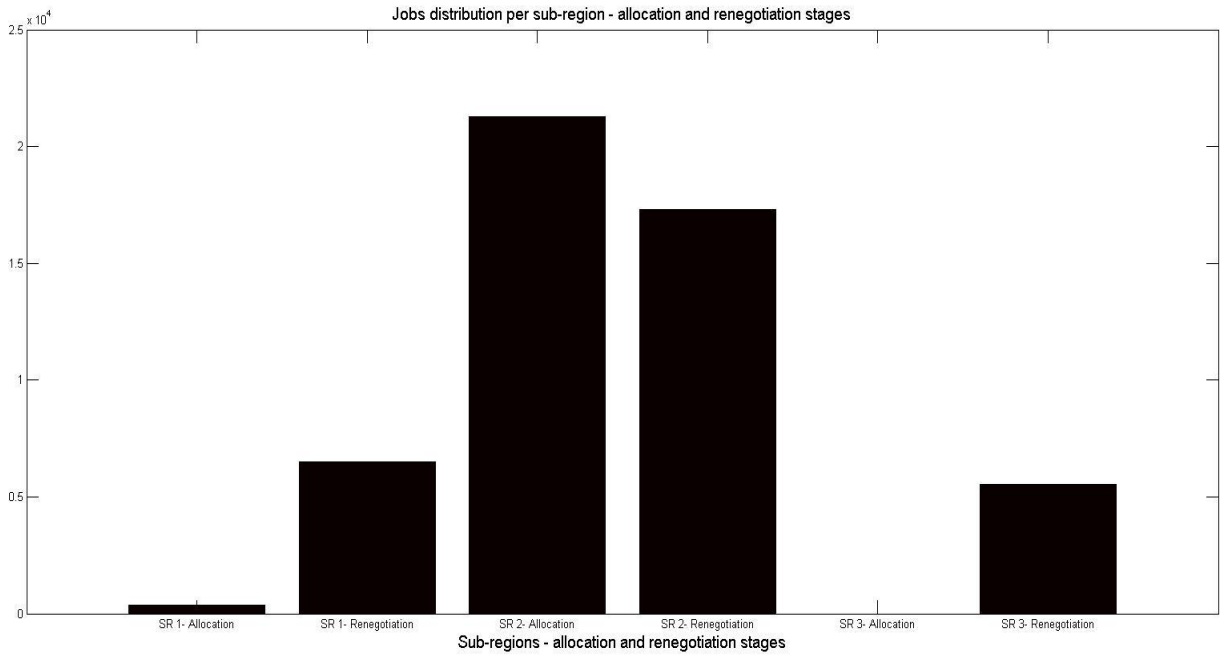


Figure 35 - Jobs assignment for each sub-region, for prediction and negotiation stages - Test Case 4



Test Case 5 - 1023 reserved plans, 947 on-demand plans and 956 flexible plans

Figure 36 - 24-hour energy distribution for the prediction stage - Test Case 5

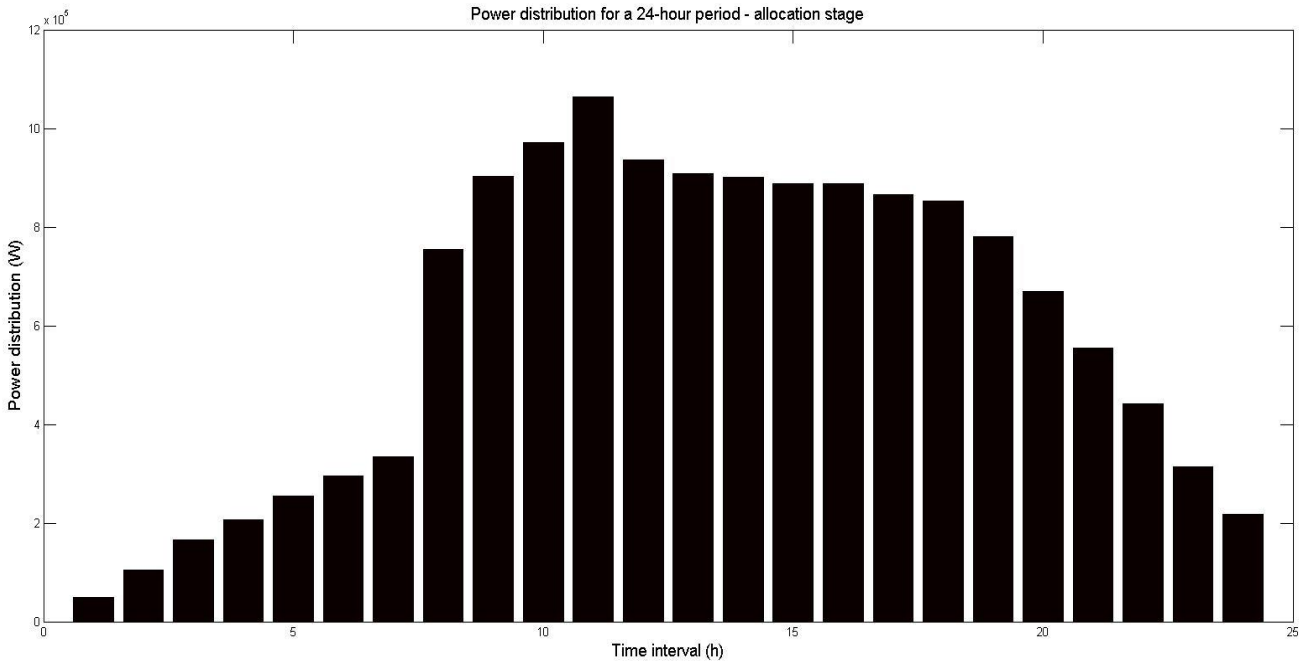


Figure 37 - 24-hour jobs assignment for the prediction stage - Test Case 5

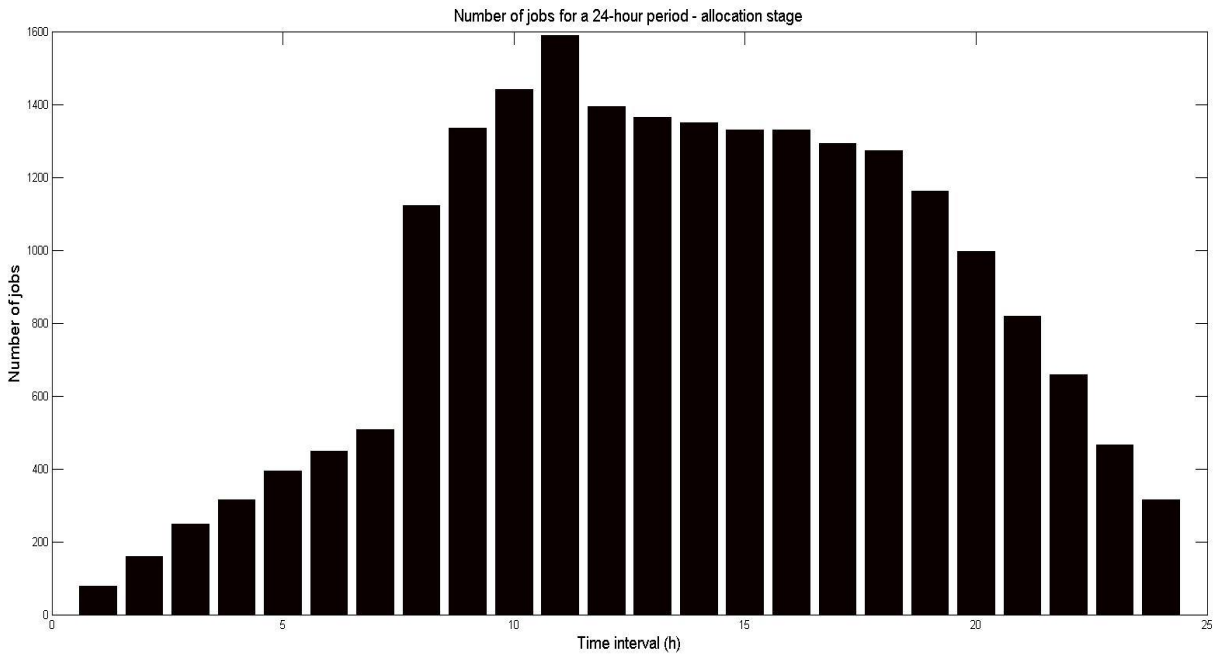


Figure 38 - 24-hour jobs assignment for the negotiation stage - Test Case 5

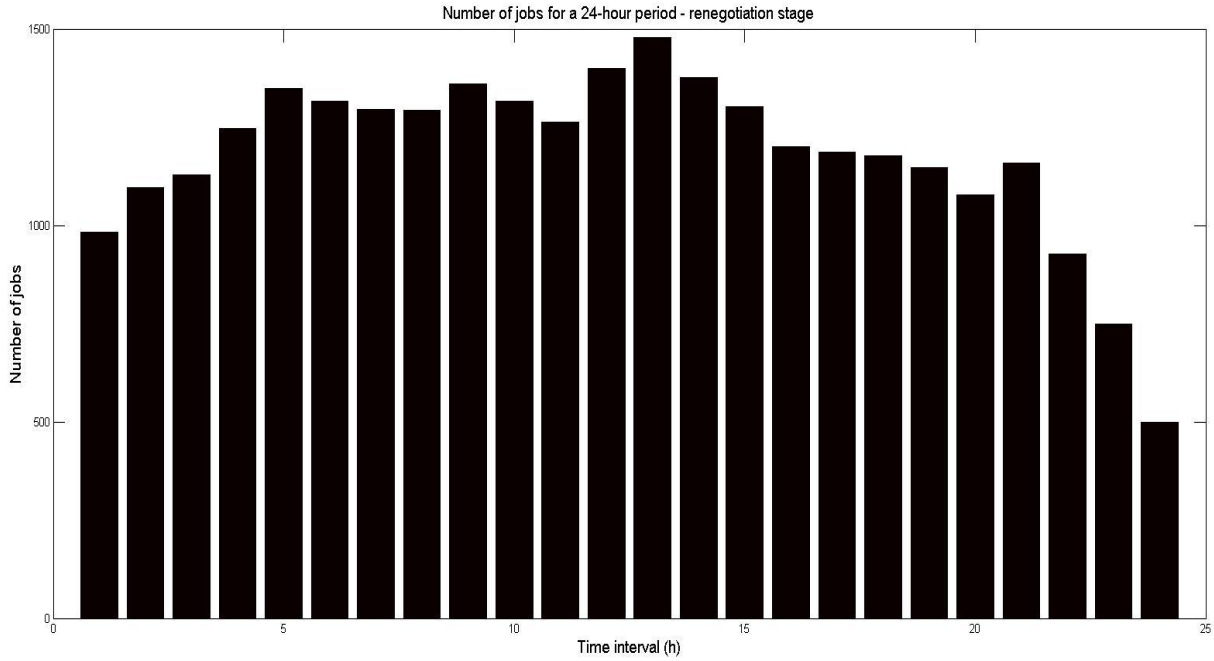


Figure 39 - Energy expenditure estimated for each sub-region, for prediction and negotiation stages - Test Case 5

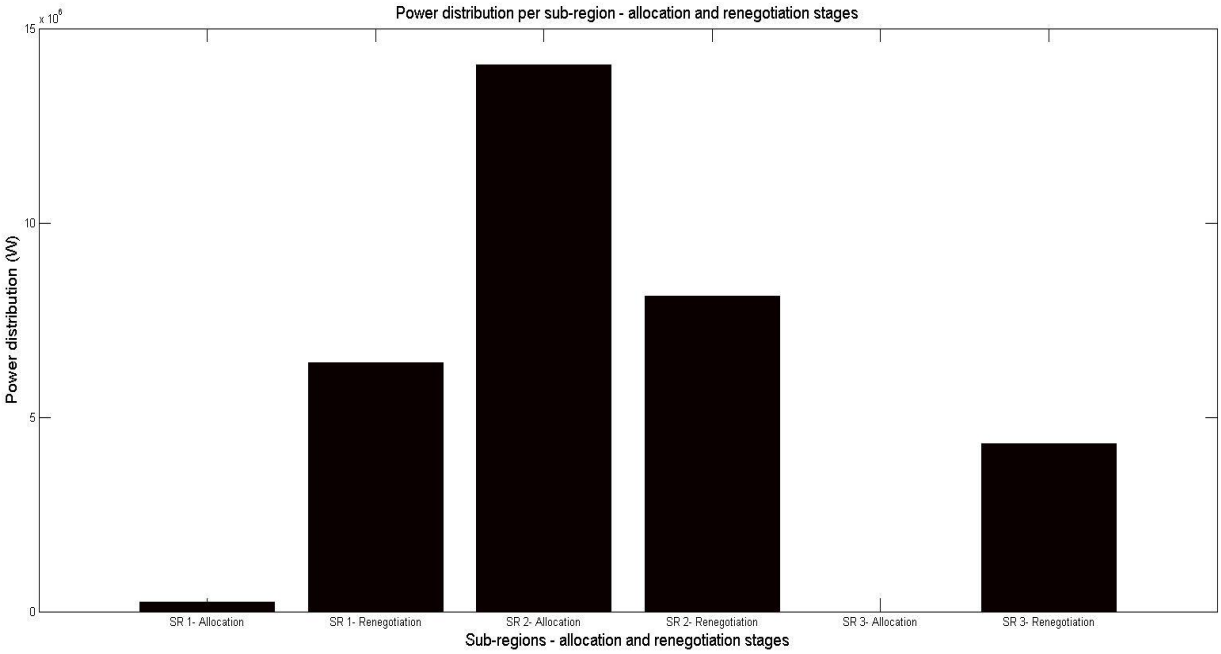
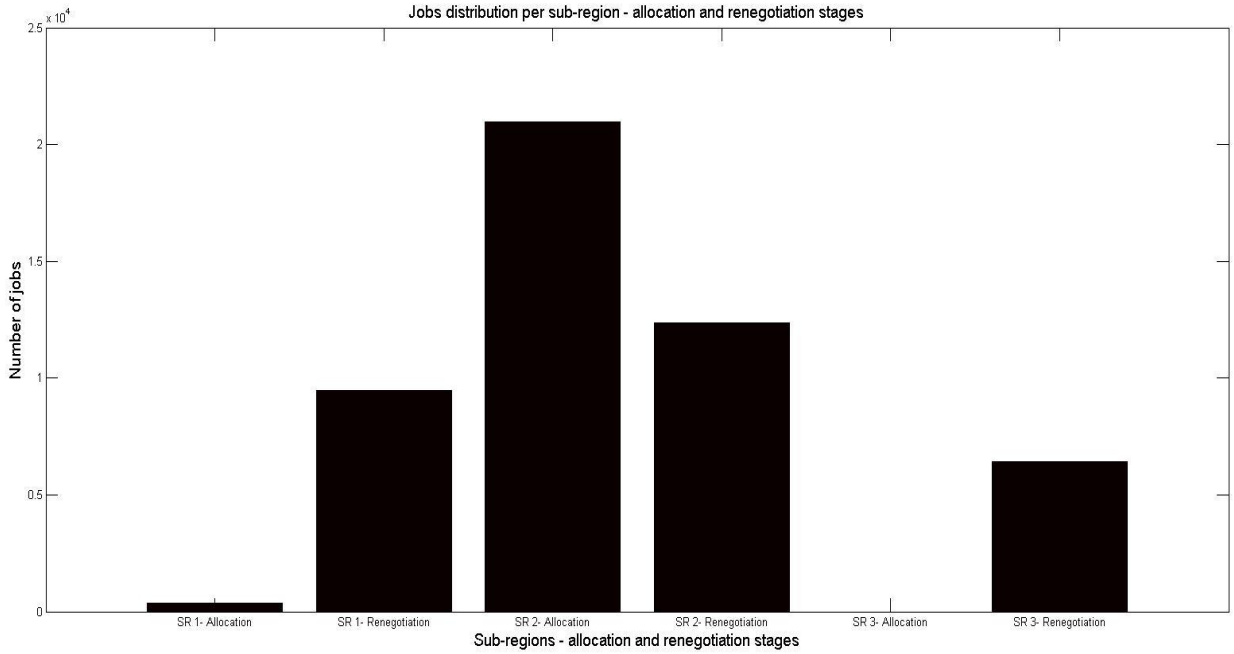


Figure 40 - Jobs assignment for each sub-region, for prediction and negotiation stages - Test Case 5



Test Case 6 - 986 reserved plans, 998 on-demand plans and 1016 flexible plans

Figure 41 - 24-hour energy distribution for the prediction stage - Test Case 6

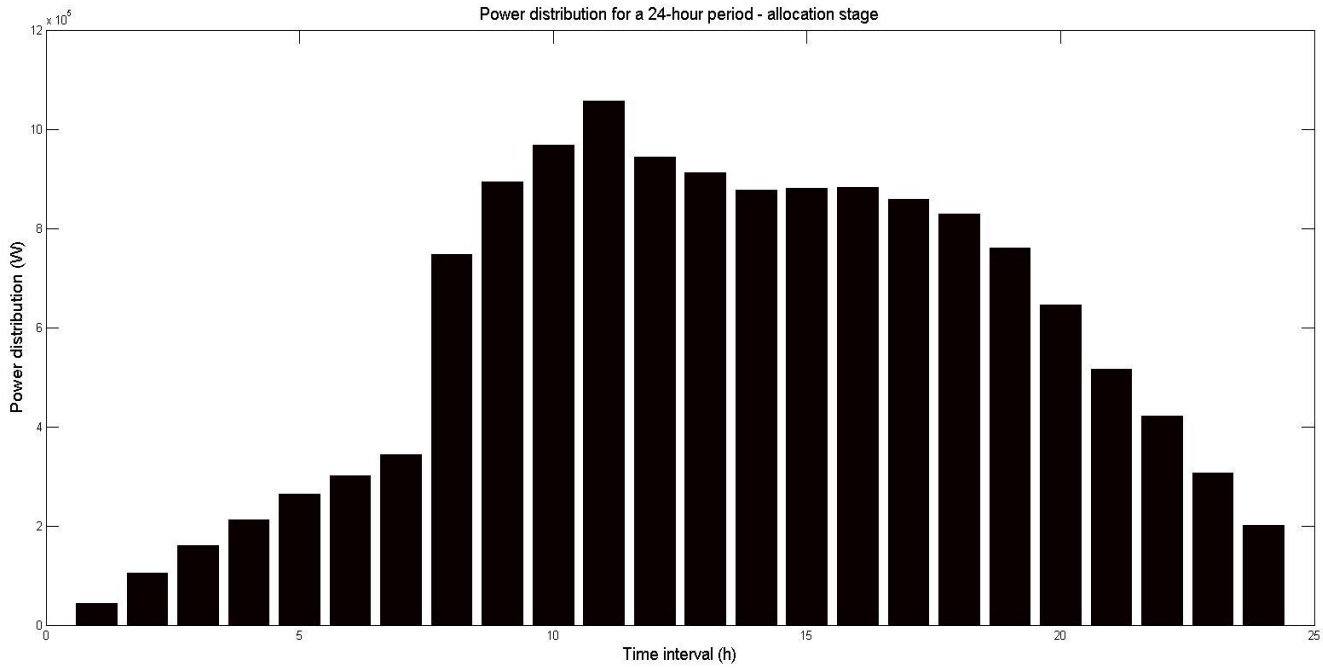


Figure 42 - 24-hour jobs assignment for the prediction stage - Test Case 6

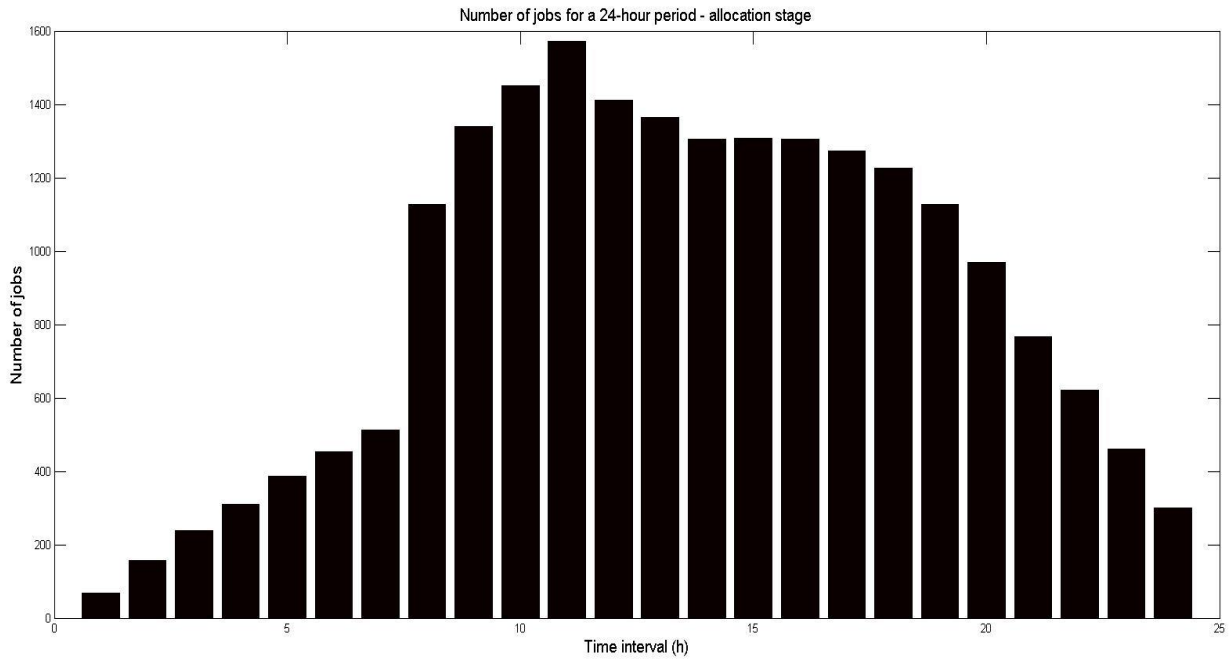


Figure 43 - 24-hour jobs assignment for the negotiation stage - Test Case 6

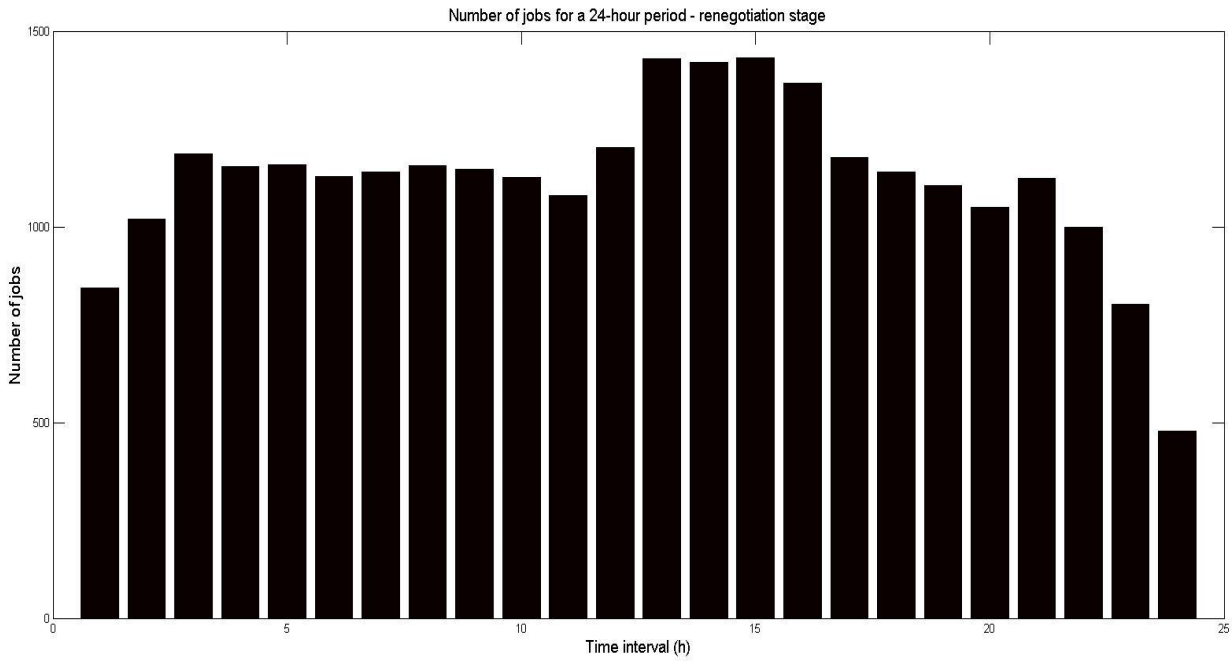


Figure 44 - Energy expenditure estimated for each sub-region, for prediction and negotiation stages - Test Case 6

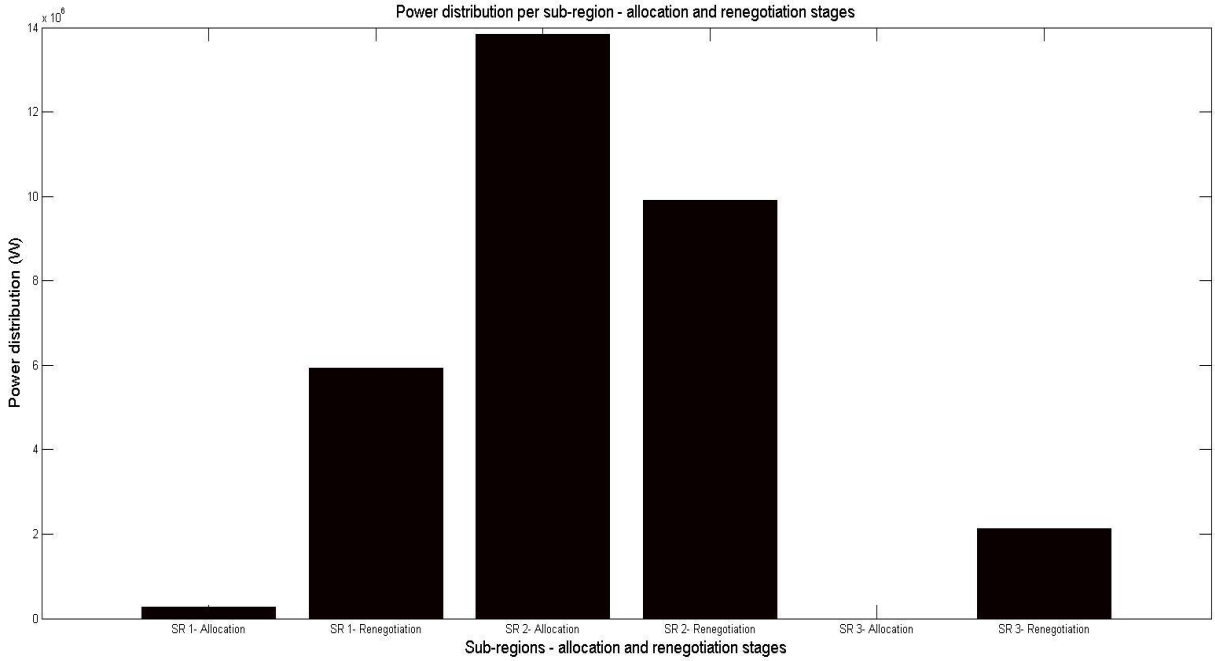
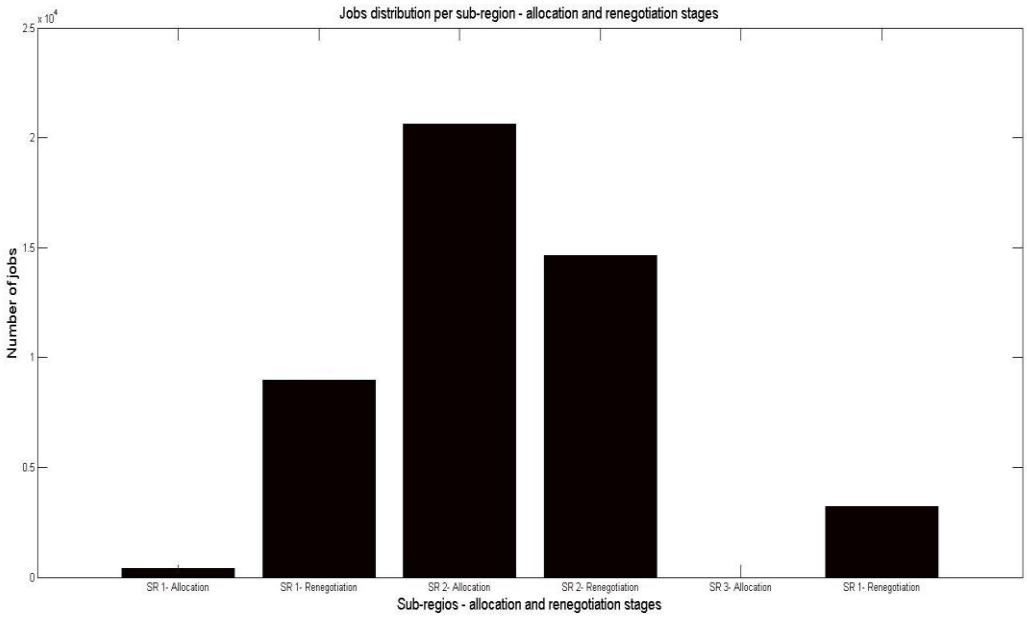


Figure 45 - Jobs assignment for each sub-region, for prediction and negotiation stages - Test Case 6



Appendix B - Results obtained for the Second Scenario of SP Use Case

Appendix B presents the obtained results for the second Use Case, according to the description in chapter 5. The appendix contains the results from Test Case 2 to Test Case 6. Section 5.2 presents the obtained graphs for the Test Case 1, although it details the results obtained for all the Test Cases proposed.

The difference between each Test Case obtained is the quantity of processing plans contracted, as referenced in Tables 5 and 7.

Test Case 2 - 1019 reserved plans, 988 on-demand plans and 993 flexible plans

Figure 46 - 24-hour energy distribution for the prediction stage - Test Case 2

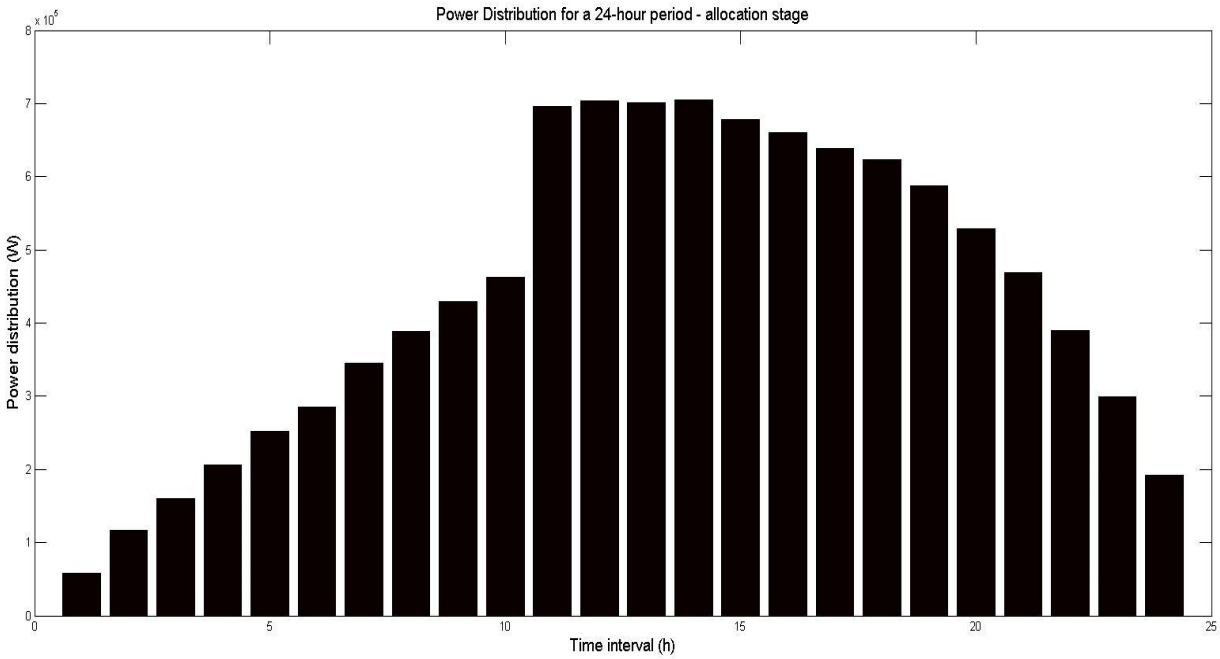


Figure 47 - 24-hour jobs assignment for the prediction stage - Test Case 2

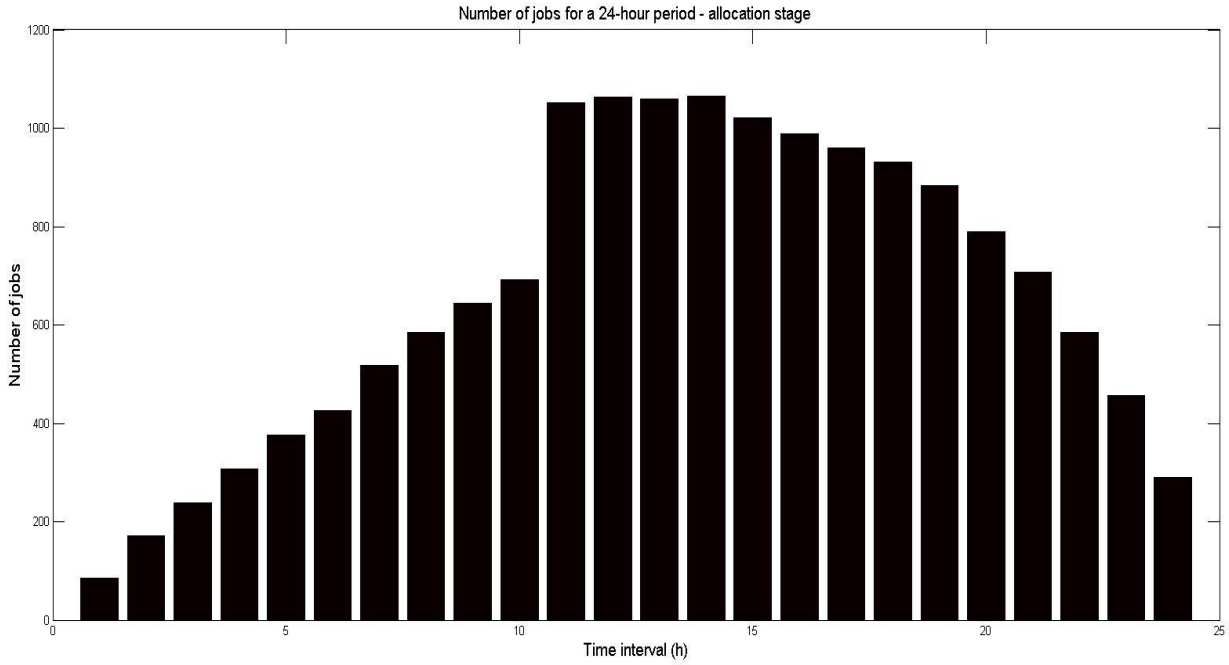


Figure 48 - Expected acquired energy for a 24-hour period, prediction stage - Test Case 2

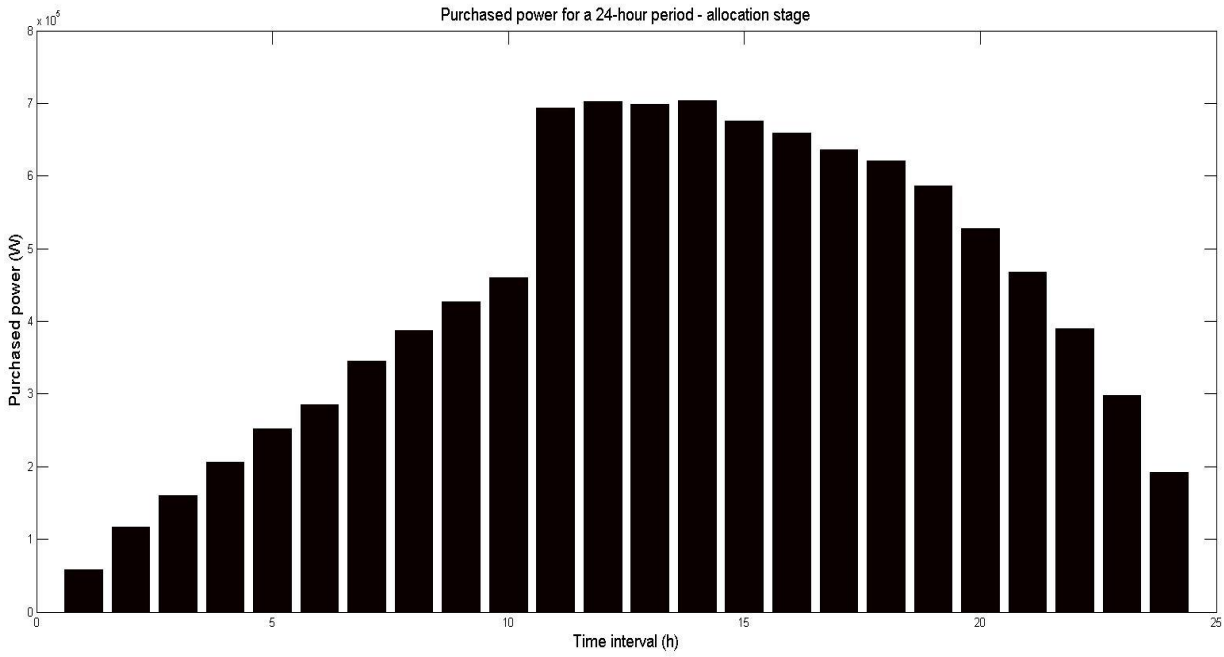


Figure 49 - 24-hour energy distribution for the negotiation stage - Test Case 2

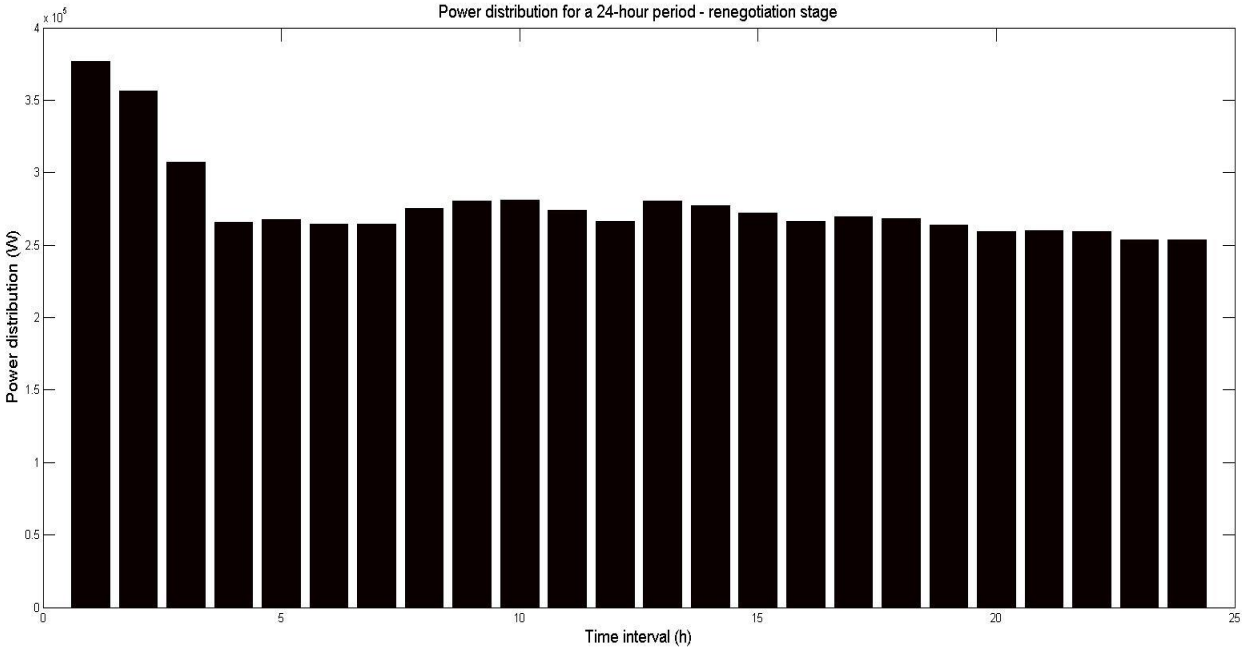


Figure 50 - 24-hour jobs assignment for the negotiation stage - Test Case 2

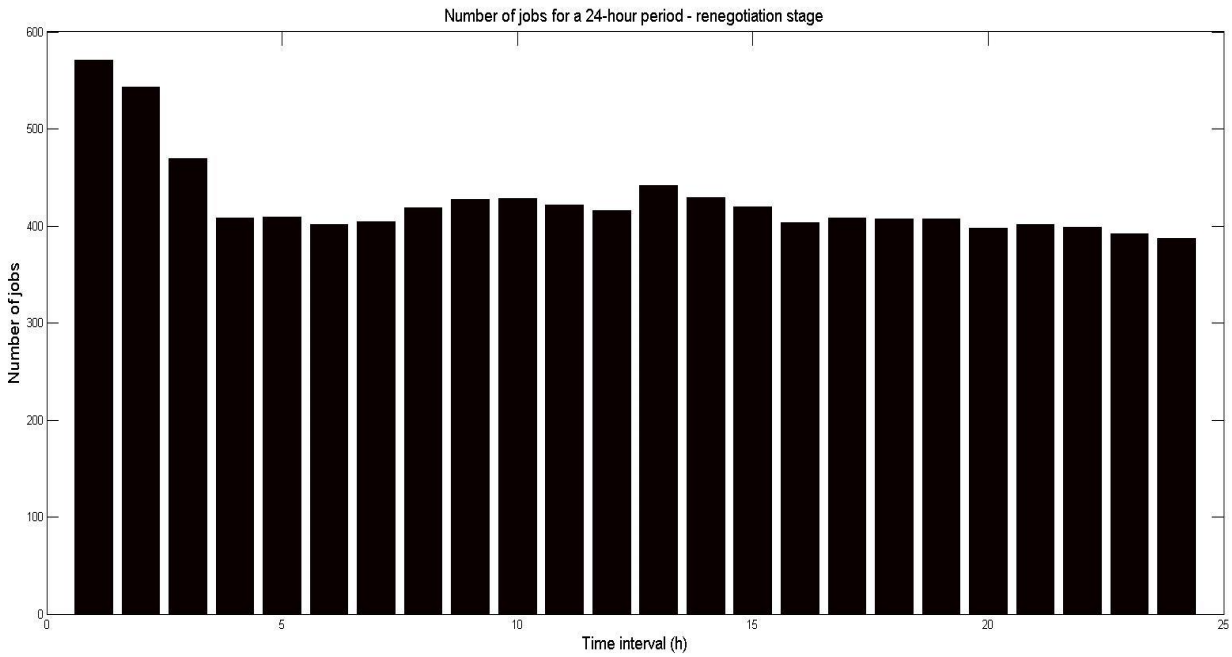


Figure 51 - Expected acquired energy for a 24-hour period, negotiation stage - Test Case 2

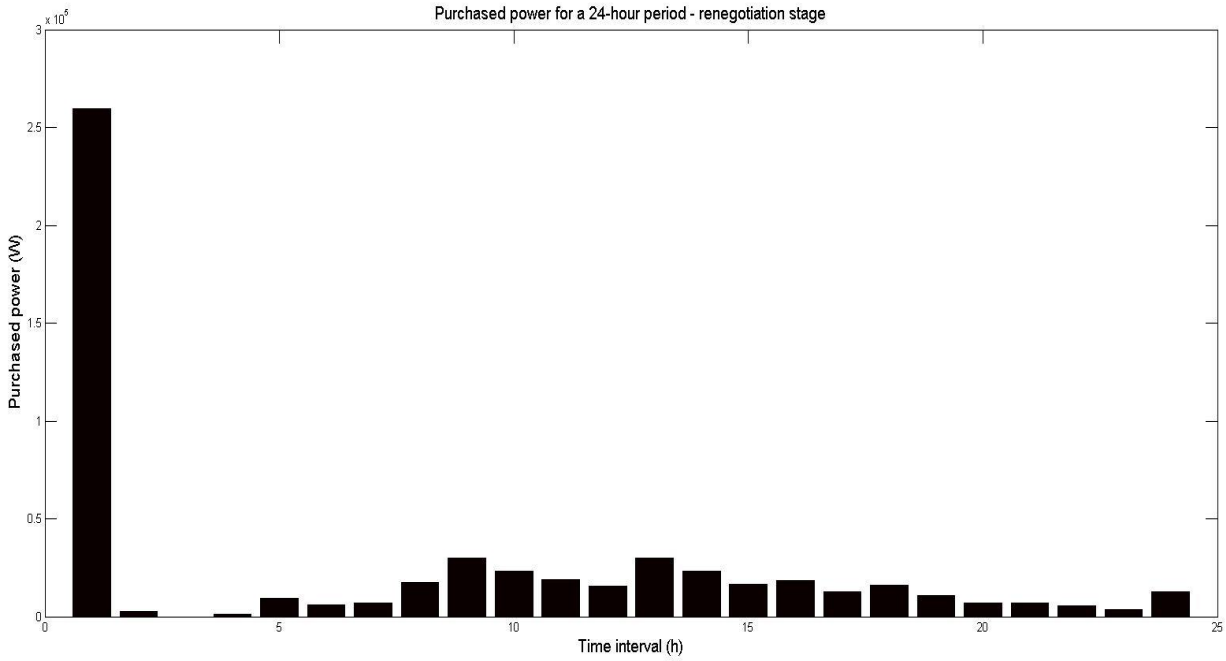


Figure 52 - Energy expenditure estimated for each sub-region, prediction and negotiation stages - Test Case 2

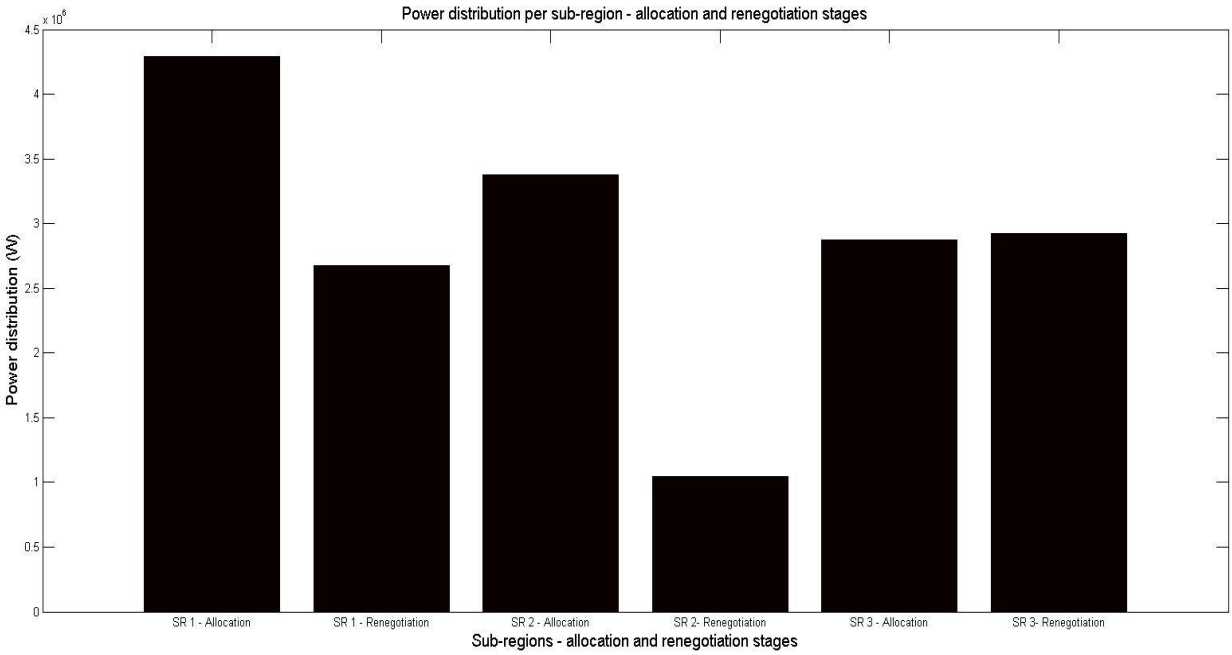
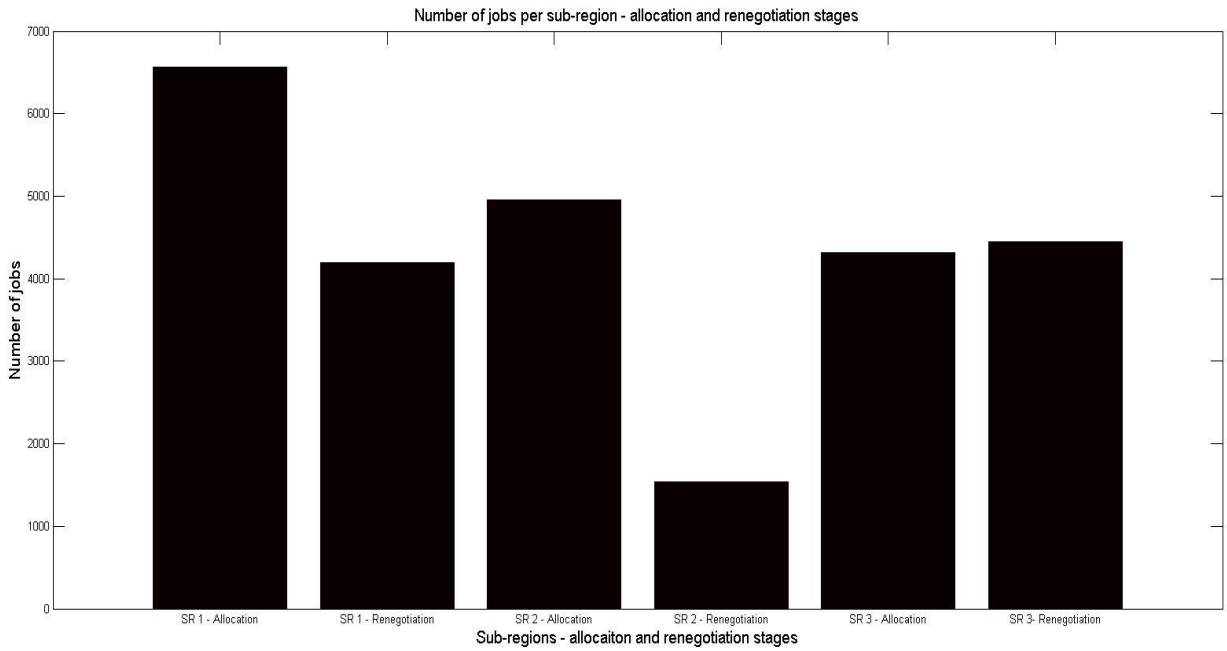


Figure 53 - Jobs assignment for each sub-region, prediction and negotiation stages - Test Case 2



Test Case 3 - 1016 reserved plans, 925 on-demand plans and 1059 flexible plans

Figure 54 - 24-hour energy distribution for the prediction stage - Test Case 3

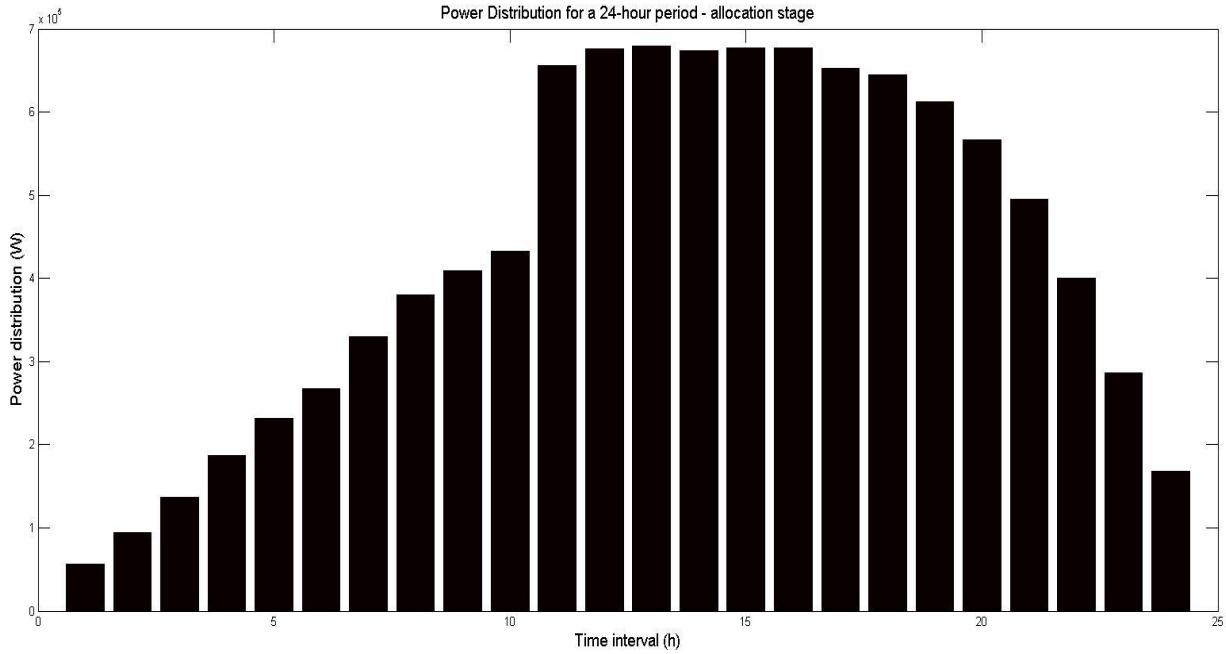


Figure 55 - 24-hour jobs assignment for the prediction stage - Test Case 3

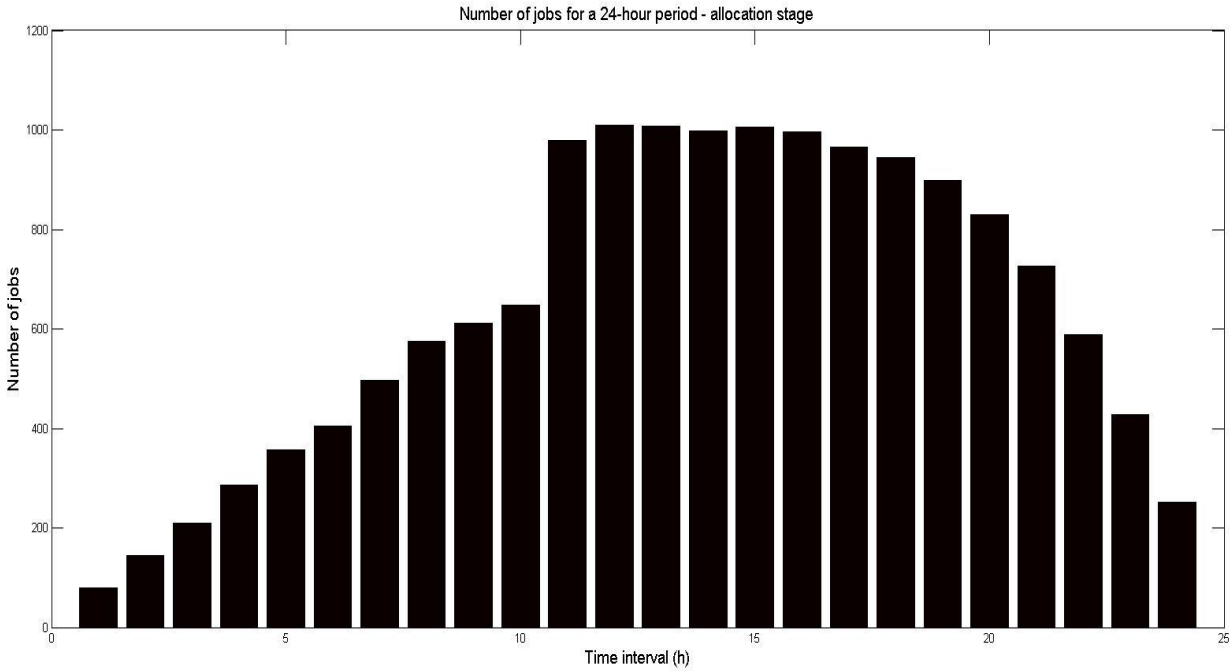


Figure 56 - Expected acquired energy for a 24-hour period, prediction stage - Test Case 3

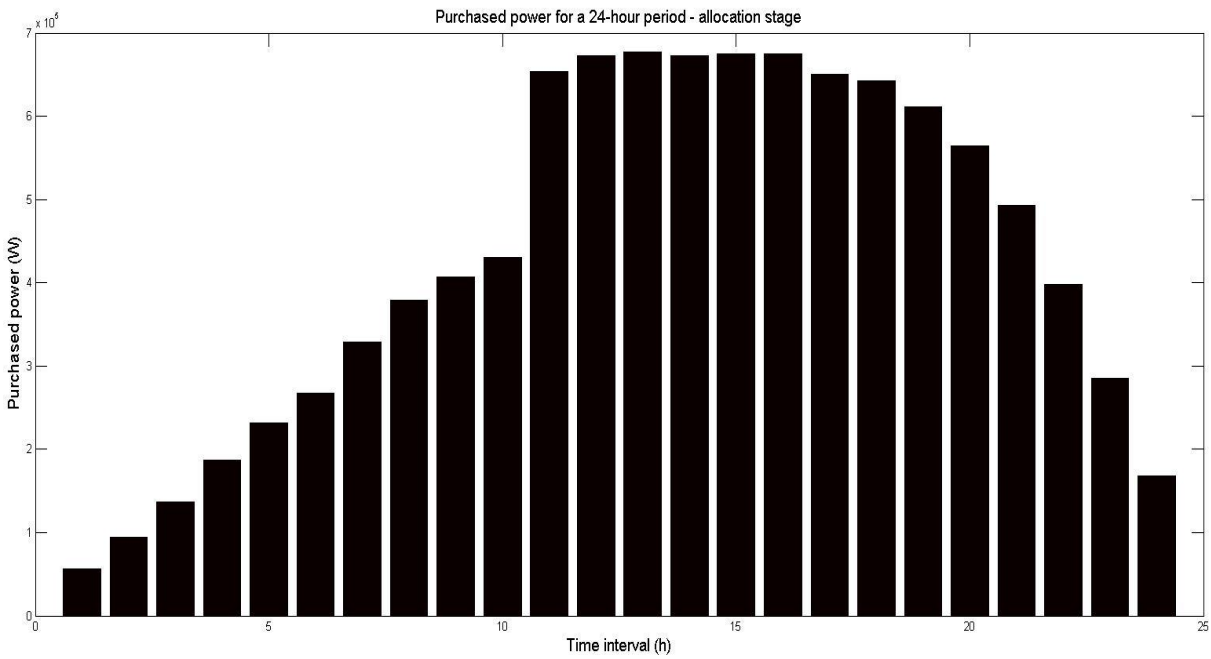


Figure 57 - 24-hour energy distribution for the negotiation stage - Test Case 3

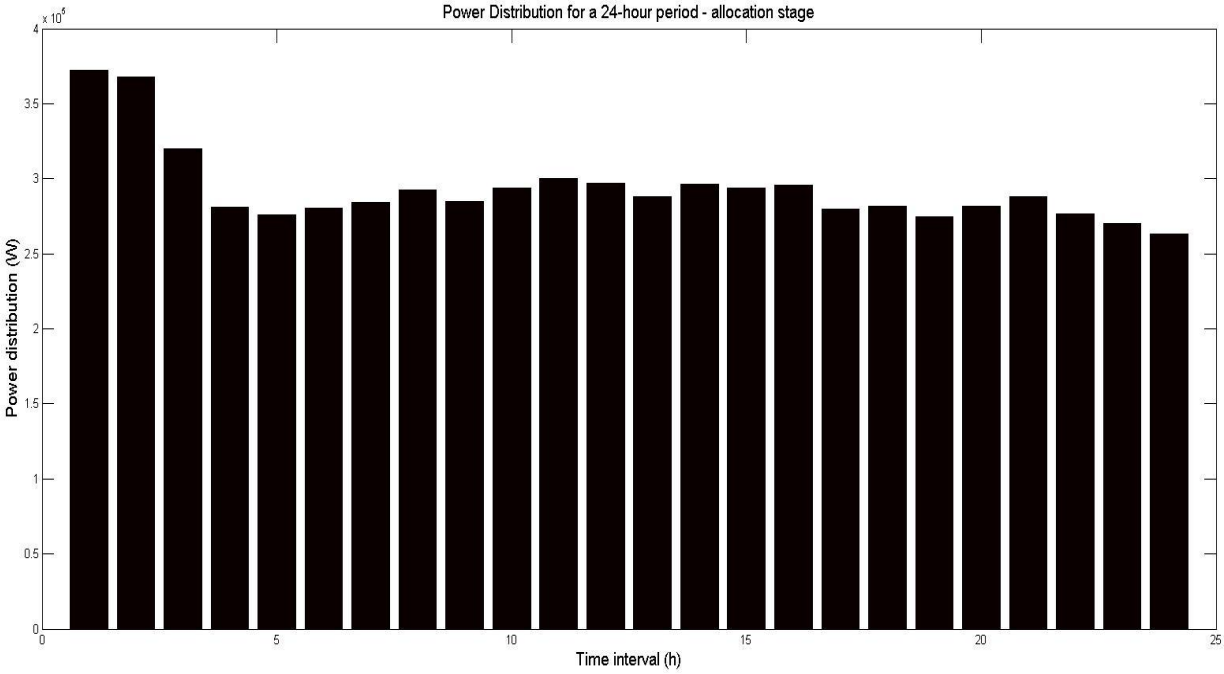


Figure 58 - 24-hour jobs assignment for the negotiation stage - Test Case 3

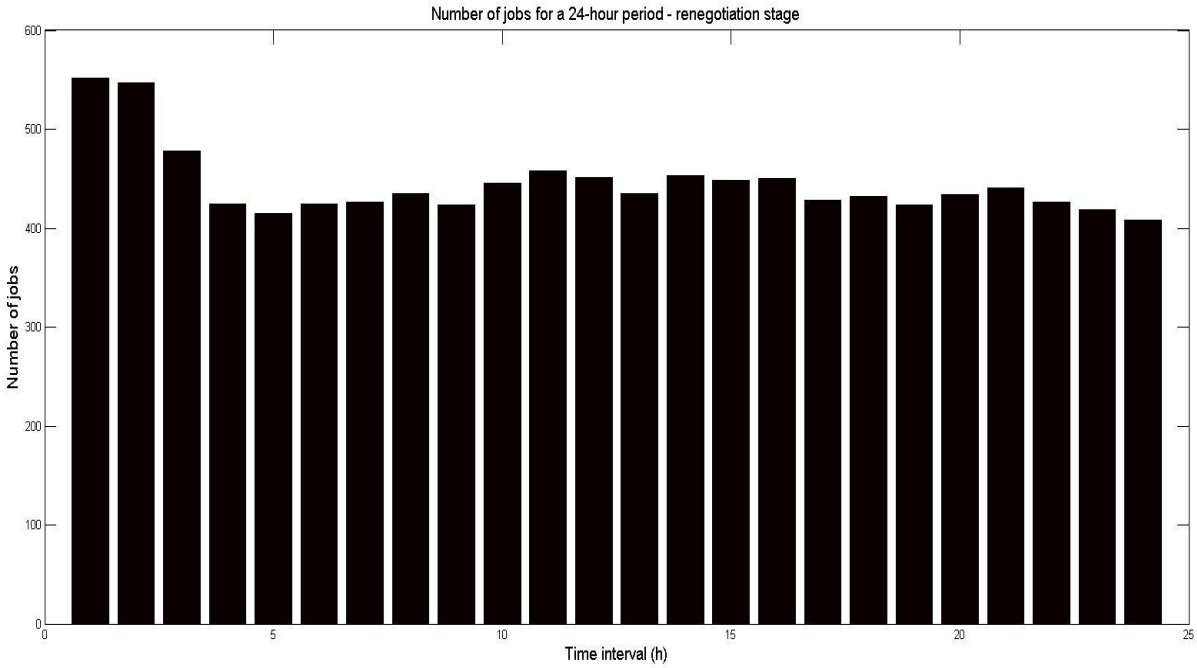


Figure 59 - Expected acquired energy for a 24-hour period, negotiation stage - Test Case 3

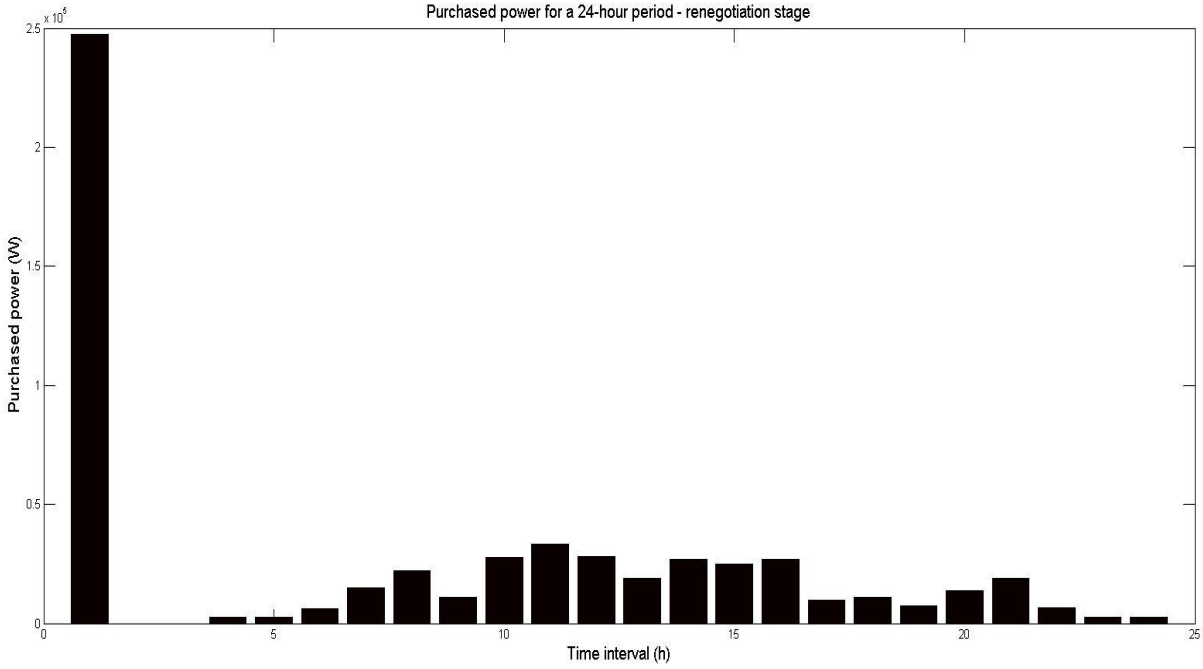


Figure 60 - Energy expenditure estimated for each sub-region, prediction and negotiation stages - Test Case 3

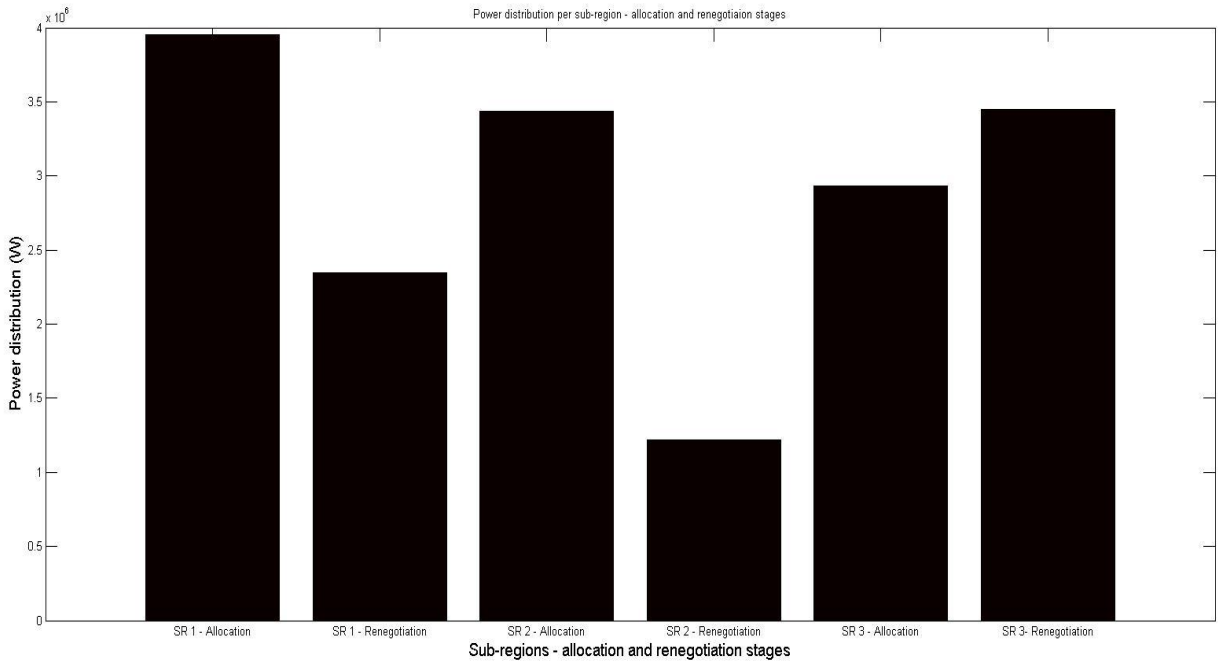
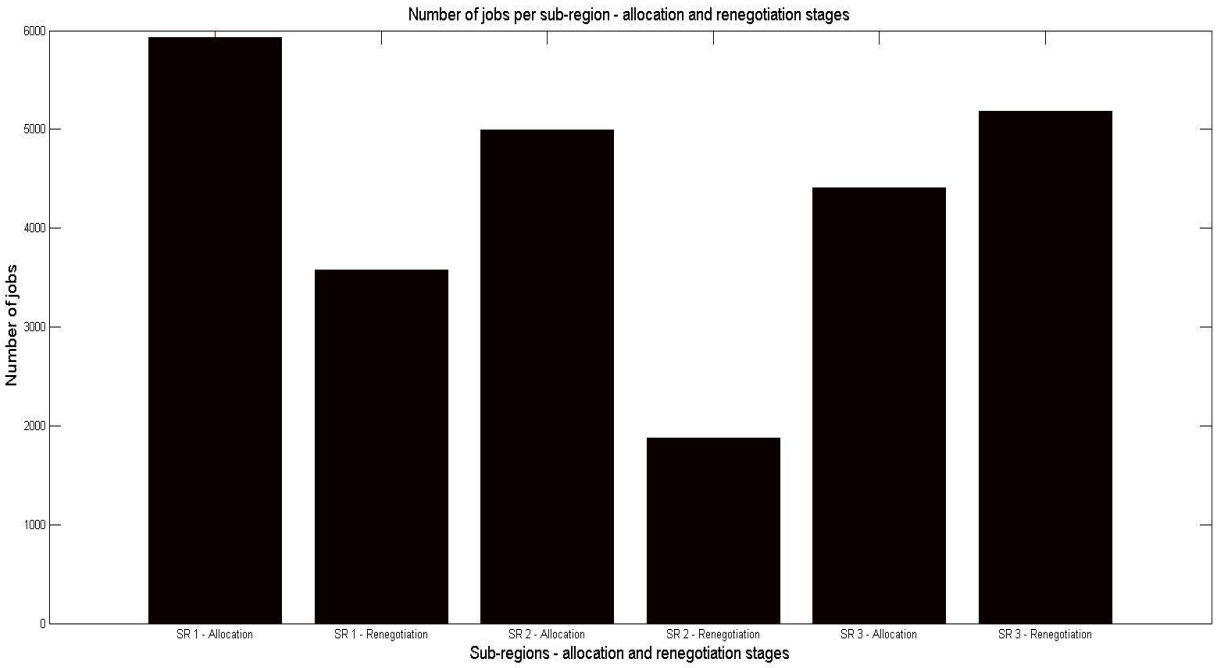


Figure 61 - Jobs assignment for each sub-region, prediction and negotiation stages - Test Case 3



Test Case 4 - 995 reserved plans, 1013 on-demand plans and 992 flexible plans

Figure 62 - 24-hour energy distribution for the prediction stage - Test Case 4

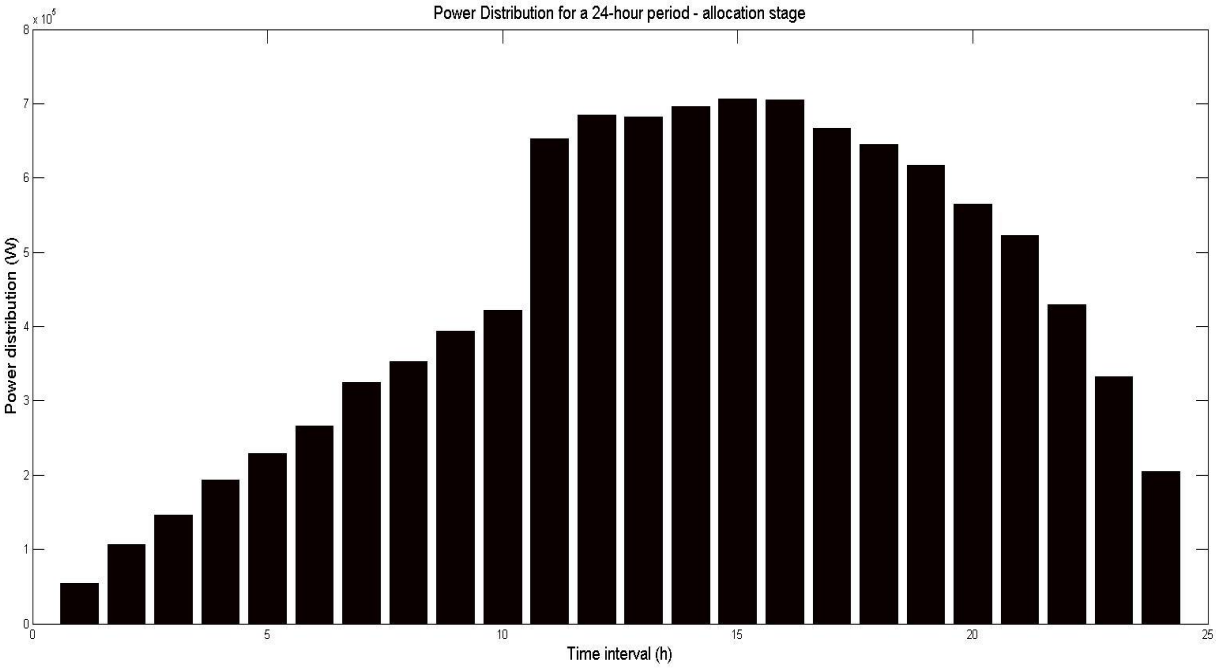


Figure 63 - 24-hour jobs assignment for the prediction stage - Test Case 4

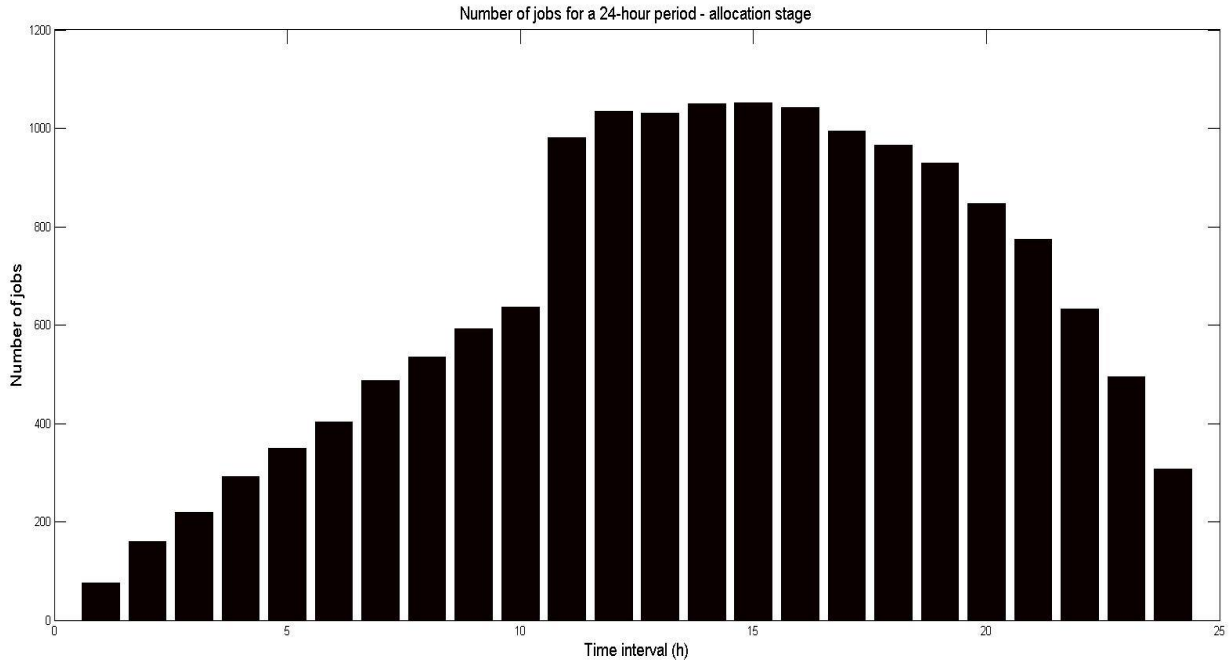


Figure 64 - Expected acquired energy for a 24-hour period, prediction stage - Test Case 4

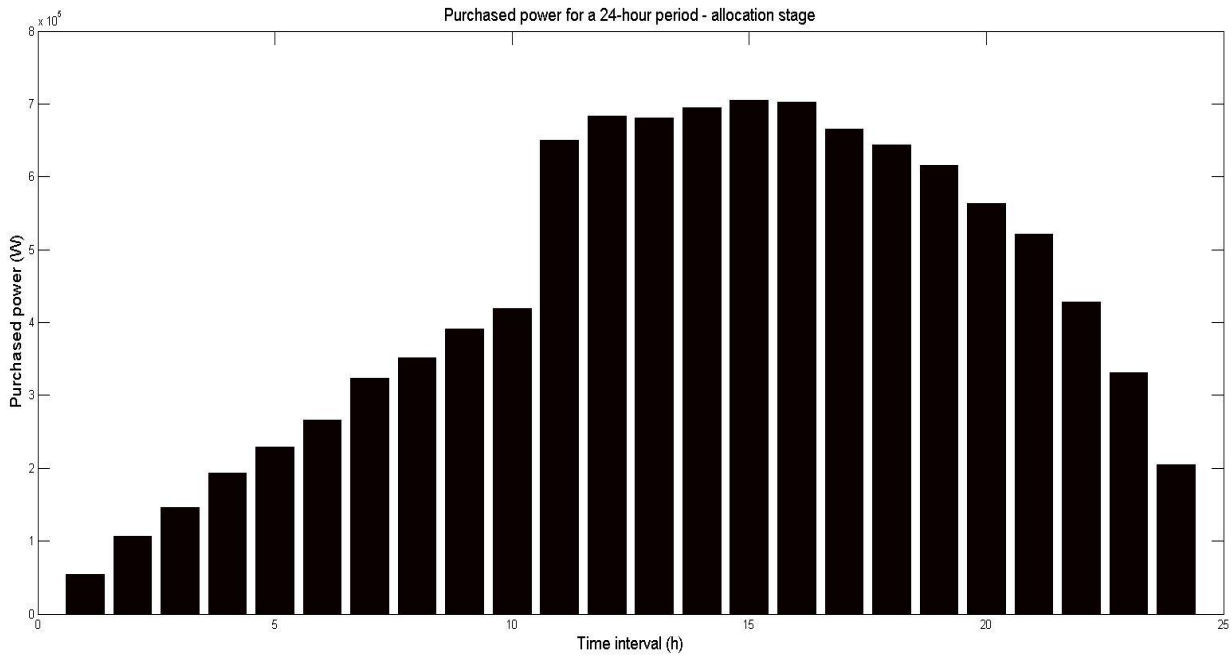


Figure 65 - 24-hour energy distribution for the negotiation stage - Test Case 4

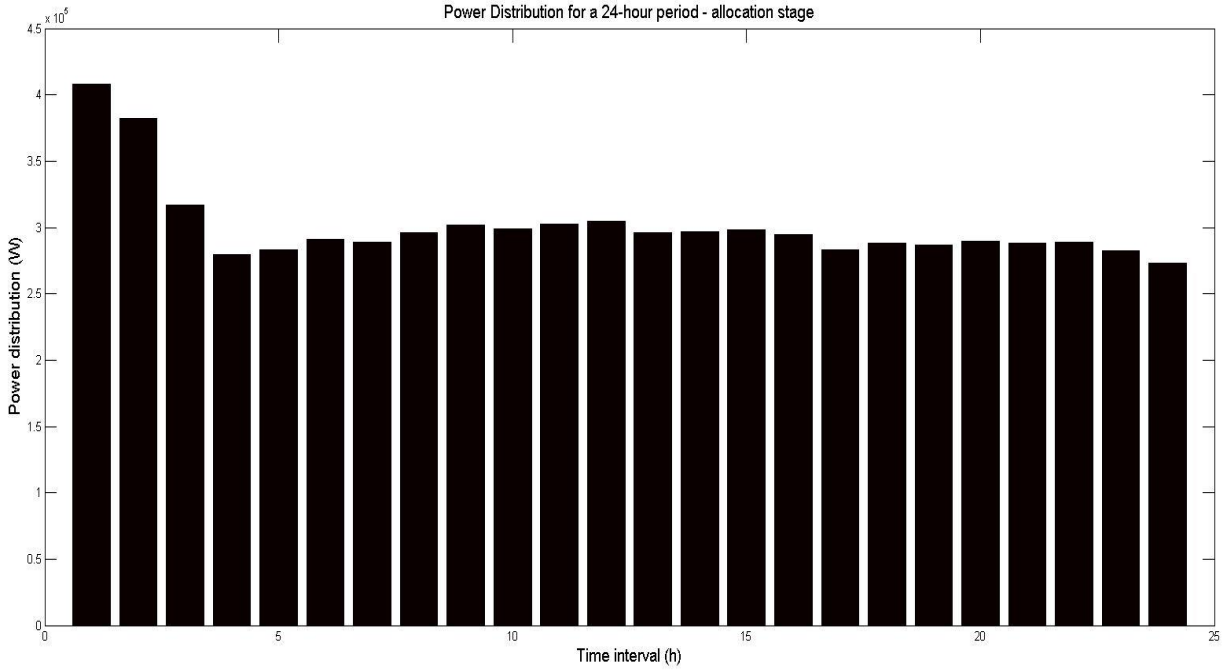


Figure 66 - 24-hour jobs assignment for the negotiation stage - Test Case 4

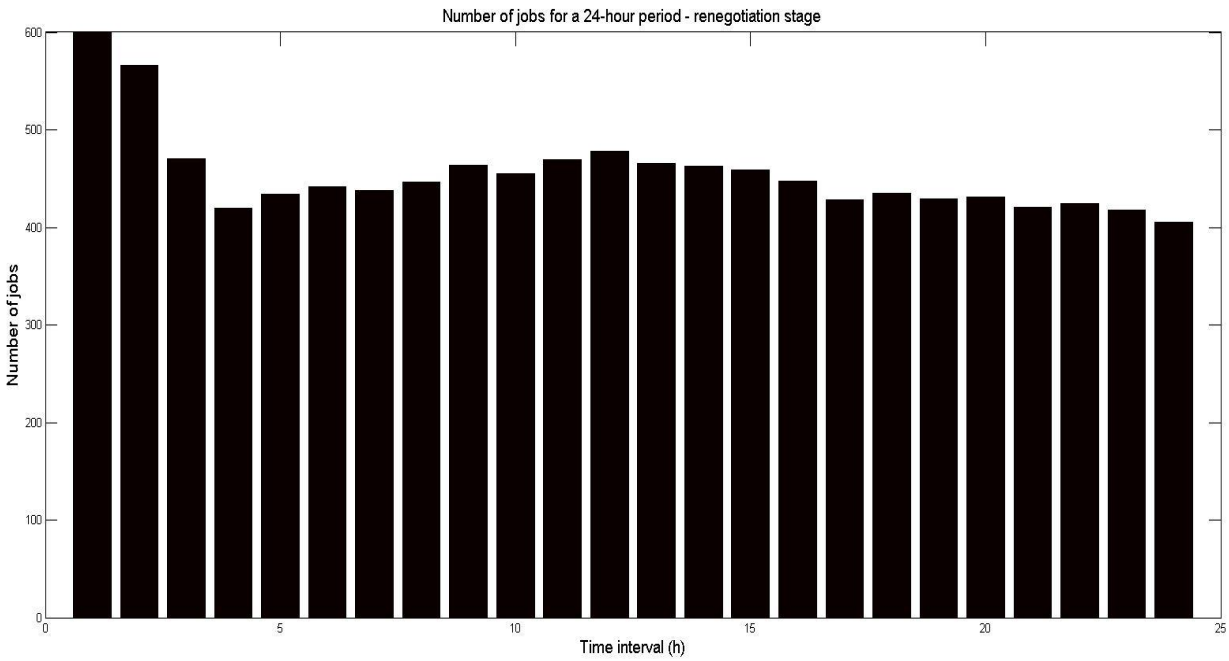


Figure 67 - Expected acquired energy for a 24-hour period, negotiation stage - Test Case 4

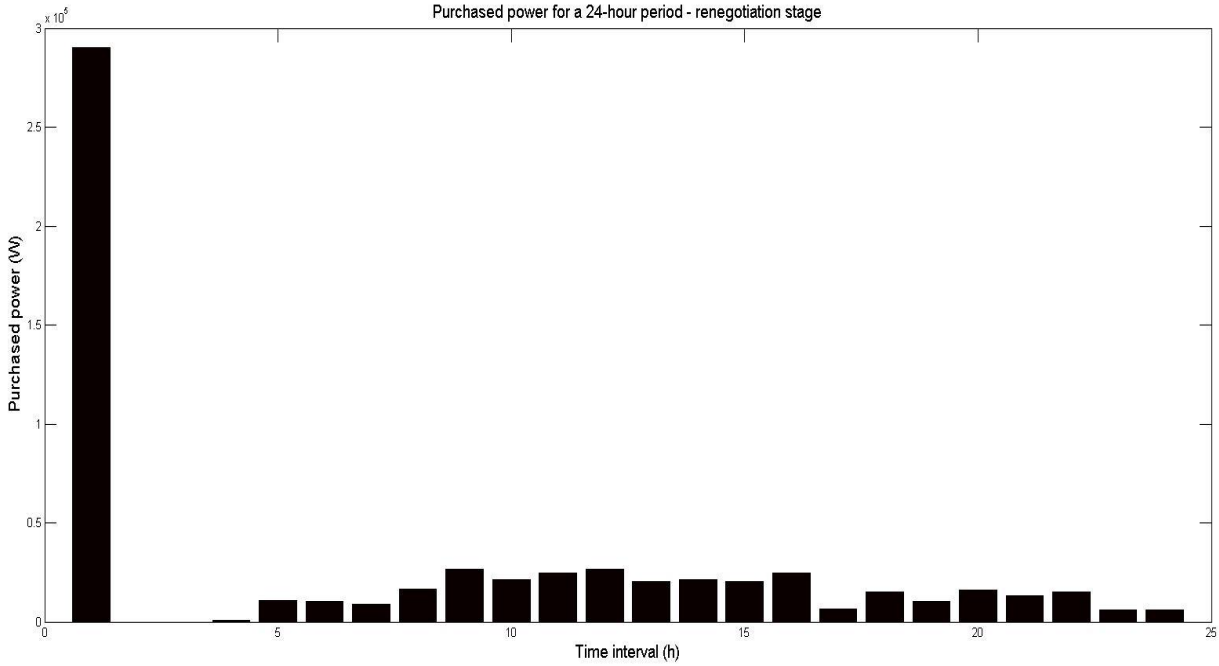


Figure 68 - Energy expenditure estimated for each sub-region, prediction and negotiation stages - Test Case 4

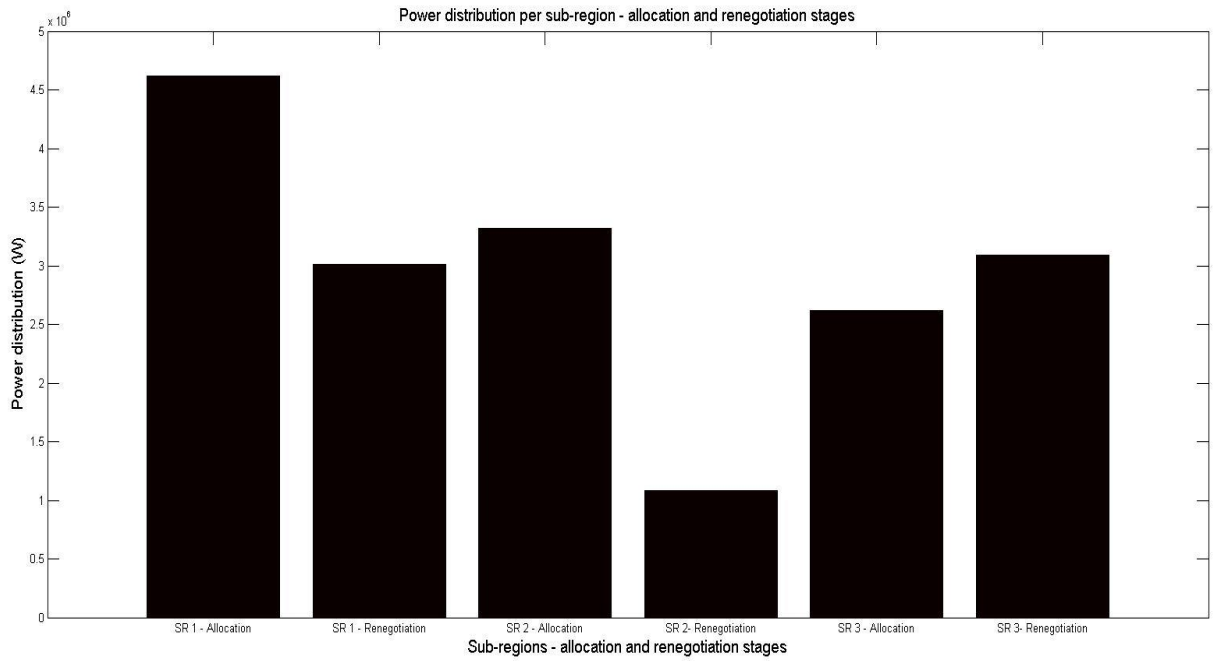
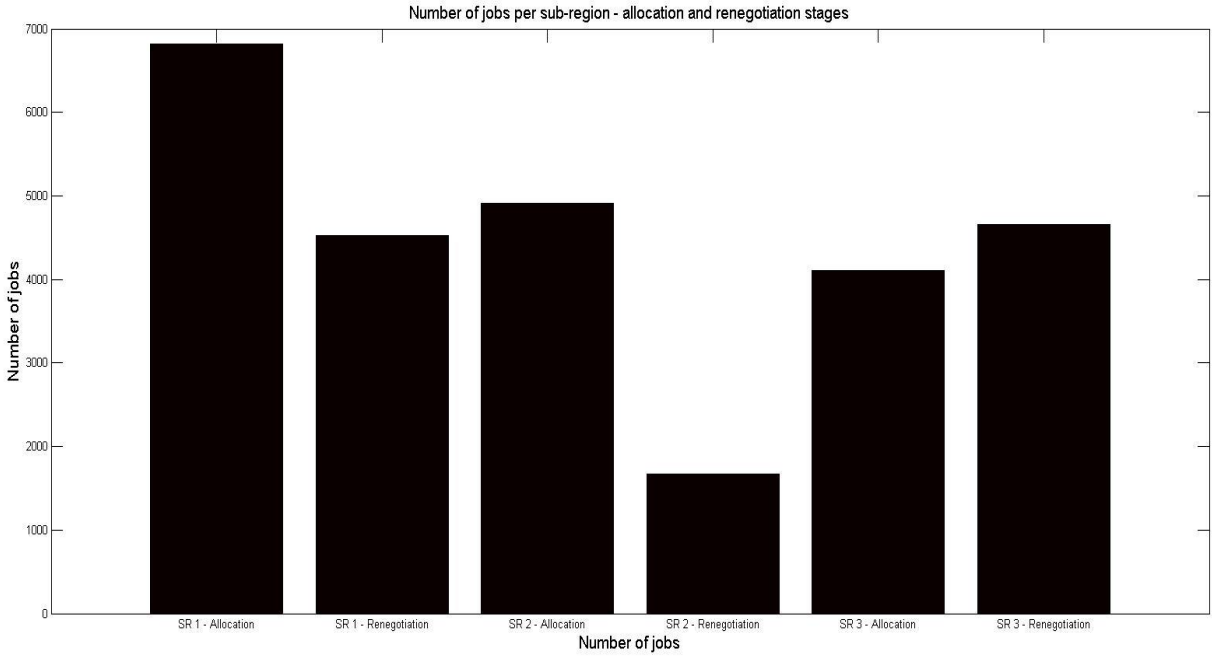


Figure 69 - Jobs assignment for each sub-region, prediction and negotiation stages - Test Case 4



Test Case 5 - 987 reserved plans, 990 on-demand plans and 1023 flexible plans

Figure 70 - 24-hour energy distribution for the prediction stage - Test Case 5

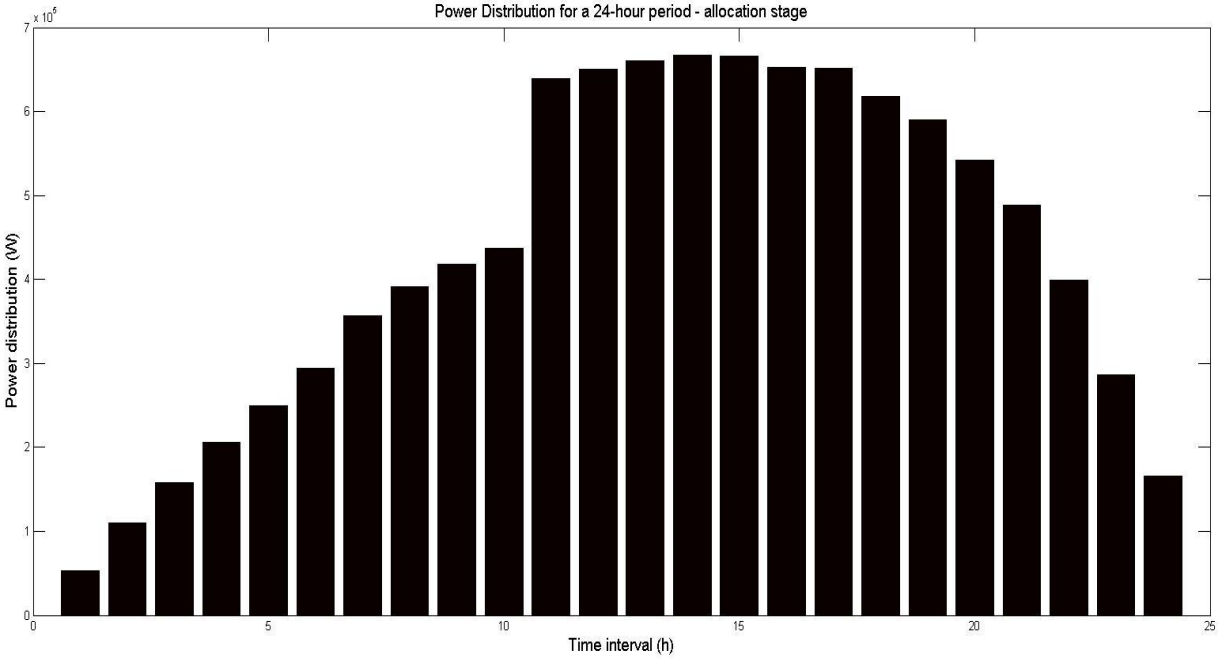


Figure 71 - 24-hour jobs assignment for the prediction stage - Test Case 5

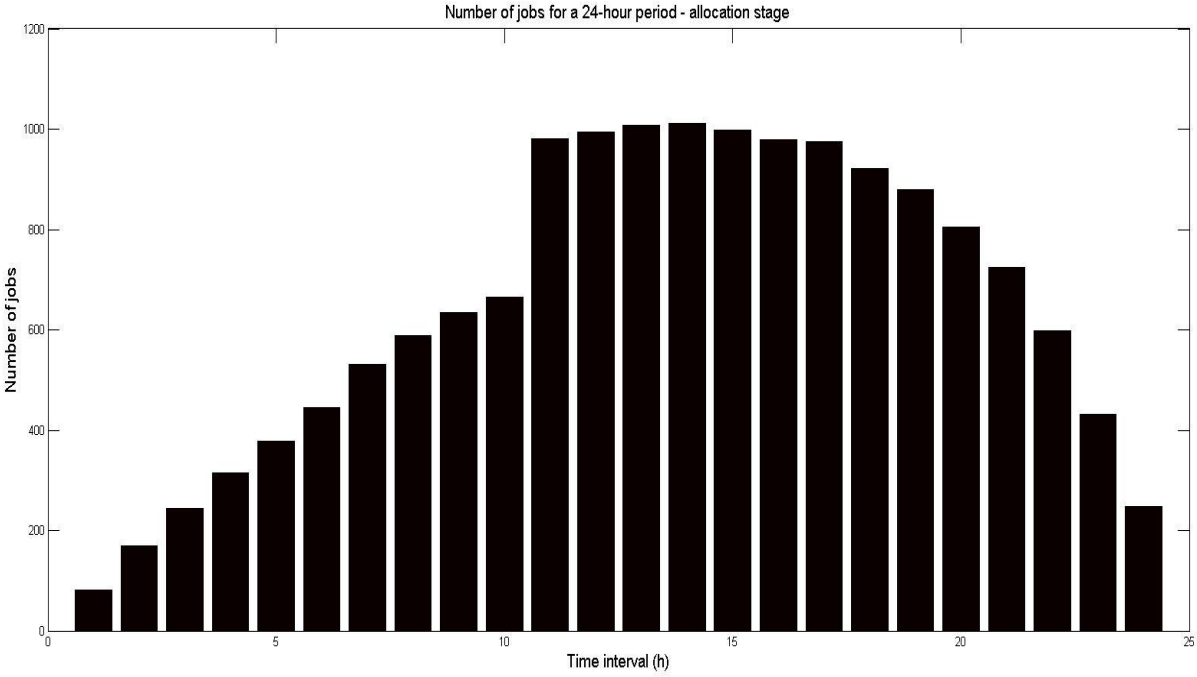


Figure 72 - Expected acquired energy for a 24-hour period, prediction stage - Test Case 5

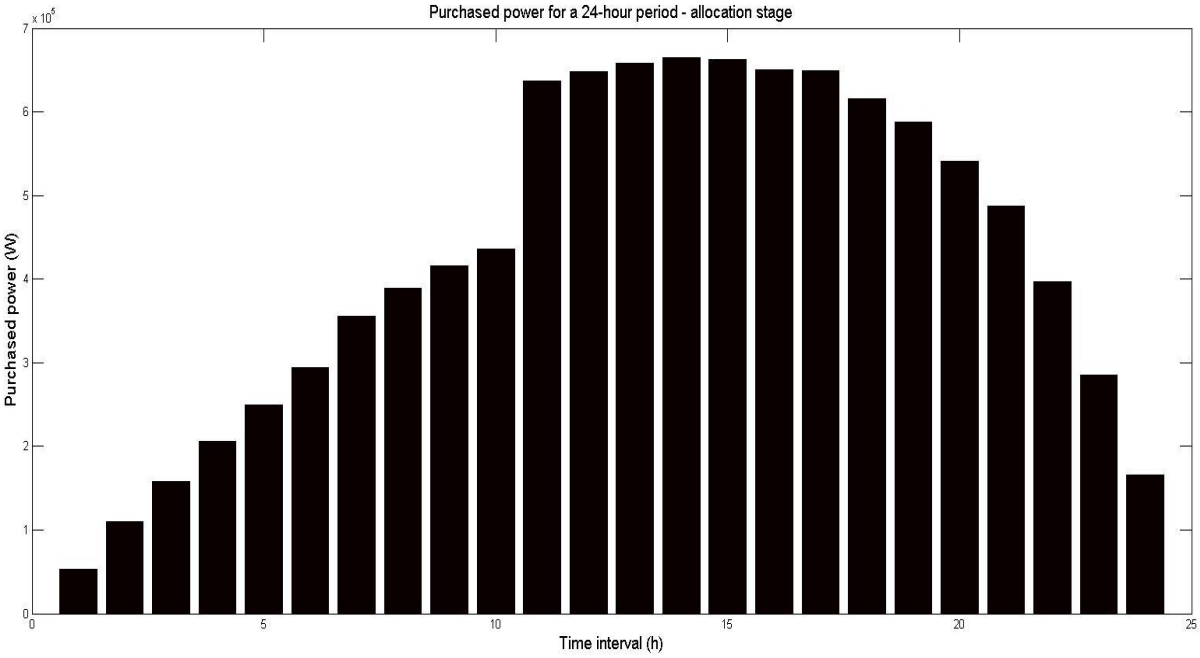


Figure 73 - 24-hour energy distribution for the negotiation stage - Test Case 5

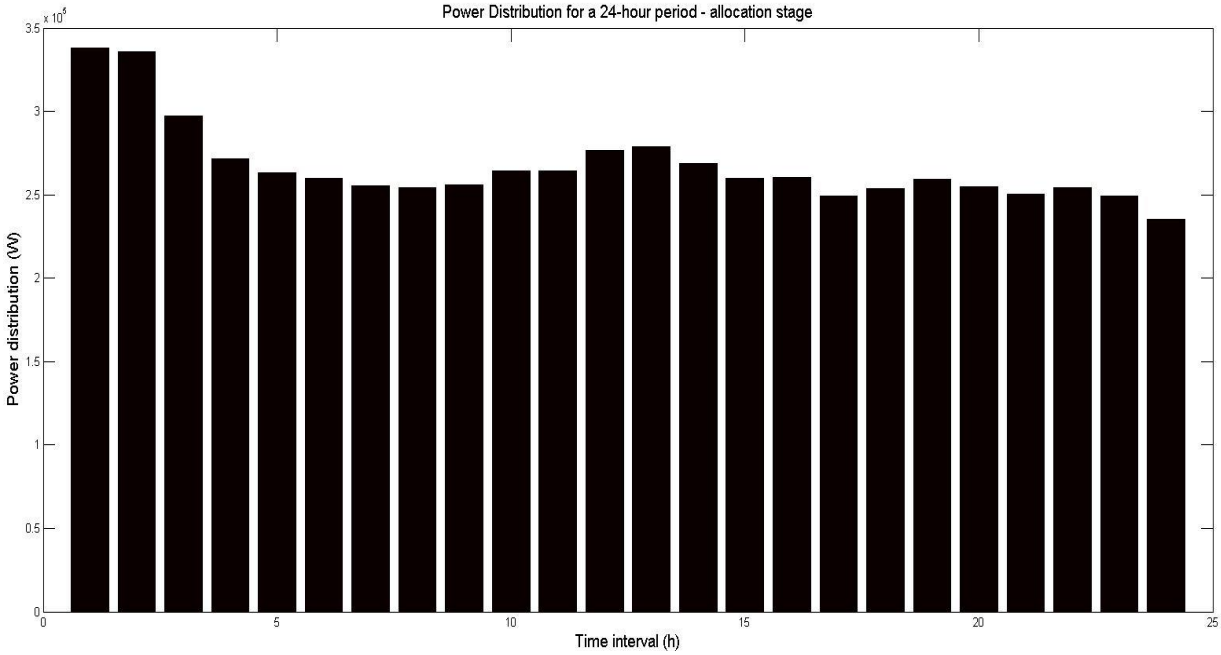


Figure 74 - 24-hour jobs assignment for the negotiation stage - Test Case 5

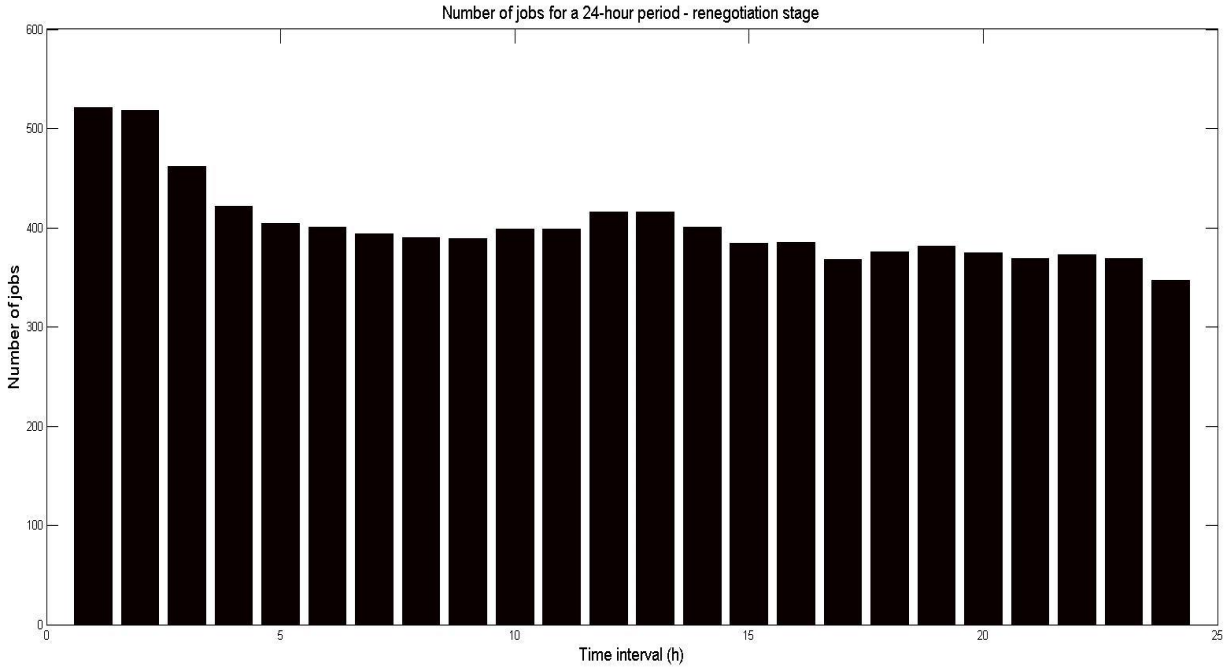


Figure 75 - Expected acquired energy for a 24-hour period, negotiation stage - Test Case 5

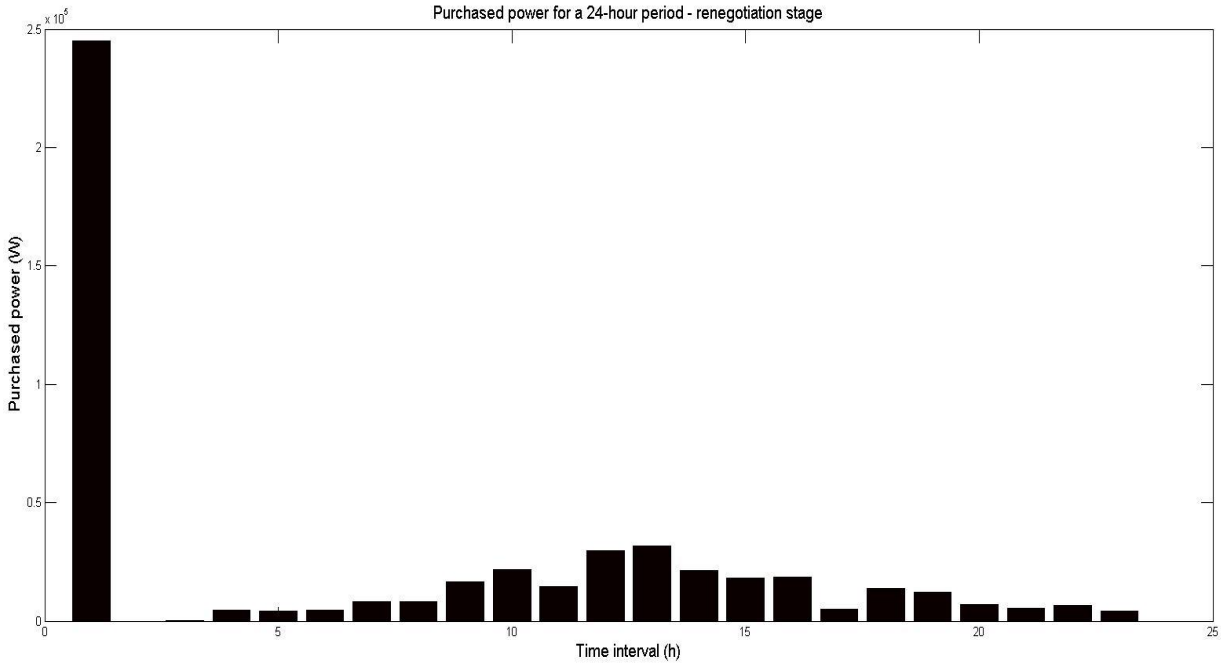


Figure 76 - Energy expenditure estimated for each sub-region, prediction and negotiation stages - Test Case 5

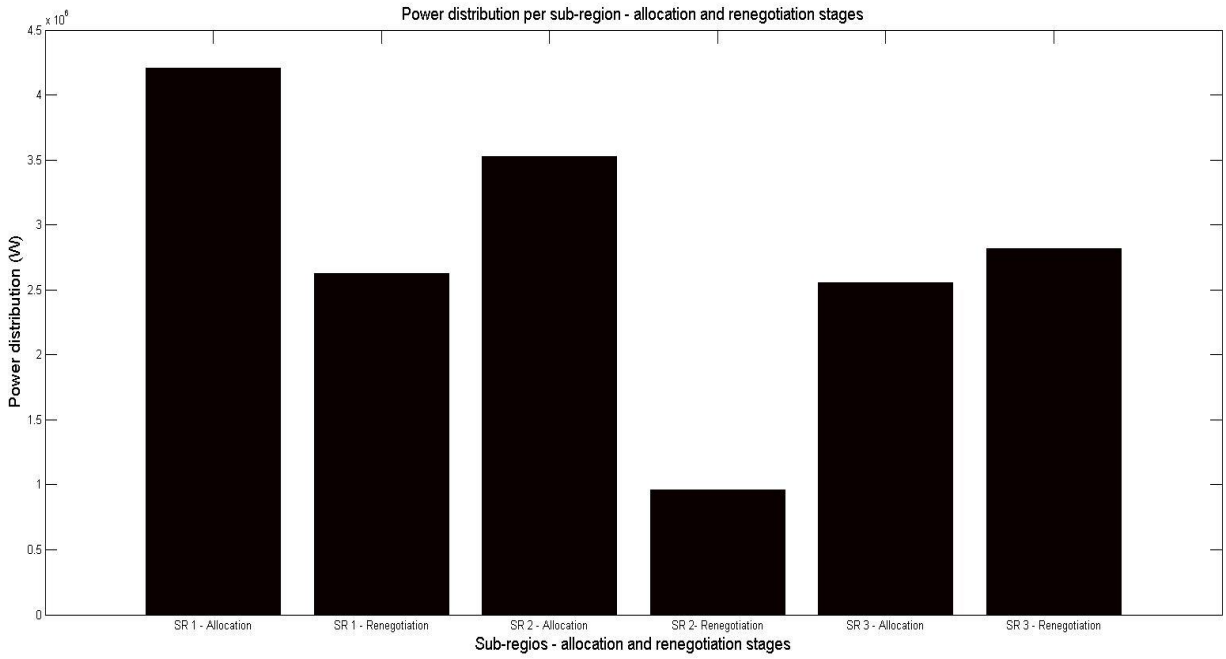
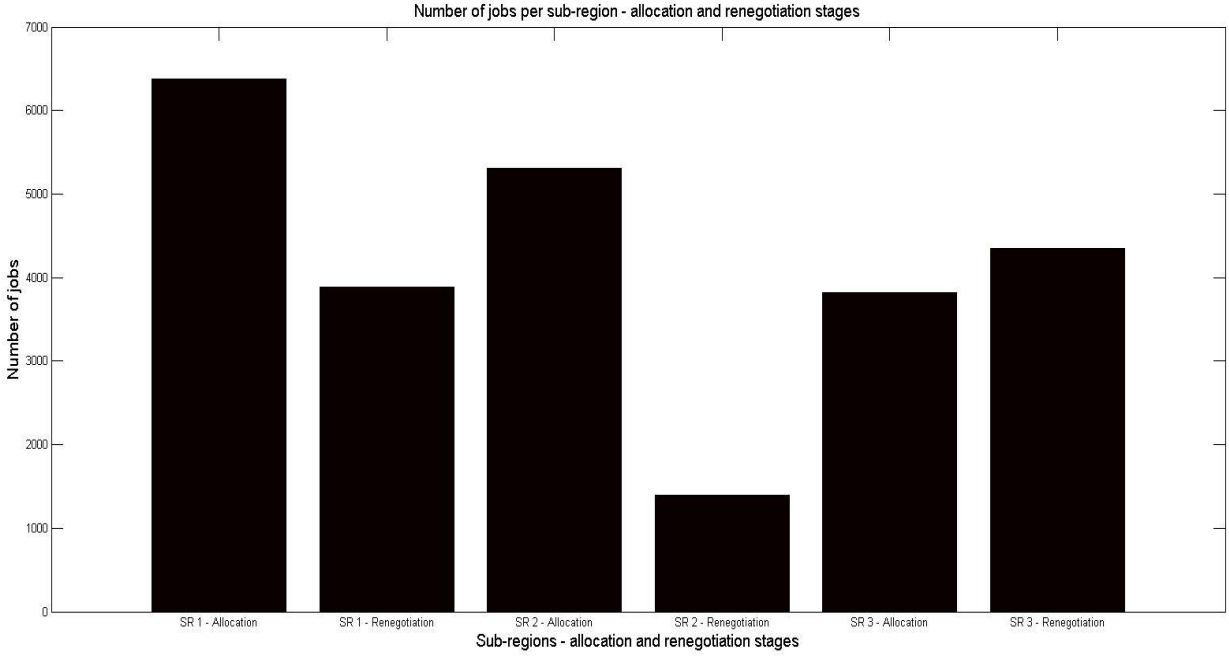


Figure 77 - Jobs assignment for each sub-region, prediction and negotiation stages - Test Case 5



Test Case 6 - 1015 reserved plans, 989 on-demand plans and 996 flexible plans

Figure 78 - 24-hour energy distribution for the prediction stage - Test Case 6

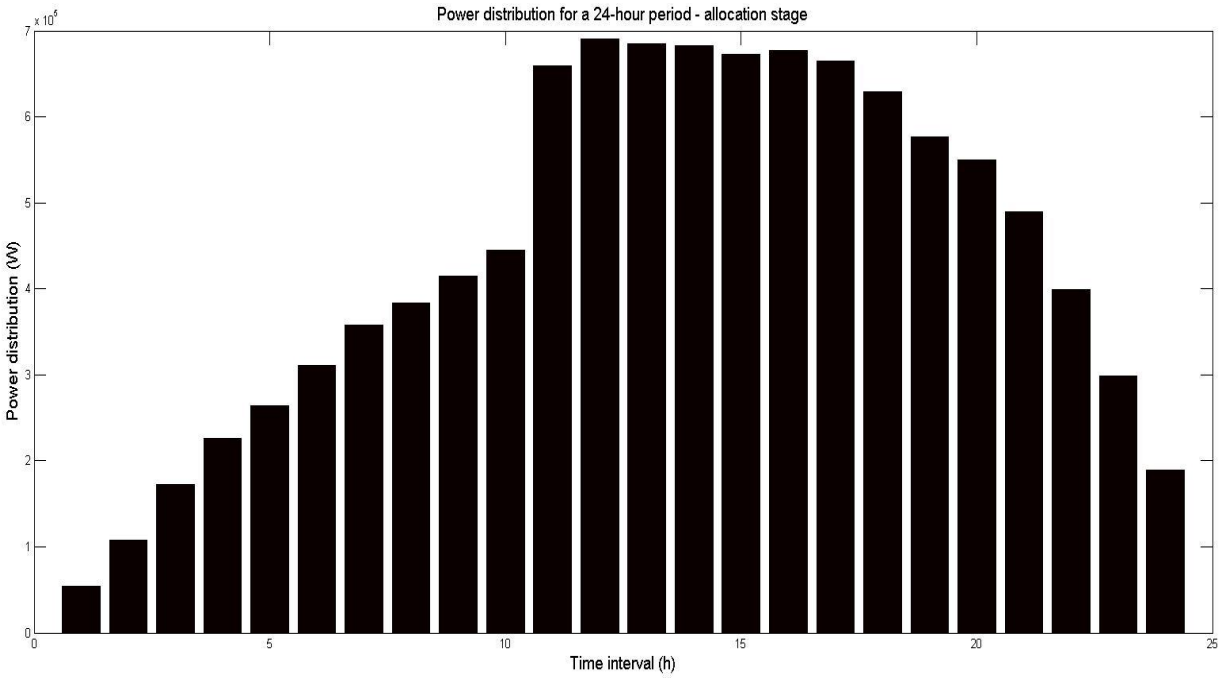


Figure 79 - 24-hour jobs assignment for the prediction stage - Test Case 6

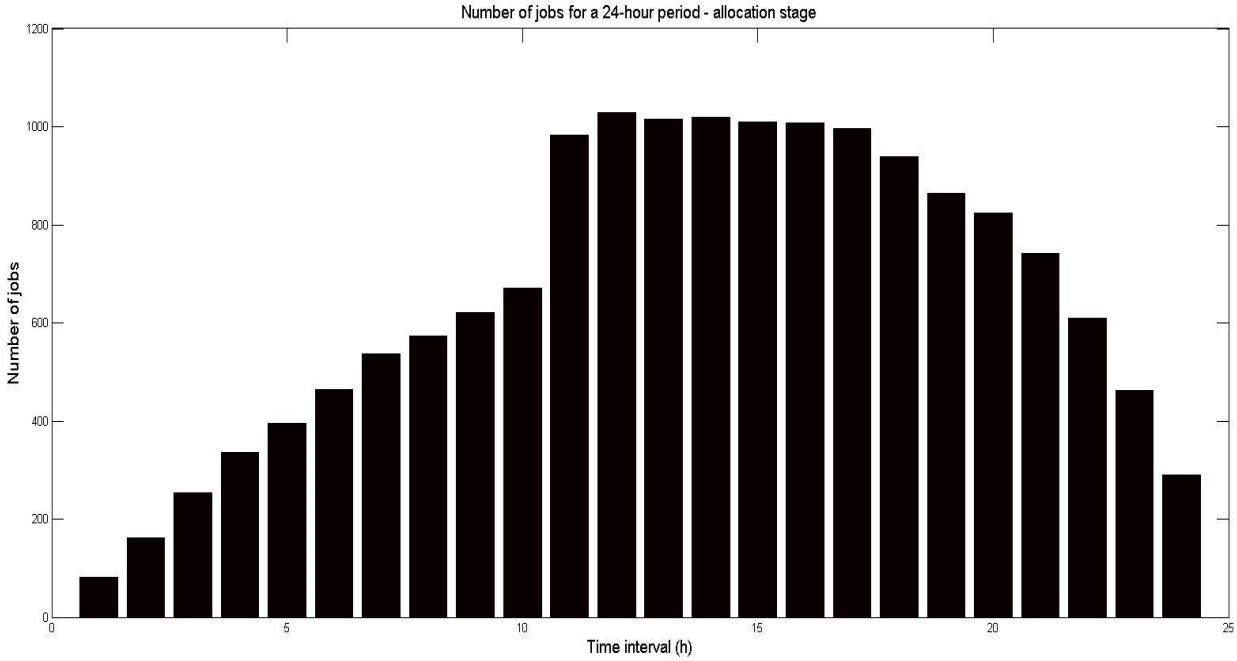


Figure 80 - Expected acquired energy for a 24-hour period, prediction stage - Test Case 6

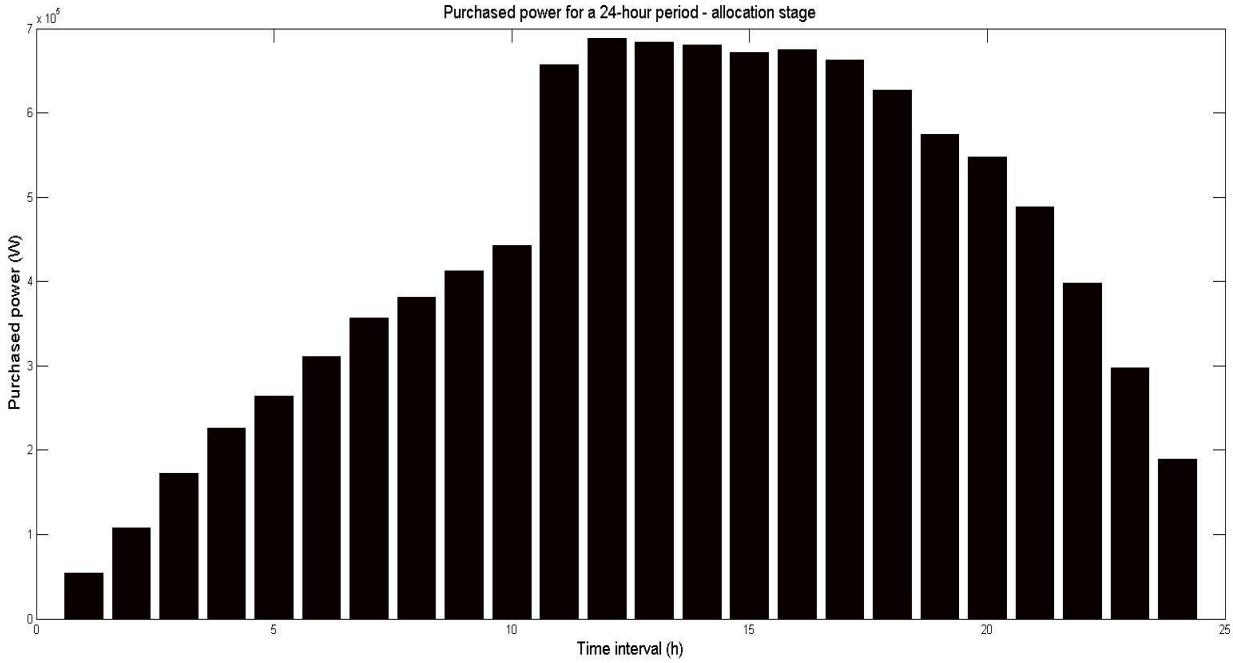


Figure 81 - 24-hour energy distribution for the negotiation stage - Test Case 6

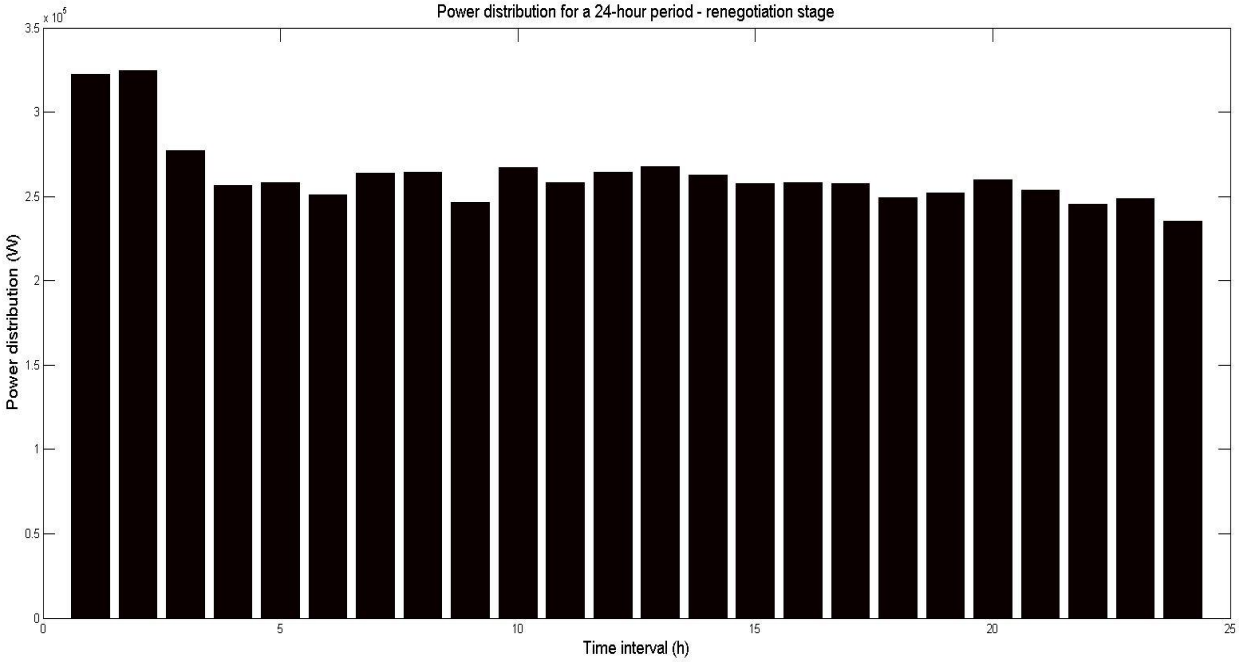


Figure 82 - 24-hour jobs assignment for the negotiation stage - Test Case 6

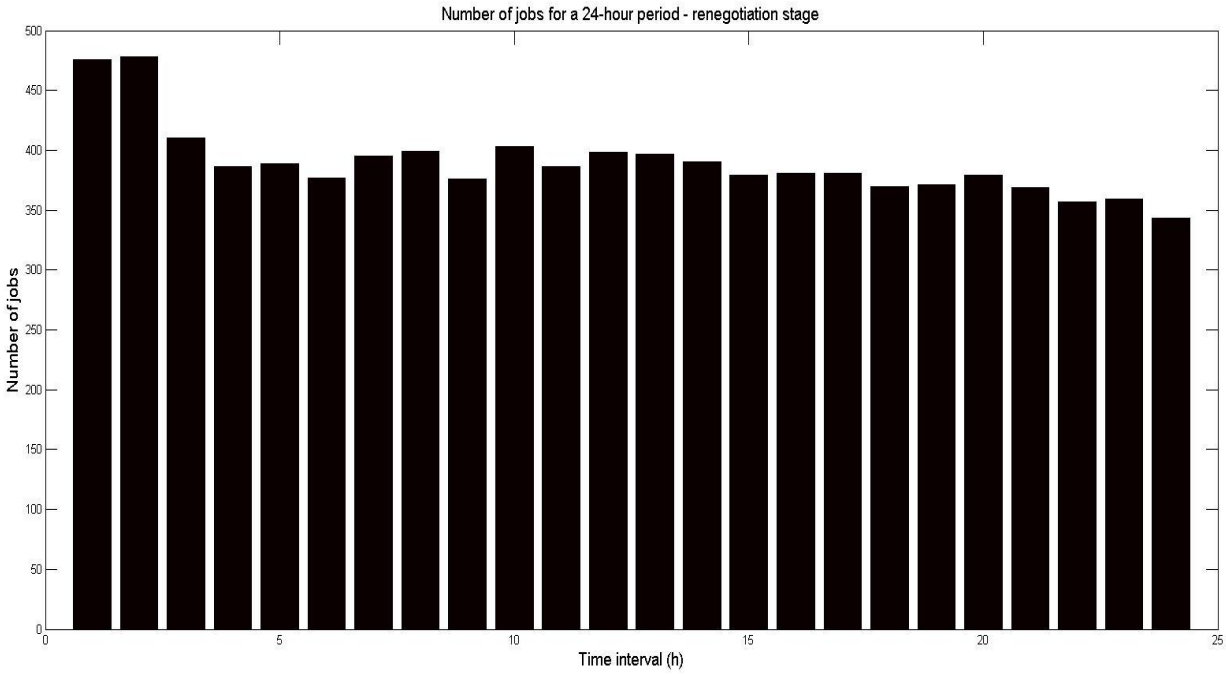


Figure 83 - Expected acquired energy for a 24-hour period, negotiation stage - Test Case 6

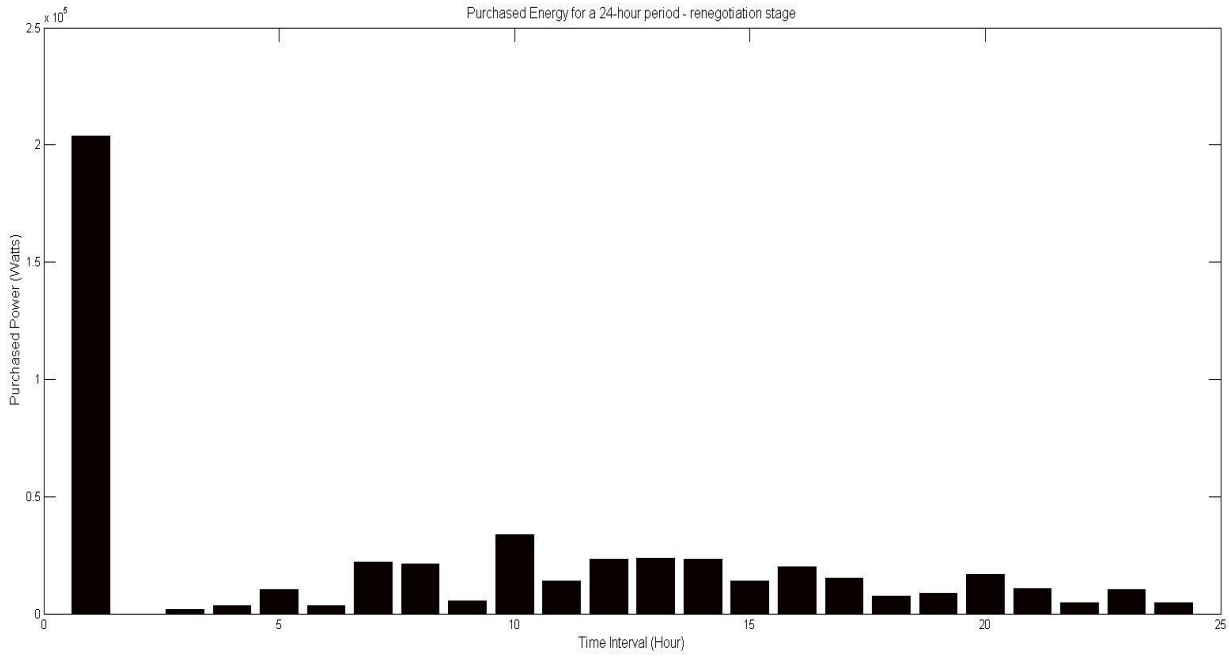


Figure 84 - Energy expenditure estimated for each sub-region, prediction and negotiation stages - Test Case 6

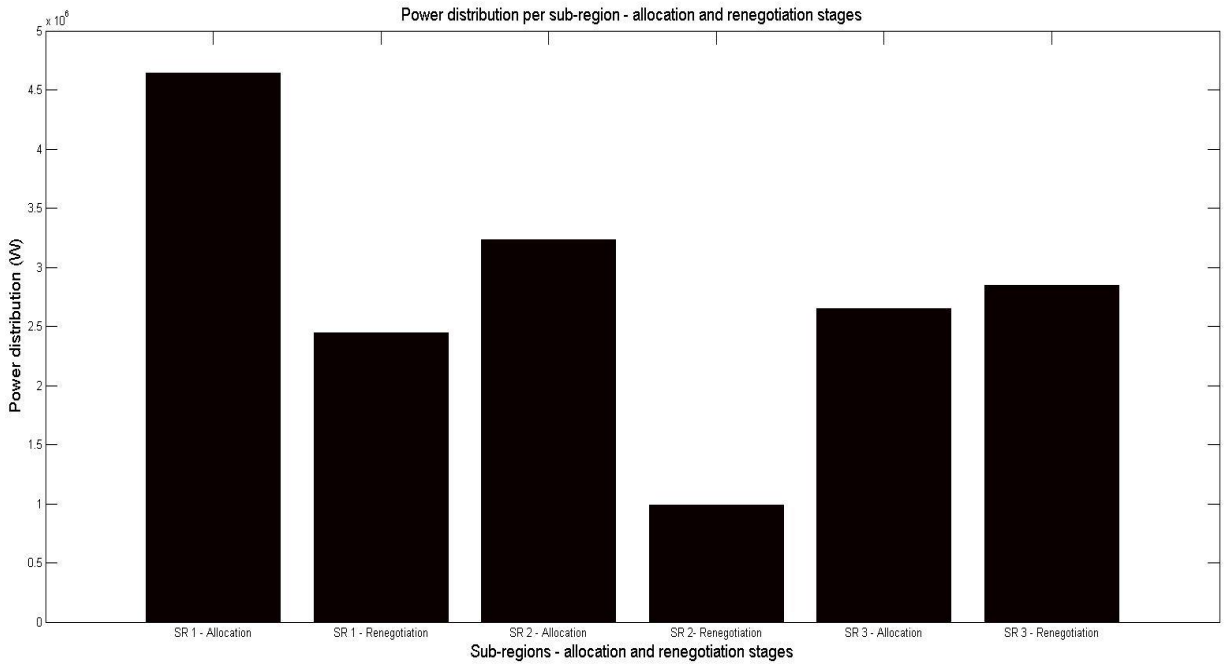


Figure 85 - Jobs assignment for each sub-region, prediction and negotiation stages - Test Case 6