Revista de Sistemas de Informação da FSMA
n. 13 (2014) pp. 30-50

SISTEMAS DE INFORMAÇÃO
http://www.fsma.edu.br/si/sistemas.html

TRILHA PRINCIPAL

Operation, Management, Security and Sustainability for Cloud Computing

Carlos B. Westphall, Carla M. Westphall, Fernando L. Koch, Guilherme A. Geronimo, Jorge Werner, Rafael S. Mendes, Paulo F. Silva, Daniel R. Santos, Ricardo F. Souza, Mauro M. Mattos, Sergio R. Villarreal, Rafael Weingartner, Leonardo Defenti, Alexandre A. Flores, Rafael R. Freitas, and Gabriel B. Brascher

***Abstract*— This paper presents some scope, context, proposals and solutions related with the following topics: Decision-Theoretic Planning for Cloud Computing; An Architecture for Risk Analysis in Cloud; Risk-based Dynamic Access Control for a Highly Scalable Cloud Federation; Challenges of Operationalizing PACS on Cloud Over Wireless Networks; Environment, Services and Network Management for Green Clouds; Provisioning and Resource Allocation for Green Clouds; and Optimizing Green Clouds through Legacy Network Infrastructure Management.**

***Index Terms* — Operation, Management, Security, Sustainability, Cloud Computing.**

I. Introduction

This paper presents a mathematical model of decision planning for autonomic Cloud Computing based on the decision-theoretic planning model. It uses Markov decision process on the cloud manager to evaluate decisions and manage the Cloud environment. Also, it contributes to the state-of-art of Cloud Computing approaching the planning phase of the autonomic process with a mathematical model, considering two important factors, (1) the uncertainty of action's results and (2) the utility of the actions. Both factors are needed when dealing with complex systems as a Cloud. [1].

Cloud computing offers benefits in terms of availability and cost, but transfers the responsibility of information security management for the cloud service provider. Thus, the consumer loses control over the security of their information and services. This factor has prevented the migration to cloud computing in many businesses. This paper proposes a model where the cloud consumer can perform risk analysis on providers before and after contracting the service. The proposed model establishes the responsibilities of three actors: Consumer, Provider and Security Labs. The inclusion of actor Security Labs provides more credibility to risk analysis making the results more consistent for the consumer [2].

Cloud Computing is already a successful paradigm for distributed computing and is still growing in popularity. However, many problems still linger in the application of this model and some new ideas are emerging to help leverage its features even further. One of these ideas is the cloud federation, which is a way of aggregating different clouds to enable the sharing of resources and increase scalability and availability. One of the great challenges in the deployment of cloud federations is Identity and Access Management. This issue is usually solved by the creation of identity federations, but this approach is not optimal. In this paper, we propose an access control system for a highly scalable cloud federation. The presented system is dynamic and risk-based, allowing the use of cloud federations without the need of identity federations. We also present results of a prototype implementation and show that it is scalable and flexible enough to meet the requirements of this highly dynamic and heterogeneous environment [3].

Clinics and hospitals are acquiring more technological resources to help providing a faster and more precise diagnostic, with the goal of making it more dynamic and effective. This is pushing health institutions to search for more modern equipment, with greater technological features. Besides last generation equipment, another problem faced by these institutions is enabling the connection of physicians to a Picture Archive and Communication Systems (PACS) from anywhere. With the use of communication resources increasingly present in everyday life, like Wireless-Fidelity (Wi-Fi), third generation of mobile telecommunications technology (3G), fourth generation of mobile telecommunications technology (4G), Worldwide Interoperability for Microwave Access (WiMax) and other wireless networks that allow the connection of mobile devices, it becomes easier and cheaper to provide quality medical services at a distance. Diagnoses that needed a doctor to be present, for instance, can now be performed from anywhere, provided there is an Internet connection. Cloud-based PACS is shown to be efficient for archiving medical images, allowing access to exams and reports from anywhere, over wireless networks, regardless of the platform used for access [4].

Green cloud computing aims at a processing infrastructure that combines flexibility, quality of services, and reduced energy utilization. In order to achieve this objective, the management solution must regulate the internal settings to address the pressing issue of data center over- provisioning related to the need to match the peak demand. In this context, we propose an integrated solution for environment, services and network management based on organization model of autonomous agent components. This work introduces the system management model, analyses the system's behavior, describes the operation principles, and presents a case study scenario and some results. We extended CloudSim to simulate the organization model approach and implemented the migration and reallocation policies using this improved version to validate our management solution [5].

The aim of Green Cloud Computing is to achieve a balance between the resource consumption and quality of service. In order to achieve this objective and to maintain the flexibility of the cloud, dynamic provisioning and allocation strategies are needed to regulate the internal settings of the cloud to address oscillatory peaks of workload. In this context, we propose strategies to optimize the use of the cloud resources without decreasing the availability. This work introduces two hybrid strategies based on a distributed system management model, describes the base strategies, operation principles, tests, and presents the results. We combine existing strategies to search their benefits. To test them, we extended CloudSim to simulate the organization model upon which we were based and to implement the strategies, using this improved version to validate our solution. Achieving a consumption reduction up to 87% comparing Standard Clouds with Green Clouds, and up to 52% comparing the proposed strategy with other Green Cloud Strategy [6].

This paper is structured as follows (section – title): Section II - Decision-Theoretic Planning for Cloud Computing; Section III - An Architecture for Risk Analysis in Cloud; Section IV - Risk-based Dynamic Access Control for a Highly Scalable Cloud Federation; Section V – Challenges of Operationalizing PACS on Cloud Over Wireless Networks; Section VI - Environment, Services and Network Management for Green Clouds; Section VII - Provisioning and Resource Allocation for Green Clouds; Section VIII - Optimizing Green Clouds through Legacy Network Infrastructure Management; and Section IX – Conclusions.

II. Decision-Theoretic Planning for Cloud Computing

A. Scope and Context

The decision-theoretic planning (DTP) problems were extensively researched during the last decades. The main problem with the decision-theoretic (or probabilistic) approach for the planning phase in autonomic computing is the need to provide extensive information about the transitions between system states. However, with the arise of Cloud Computing (CC), sensor networks and other technologies that enabled the monitoring and collection of large volumes of data, the information became abundant and the recommendation of utility to solve the contradictions between rules on large rule-based polices [8], [9] must be taken seriously. On big data environments, the DTP problems no longer exist, enabling its application for the planning phase on the autonomic loop.

This work presents a model that plans actions for CC management systems using a decision-theoretic approach. It contributes to the state-of-art in CC research by:

• (i) Adapting the decision-theoretic models, which was based on Markov Decision Process (MDP), to use in the planning phase of the autonomic management loop;
• (ii) Introducing decision-theoretic and MDP for planning in CC;
• (iii) Applying mathematical models on a concrete decision making scenario for self-configuration of CloudStack [10].

B. Literature Review

After some years, the definition of CC that has grown in acceptance was created by NIST [11]:

Cloud computing is a model for enabling ubiqui- tous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. This cloud model is composed of five essential characteristics, three service models, and four deployment models.

Another important contribution to CC also can be found in [12]; "Cloud Computing: the need for monitoring", where are stated some useful concepts to understand the fundamental elements of a Cloud.

As stated in [13], to deal with a complex system like a Cloud, it is necessary to be able to accurately capture its states.

Beyond the well-known CC characteristics, like on-demand self-service, broad network access, resource pooling, rapid elasticity, measured service, etc. [11], [14], it is important to highlight the stakeholder heterogeneity characteristic. This characteristic is poorly defined and appears in some works like stakeholder, actors or roles.

In [14] the stakeholders are defined as roles:

• Cloud Consumer;
• Cloud Provider;
• Cloud Auditor;
• Cloud Broker;
• Cloud Carrier.

Litoiu at al. [15] presents four type of stakeholders: infras-tructure providers, platform providers, application providers and end users, although, it does not describe these stakeholders roles. In the same paper [8], there is a change on the terms used to present the stakeholders; the term actors is used instead of stakeholders. It places the actors in function of service models: Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS); introducing a whole different set of actors as layer owners, that are: IaaS owner, PaaS owner, SaaS owner and end users.

Leimeister et al. [16] defines five actors in the CC value net- work: Customer, Service providers, Infrastructure providers, Aggregate services providers (aggregators), Platform provider and Consulting.

In [17], it can be observed a different definition of roles on Cloud environments. The work defines these roles as stakeholders and present the following concepts: Consumers,

Providers, Enablers and Regulators.

Letaifa et al. [18] present a definition of actors and roles in cloud computing systems as: Vendors, Developers and End users.

In Tan et al. [19], although the work focus is adoption (or not) of SaaS, it classifies stakeholders in three categories: SaaS infrastructure provider, SaaS provider and SaaS consumer.

There is no concise definition of CC stakeholders and interests. Furthermore, those distinct definitions may indicate that each Cloud implementation requires a specific stakeholder's modeling.

C. *Autonomic Computing*

The Autonomic Computing (AC) concept was based on the human nervous system, which regulates critical functions such as heart rate and body temperature, in the absence of a conscious brain [20]. AC systems have many common points with Expert Systems (ES) but are less generic, applied to management and control of wide computational systems, while the ES are applied in a more generic way. The AC differs from ES principally when it addresses the "action taking", that was unusual in ESs, as stated in [21].

AC systems are based on MAPE-K control cycle, that consists in Monitor, Analyze, Plan, Execute and Knowledge elements. Figure 1 shows the MAPE-K life cycle.

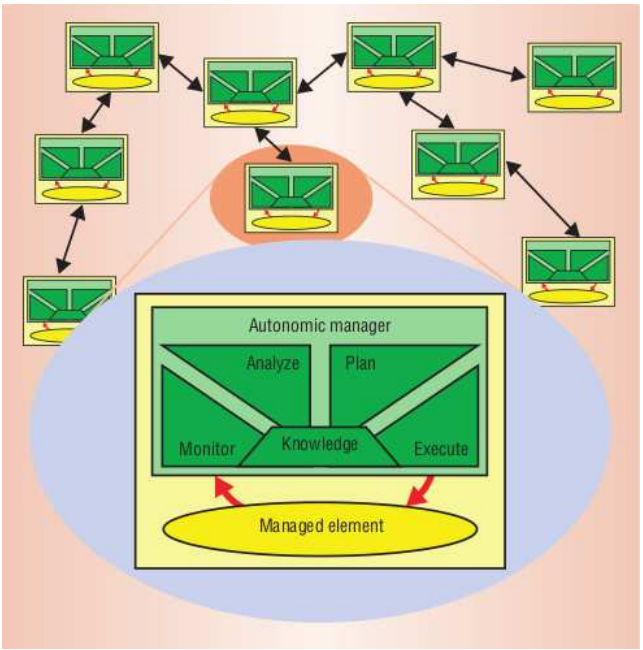

Figure 1. Structure of and autonomic agent [1].

An autonomic system, as shown in [20], to be able to perform self-management, must present four main abilities:

• self-configuration - the ability of configure itself according to high-level policies;

• self-optimization - the capacity of optimize its use of resource;

• self-protection - autonomic systems must protect itself from malicious or incorrect user behavior;

• self-healing - the ability of detect, diagnoses and fix problems.

In [9], the abilities were extended, adding four attributes of autonomic systems:

• self-awareness - the system must be aware of its internal state;

• self-situation - it should detect its current external operating conditions;

• self-monitoring - it has to detect changing circumstances;

• self-adjustment - it has to adapt accordingly to external or internal changes.

D. *Markov Decision Process*

Broadly speaking, it can be said that the planning techniques developed in the Artificial Intelligence domain are concerned to obtain a course of actions which conducts the agent to a goal state or to an improvement in its condition. In deterministic planning approaches, each action leads to a single state. On the other side, the DTP is a non-deterministic way of modeling the decision taken problem where each action (or exogenous event) can lead the system state to more than one possible states with a certain probability.

To deal with probabilistic non-determinism, many mathematical tools must be used. A common framework used as underlying model to DTP is the MDP [22] that exposes the probabilistic relation between the system's states. Another framework is the decision theory [23] which combines the probability theory with utility theory.

In order to model the planning problem for a stochastic dynamic system, it is necessary to present a basic problem formulation using a MDP. This work will model the problem according to [22], that presents the follow key elements:

• a set of decision epochs;
• a set of system state;
• a set of actions;
• a set of transition probabilities (state X action);
• a set of rewards or costs for transitions.

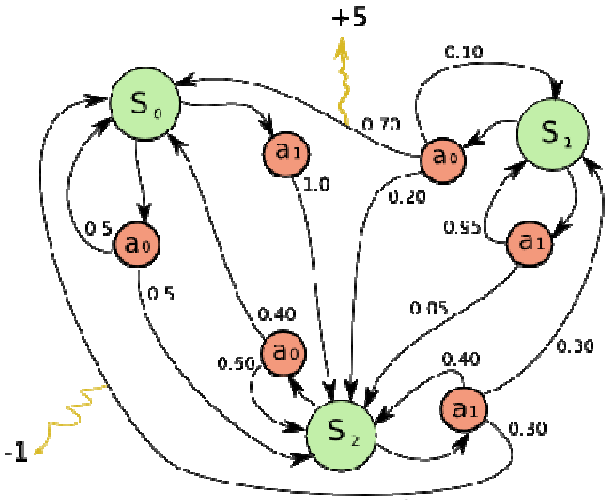

Figure 2. A graphical representation of MDP [24].

The Figure 2 shows a graphical representation of a MDP, where the green circles are the states, the red circles represents the actions, the arrows are the transitions between states, the

numbers over the arrow are the probabilities to achieve a state, and the numbers indicated by the yellow arrows are the reward value of the transition.

E. Stakeholders, Interests and Cloud Computing

This work introduces the idea of interests. Interests have been implicitly referenced in many works that address the CC, like [25]. It is relevant to explain stakeholder's concerns in a way that they lead the decisions on a cloud autonomic system.

Sharma et al. [26] presents two approaches on decisions for dynamic provisioning: cloud provider centric and customer-centric. The paper differentiates them as follows:

Cloud provider centric approaches attempt to maximize revenue while meeting an applications SLA in the face of fluctuating workloads, while a customer- centric approach attempts to minimize the cost of renting servers while meeting the applications SLA.

This definition states important aspects of decision on CC management:

• different types of stakeholders have different interests over the Cloud;

• any decision method inherently carries the ability to benefit the interests of some stakeholders and harm another;

• the stakeholder's interests are not always reconcilable given certain constraints of resources and service demand;

Therefore, it is possible to postulate that an autonomic system that intends to manage a Cloud must guide its decisions in order to maximize the total satisfaction of the stakeholders.

III. ARCHITECTURE FOR RISK ANALYSIS IN CLOUD

A. Scope and Context

Cloud computing brings several challenges for the scientific community of information security. Major challenges cited are [27] [28] [29] [30]: data privacy of users, protection against external and internal threats, identity management, virtualization management, governance and regulatory compliance, Service Level Agreement (SLA) management and trust gaps.

A strategy to meet the challenges of information security in cloud computing is the risk analysis [31]. Several papers have worked on risk analysis on cloud computing [32] [33] [34] [35] [36] [37] [38], focusing on specific techniques for identifying and assessing risks.

Current solutions for risk analysis in cloud computing does not specify the agents involved and their responsibilities during the implementation of risk analysis. This uncertainty creates deficiencies in risk analysis, as:

• Deficiency in scope: occurs when the selection of security requirements is performed by the Cloud Service Provider (CSP) or an agent without sufficient knowledge. The CSP can specify security requirements vicious in their own environment, thus defrauding the results of the risk analysis. An agent unprepared may specify wrong or insufficient requirements, thus creating an incorrect risk analysis;

• Deficiency in adhesion to Cloud Consumer (CC):

occurs when the agent responsible for defining impacts ignores the technological environment and business nature of the CC. In this case, the specification can disregard the impact scenarios relevant to the CC or overestimate scenarios that are not relevant, thus creating an incorrect risk assessment;

• Deficiency of reliable results: occurs when the quantification of the probabilities and impacts is performed by an agent who is interested in minimizing the results of the risk analysis. For example, if the analysis is performed solely by CSP, he can soften the requirements and evaluation of impacts, thus generating a satisfactory result for the CC. However, such results are incorrect.

The deficiencies outlined above can generate a lack of trust on the part of CCs in relation to risk assessments, as in current models where CSPs are performing their own risk analysis, without the participation of CCs or any other external agent.

This paper proposes a model of shared responsibilities for risk analysis in cloud computing environments. The proposed model aims to define the agents involved in the risk analysis, their responsibilities, language for specifying risks and a protocol for communication among agents.

B. The Proposed Architecture

The proposed architecture defines the sharing of responsibilities between three agents during the risk analysis. Information Security Labs (ISL) is an agent that represents a public or private entity which specializes on information security, eg an academic or private laboratory. The CC is an agent that represents the entity that is hosting their information assets in the cloud. The CSP is an agent that represents the entity being analyzed.

The three agents defined by the proposed architecture divide the responsibilities of running a risk analysis, according to the concepts defined by ISO 27005. In this context threats exploit vulnerabilities to generate impacts on information assets [31].

A risk analysis works with many variables. The variables used on proposed architecture are: (i) DE – Degree of Exposure, defines how the cloud environment is exposed to certain external or internal threat, (ii) DD – Degree of Disability, defines the extent to which the cloud environment is vulnerable to a particular security requirement, (iii) P – Probability, defines the probability of an incident occurrence, ie, a threat exploiting a vulnerability (iv) I – Impact, defines the potential loss in the event of a security incident, (v) DR – Degree of Risk, defines the degree of risk for a given scenario of a security incident.

The risk analysis of the proposed model is organized on two well-defined phases: risk specification and risk assessment.

The risk specification phase defines threats, vulnerabilities and information assets that will compose the risk analysis. At this stage it is also defined how to quantify the threats, vulnerabilities and assets specified.

The risk assessment stage comprises the quantification of the variables DE, DD and I, for threats, vulnerabilities and information assets, respectively. In this phase the quantification of variables of P and DR for each incident scenario is also performed (a combination of threat, vulnerability and asset information).

Figure 3 illustrates the flow of interactions between

components of the architecture and the ISL, CSP and CC agents in the risk specification phase. Initially each agent must register on their respective registry component (Fig. 3 a, b, c). After their registration the ISL is responsible for identifying threats and vulnerabilities in cloud computing environments. Then the ISL specifies how to quantify threats and vulnerabilities.

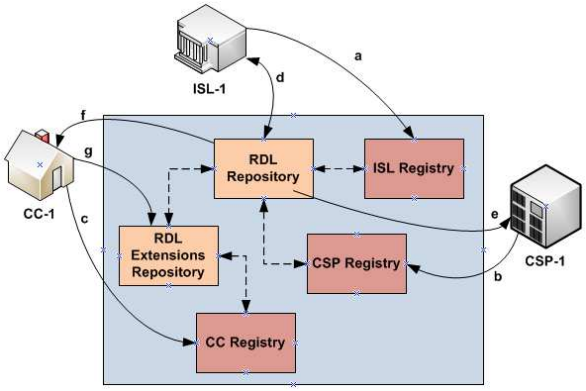

Figure 3. Risk specification phase [2].

The architecture provides a language for the specification of risk, the RDL – Risk Definition Language. This language is used by ISL to specify threats and vulnerabilities. The RDL is specified in XML and contains information such as: risk ID; ISL ID; threat and \ or vulnerability ID and reference to a WSRA – Web Service Risk Analyzer. The WSRA is a Web Service specified by ISL to perform the quantify the Degree of Disability (DD) and Degree of Exposure (DE).

After developing its RDLs and WSRAs the ISL exports the records for the RDLs repository (Fig.3-d) and publishes WSRAs.

The responsibility of the CSP on the specification phase of risk consists in importing RDLs and implementation of calls to WSRAs (Fig.3-e).

ISL is responsible for the correct identification of threats and vulnerabilities. CSP is responsible for the correct execution of the quantification of threats and vulnerability. The CC agent is responsible for the identification of information assets and the quantification of impact, as this is the most fitting agent to express the cost of an information security incident.

In order to perform the identification of an information asset and quantifying an impact on this asset a CC must import the RDLs (Fig.3-f) and extend them including information on information assets and their impacts.

The method of quantification of impacts may be static or dynamic. In the static method the CC determines a fixed value for the impact and in the dynamic method the CC specifies a Web Service to quantify the impact. After specifying their information assets and their impacts the CC exports the extension to the RDL Extensions Repository (Fig.3-g).

C. Results and Discussions

With the information from the risk analysis the CC may decide to allocate or not their information assets in a particular CSP .

The proposed model aims to reduce the three main deficiencies presented by current models of risk analysis in the cloud: deficiency in scope, deficiency in adhesion and deficiency of reliable results.

The reduction in adhesion deficiency occurs when the proposed model includes the CC as a key agent in the process of risk analysis. The CC agent has an important role in risk analysis, defining information assets and quantifying impacts on these assets.

The CC is the most suitable agent for the definition of impacts. It is the agent which best understands the relevance of each information asset within its area of expertise. CSP and ISL agents are not able to identify or quantify the impacts on information assets. They are not experts in the business of CC.

The proposed model acts to reduce the deficiency in scope by adding the ISL agent. ISL is an agent specializing in information security. It is the entity best suited to define security requirements, threats and vulnerabilities (specification of the RDLs), as well as to define how to qualify such threats and vulnerabilities (specification of the WSRAs).

The proposed model acts on the deficiency of reliable results because in our model the CSP has more restricted responsibilities than in models traditionally presented by related works.

Traditionally, the CSP is responsible for defining security requirements and the tests that are applied to evaluate the risk of their own environment. In this scenario, the risk assessment can be smoothed by CSP. The inclusion of the ISL agent removes responsibilities which are traditionally assigned to the CSP, such as the identification and quantification of threats and vulnerabilities, thus the result of the risk analysis more reliable.

The proposed model allows multiple ISLs defining RDLs and WSRAs jointly (Fig. 3). Thus, the definitions of risk can come from different sources and can be constantly updated in a dynamic and collaborative way, forming a large and independent base of risk definition for cloud.

The way WSRAs are specified is also a feature that impacts the improvement of scope. The use of Web Services to specify safety requirements allows them to be platform independent. It also allows the use of a wide variety of techniques for quantifying the threats and vulnerabilities because the only limit is set by the programming language chosen for implementation of WSRA.

Related works of risk analysis in the cloud do not consider the role of the CC agent on risk analysis. These works usually focus on the vulnerability assessment by the CSP, without considering the impact it will have on the vulnerability of the different information assets of the CC. The proposed model assigns the responsibilities of the identification and quantification of impact to the CC. Thus, the performing of risk analysis is shared among different agents, so the responsibility for quantifying the variables of risk analysis is not centered on a specific agent.

The CSP is the agent that will be analyzed; therefore it is not able to set any of the variables of the risk analysis, as this could make the results of risk analysis incorrect. The role of

CSP is only to inform the data requested by ISL, so to the own ISL performs the quantification of each information security requirement.

A CC can perform analysis on multiple CSPs before deciding to purchase a cloud service. It is also possible to perform periodic reviews of its current provider and compare them with other providers in the market, choosing to change CSP or not.

IV. Risk-based Dynamic Access Control

A. Scope and Context

Cloud computing is a model for enabling on-demand network access to a shared pool of computing resources [39]. It is widely adopted and provides advantages for customers and service providers.

As cloud computing grows in popularity, new ideas and models are developed to exploit even further its full capacity, increasing efficiency and scalability. One of these ideas is the deployment of cloud federations [40] [41]. A cloud federation is an association among different Cloud Service Providers (CSPs) with the goal of sharing data and resources [42].

However, to make such a scenario feasible it is necessary to develop authentication and authorization models for largely distributed, dynamic and heterogeneous environments.

This problem is usually treated by the deployment of identity federations. An identity federation is a model of identity management where identity providers and service providers share users' identities inside a circle of trust [43].

This solution, nevertheless, is not optimal, since identity federations present problems such as the necessity of attribute and trust agreements, interoperability issues and, in practice, show limited scalability [44]. This paper shows that it is possible to provide authorization in cloud federations without the need for an identity federation.

The difference between cloud federations and identity federations is that cloud federations are built to share resources and identity federations are built to share users and identity information.

In this paper, we propose to use a risk-based dynamic access control to enable authorization in a cloud federation without the necessity, but allowing the possibility, of using identity federations.

B. Proposals and Solutions

In this paper we propose that it is possible to provide a way to establish cloud federations without the need for identity federations, by using risk-based access control. This can increase the scalability of this model and handle exceptional requests.

A.Cloud Federations

Figure 4 presents an overview of the cloud federation architecture that we are considering. This architecture is based on the common points found in the main federation projects currently being developed, some of which were described in Section II.

The main application scenarios for such federations are medical, military and scientific collaborations, which require large storage and processing capabilities, as well as efficient information sharing.

In this architecture we have the following components:

CloudProvider: this is the Cloud Service Provider (CSP) itself, who provides the infrastructure over which the virtual resources are allocated (they are represented by the clouds in the figure);

CloudManager: responsible for attaching a CloudProvider to the federation. It is composed of several services that deal with users, resources, policies, service-level agreements, security and the CloudProvider. It is modular so that it can be attached to different cloud management software just by changing one of its services.

FederationManager: responsible for coordinating the federation. It acts as a naming service and is also responsible for message passing.

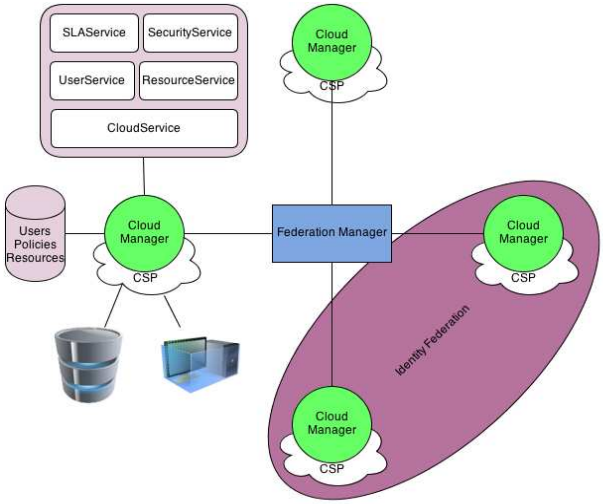

Figure 4. Overview of the federation [3].

As shown in Figure 4, some of the participating clouds may form identity federations among themselves.

Under the point of view of a user there are two types of clouds in this architecture: a home cloud (his original CSP) and foreign clouds (the other clouds in the federation). Users can deploy and access resources in both types of cloud, but access control behaves differently for each case.

When users deploy a resource in their home cloud they may choose if it will be available for users of foreign clouds. In any case the user must upload an XACML policy file together with the resource, which will be used for ABAC.

Users may also deploy resources in a foreign cloud and it will automatically be available to every user of the federation. Finally, users may access resources in their home cloud or shared resources in foreign clouds.

When a user tries to access a resource in their home cloud, this request is handled by a classical ABAC model. Based on user attributes and XACML policies the system grants or denies the requested access.

When a user tries to access a resource in a foreign cloud, the system first verifies if both clouds are in an identity federation, in which case the access will also be handled by ABAC, but if there is not an identity federation between them, the "break-the-glass" mechanism is activated and the risk-based access control Policy Decision Point (PDP) is called.

The PDP is located in the cloud handling the access request (foreign to the requester) and the metrics and parameters of risk estimation are defined by the administrators of this cloud and the users who own the resources.

These metrics are informed in an eXtensible Markup Language (XML) file, containing definitions of risk metrics and how to measure and aggregate them, as well as a threshold level for granting access to the resource and possible obligations that users will have to follow. This file is known as a risk policy.

Each cloud provider must provide a set of basic metrics with their quantification rules. Those will be used to create a baseline risk policy for the provider. This guarantees that a cloud provider is able to maintain their minimal security requirements.

Each resource has its own risk policy, which must respect what is defined in the baseline policy, but may be extended to become more or less restrictive as the user desires. The XML file of the policy must be uploaded by users when they choose to deploy a shared resource.

If a user chooses to define a risk metric that is not available in the server, he/she must provide a way for the CSP to quantify this risk. This is done by defining a Web Service that will be called by the PDP upon the evaluation of the access request. The PDP will forward the access request to the Web Service, which will have to parse it, process it and return a numeric value representing the associated risk for the metric being evaluated.

To handle the access request for a given resource all of the metrics are valued, based on the rules defined by the CSP and the Web Services defined by the user. The chosen aggregation engine is used to reach a final risk value. This value is then compared to the defined threshold and, if lower, the subject is given special access.

Before granting access, however, the policy is analyzed in search of obligations that were defined by the user. Those obligations are stored in a system monitor, which will watch and log every user action once the access is granted.

C. *Results*

To validate our proposal and measure some performance characteristics we implemented the key parts of the federation system and the whole access control system.

The implementation used the Python programming language, the zeromq library to handle message passing, MySQL for persistence, the ndg-xacml library for XACML evaluations and the web.py framework for the web services.

The infrastructure over which the federation was deployed was composed of two OpenNebula clouds running on a laptop with a 2,53GHz Core i5 processor and 4GB of RAM. All of the experiments were repeated 50 times to obtain the averages and all of the times shown here refer only to the execution of the access control decision function, ignoring message passing between the clouds. Table I shows four different cases of access request. Case A represents 10 requests handled by local XACML only; case B represents a risk decision that involves 10 risk quantification rules performed locally; case C uses 5 local rules and 5 external (web service) rules; and case D represents a risk policy with 10 external risk quantification rules.

It is possible to see that the use of local risk quantification rules has no significant impact on performance, while the use of web services does affect performance, as expected, because of the HTTP invocations that must be performed for each metric.

Figure 5 shows the growth in time spent reaching an access decision as we increase the number of metrics which call web services in a risk policy file.

TABLE I. COMPARISON OF DIFFERENT ACCESS REQUEST CASES [3]

	min. (ms)	max. (ms)	average (ms)
A	1.057	9.372	1.46
B	1.824	15.564	4.574
C	1556.182	2813.56	1726.71
D	3247.563	10350.5	4220.6

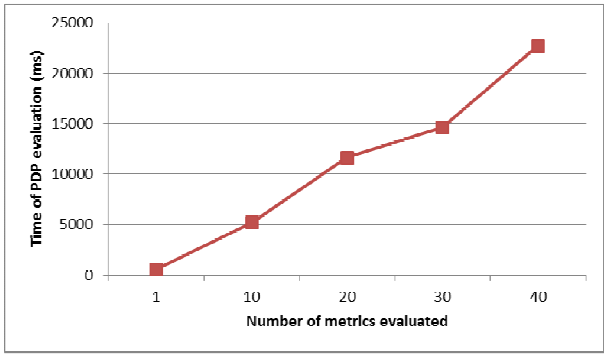

Figure 5. Performance comparison with a varying different number of external metrics [3].

V. CHALLENGES OF OPERATIONALIZING PACS ON CLOUD

A. *Scope and Context*

Medicine is being perfected through the use of innovative technological solutions in various equipment with diverse applications, such as image processing, blood analysis, surgical assistance and distance patient control.

Inside hospitals and clinics for diagnostic imaging, it is usual to find PACS. PACS have the goal of managing the storage and exhibition of medical images. Through workstations, doctors have access to the PACS system where they manipulate images independently of their physical location.

Specialized physicians achieve diagnosis through the analysis of images or the reading of reports. But these physicians are not always present where the exam was

performed, especially in cases where the participation of a second doctor is necessary or in the case of training for resident physicians. The involvement of these doctors can happen with telemedicine.

Telemedicine comprehends the offering of services related to health care in cases where distance is a critical factor; such services are performed by health professionals using communication and information technologies for the interchange of information valid for diagnostics, prevention and treatment of illnesses and the continual education of health service providers, as well as research [45].

The practice of telemedicine is only made possible because of significant advances in communication systems. The possibility of connectivity to the World Wide Web from mobile devices, a constantly evolving technology, allows patients to obtain adequate medical care in less-favored regions, where there are no doctors or wired Internet connection available. Wi-Fi Networks, 3G, 4G, WiMax and other wireless networks are being constantly improved with higher data transmission rates, allowing access to content not explored before, which aim to improve, simplify and complement the services related to patient care and make them more efficient.

Cloud computing is currently the main theme of a lot of research in information technology. The possibility of sharing resources through clusters, virtualization and the ease of access to information attracts more and more information technology researchers. This technology is also a powerful tool to promote the homogenization or virtualization of space [46].

Images from radiological exams are used in clinics and hospitals for medical diagnosis. The inter-relation among clinics, hospitals and radiology departments are increasingly dependent of the accessibility of these images, from anywhere inside or outside of the health care unit [47].

The idea is to use cloud as a model for applications being delivered as services over the Internet. Cloud services are built in such a way that if a machine fails, the system resets, in order to prevent the service to crash or that the contractor knows that there was some kind of problem. Cloud computing enables the growth of processing and storage infrastructure for hospitals and clinics without causing much impact. Thus cloud based PACS enables medical activity from anywhere using computers or devices connected to the Internet.

B. Interoperability Challenges

Health charts, medical and laboratory reports, medical images and prescribed medicine are some of the items in a medical record, and those records are becoming more and more complex. Physicians in hospitals and clinics need a flexible resource that allows them accessing information and history for each patient, because they work and meet patients in several places; they need to be frequently following exams and giving support to several people.

Medical records occupy a great storage space and the management of these data is a challenging task for hospitals and clinics [48]. To solve those issues, these organizations invest large amounts of money in infrastructure for communication, processing and storage of exams. Inside this infrastructure are equipment for ultrasonography, MRI, CT scans and radiography. Figure 6 shows the sending of medical images to the PACS.

The modalities, as known equipment's, send images using Digital Image and Communication in Medicine (DICOM), a digital standard to store and transmit medical images, over a Wireless Local Area Network (WLAN) to a PACS located on a cloud. Communication from the medical equipment to the WLAN is achieved by using a device that connects to a Local Area Network (LAN), which in turn accesses the PACS that is on the cloud using the Internet.

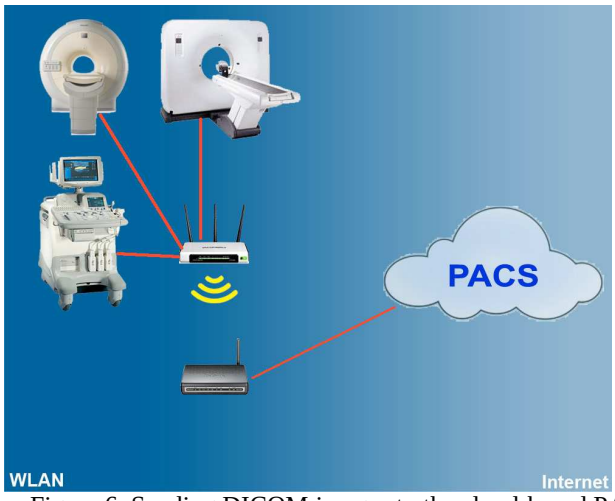

Figure 6. Sending DICOM images to the cloud-based PACS [4].

C. Proposed Scenario

This work shows a cloud-based PACS solution, allowing exams to be performed in various equipment and transmitted over wireless networks, so that from any connected mobile device a doctor may have access to PACS exam images. Figure 7 shows diverse equipment such as cellphones, tablets and laptops accessing the cloud-based PACS over wireless connections, such as Wi-Fi, 3G, 4G or WiMax.

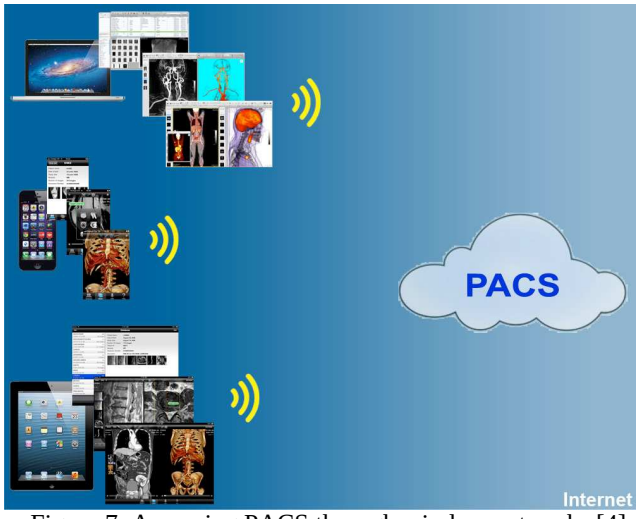

Figure 7. Accessing PACS through wireless networks [4].

The PACS system in the implementation is DCM4CHEE [49] [50], with a PostgreSQL database. The devices for test access were a laptop, a smartphone and a tablet, all of them with the Osirix application. All of them are using publicly available 3G and Wi-Fi connections. In every case the connection is successful, showing a mean time of 20 seconds to download a DICOM exam of a lumbar spine containing 100 images.

D. Identity Management Implementation

The PACS system was installed and configured on a Windows 2008 Server R2 virtual machine. On this same machine the Shibboleth, Apache 2.2, Tomcat 7, OPENLDAP, JASIG CAS, JDK 1.7 and PostgreSQL 9.2 services were configured.

Apache was configured to allow the use of Secure Sockets Layer (SSL) connections, and also to proxy its requests to Tomcat. Tomcat was configured to run the authentication and identity management applications.

The authentication server enables JASIG CAS SSO authentication via a web interface passing authenticated users to Shibboleth. CAS was set up to search for users in a Lightweight Directory Access Protocol (LDAP) directory.

With the server configured, Shibboleth was installed. The IDP application was installed and configured in Tomcat. To act as the service provider, the Shibboleth application must belong to a federation. TestShib was chosen as the federation, because it was created to test Shibboleth configurations (SPs and IDPs). An IDP was registered informing the hostname and digital certificate created earlier. Shibboleth was then configured to use TestShib's metadata. The CAS Client receives shibboleth authentication. With this process ready Shibboleth is configured and authenticating users from JASIG CAS.

VI. Environment, Services and Network Management for Green Clouds

A. Scope and Context

The load prediction models in traditional architectures and cloud computing environments are based on the analysis of historical data and demand increments from business models. This information makes it possible to pre-allocate resources. However, load prediction models are challenged (and frequently broken) when unexpected peaks of demand occur.

Approaches to dealing with the problems of load prediction models include the following: allow for a margin of on-line resources, i.e., over-provision resources; to turn on idle resources; and to temporarily use external resources on-demand (i.e., federated clouds), and others. Each of these approaches has its advantages and disadvantages. The challenge in green computing, as described by [51], is to exploit the balance between these approaches in order to address the pressing issue of data center over-provisioning related to the need to match the peak demand.

We propose a solution based on integrated environment, services and network management that promotes: equitable load distribution through techniques like virtual machines; predictive resource allocation models through historical load analysis and pro-active allocation methods; aggregate energy management of network devices; and integrate control over the environmental support units, which represent the larger share of energy consumption.

The objectives are the following: to provide flexibility of the system configuration that allows for the easy introduction of new elements in the managed environment and the configuration processing distribution among services; to provide a level of availability that keeps to higher standard SLA (Service Level Agreement) compliance rates and which contributes to system's stability and security; to reduce cost in both capital and operational costs (CAPEX and OPEX), [52], to support the business predicates, and thus promote the acceptability of the proposed method; and to provide sustainability by using methods to reduce energy utilization and carbon emission footprints.

To achieve our objectives we propose an OTM (Organization Theory Model) for integrated management of a green cloud computing environment. It works based on organization models that regulate the behavior of autonomous components (agents) that view the environmental elements, network devices (e.g. switches, cards and ports) and service providers (e.g. processing servers, load distribution services, task processors and temperature reduction services). For example, the management system is able to turn off unused network devices and servers, turning off the environmental (cooling) support units. This is reactive to characteristics of the predicted system load. The controlling elements are able to coordinate between themselves aiming at a higher-level system's objective, e.g. to keep overall energy utilization and SLA compliance metrics.

Our research advances the state of the art as follows: it introduces an organization theory model for integrated

management of the green clouds based on the concepts of organization models, network management, and distributed computing; it analyses the network and system's behavior and operational principles; it validates the proposal demonstrating the system's added-value in a case study scenario; and it improves a simulator (the CloudSim framework) to validate the green cloud computing management approach.

Our research was motivated by a practical scenario at our university's data center. In the (not so distant) past, we applied the "traditional architecture" which was composed of diverse processing clusters configured to process different services. We faced the usual issues encountered in large data centers at that time: lack of rack space, which impacted flexibility and scalability; an excessive number of (usually outdated) servers, which impacted operation costs; the need of an expensive refrigeration system; and an ineffective UPS (Uninterruptible Power Supply) system, which was problematic to scale due to the number of servers involved.

With the use of cloud computing, we managed to consolidate the number of servers using virtualization techniques. Using this technology, we concentrated the predicted load on a few machines and kept the other servers on standby to take care of peak loads. The immediate results were very positive: reduction of rack space utilization; lower heat emission due to the reduction in server utilization, with consequent optimization of the cooling infrastructure, and, a quick fix for the problematic UPS system because we had less active servers.

As part of an institutional initiative towards sustainability and eco-friendliness, our next step was to optimize energy utilization [53] and reduce carbon emission. For this, we looked at solutions from the fields of green computing and, more specifically, green cloud computing. We noticed that there was room for improvement as we consolidated resources using cloud computing. For instance, there were periods in time when the VMs (Virtual Machines) were idle and the servers were underutilized. Based on the principles established by [54], our goal was to promote energy-efficient management and search for methods to safely turn off unused servers using an on-demand basis. The intuitive approach was to concentrate the running applications (configured per VMs) in a few servers and recycle server capacity.

Although appealing, this approach led to a major issue: service unavailability! A quick analysis concluded that it was related to the time required to bring up the servers during unpredictable peak loads. We concluded the following: the dimensioning is based on historic intra-day analysis of services demand. More specifically, it is based on the analysis of previous day's demand plus a margin of the business growth that can be estimated as the amount of resources required for one service in a period of time; however, when dealing with services with highly variable workloads, that prediction becomes complex and often immature. Moreover, external factors can lead to unexpected peaks of demand. For that, we left a safety margin of resources available (e.g. 20% extra resources on standby). Besides the excessive energy utilization, this approach fails when the demand surpassed that threshold; as a solution, we needed to bring up turned-off resources. The lapse of time between the detection of the situation and the moment that processing resources become available caused the service unavailability.

We analyzed several alternatives to overcome this issue that implements an OTM (Organization Theory Model) for integrated management of the green clouds focusing on: optimizing resource allocation through predictive models; coordinating control over the multiple elements, reducing the infrastructure utilization; promoting the balance between local and remote resources; and aggregating energy management of network devices.

Cloud computing is based on server virtualization functionalities, where there is a layer that abstracts the physical resources of the servers and presents them as a set of resources to be shared by VMs. These, in turn, process the hosted services and (may) share the common resources. The green cloud is not very different from cloud computing, but it infers a concern over the structure and the social responsibility of energy consumption [55], hence aiming to ensure the infrastructure sustainability [56] without breaking contracts.

B. Proposals and Solutions

To understand the problem scenario, we introduce the elements, interactions, and operation principles in green clouds. Green clouds emerged as a solution to save power by utilizing server consolidation and virtualization technologies. Fine tuning resource utilization can reduce power consumption, since active resources (servers, network elements, and A/C units) that are idle lead to energy waste. The target in green clouds is: how to keep resources turned off as long as possible?

The interactions and operation principles of the scenario are described below:

• There are multiple applications generating different load requirements over the day.

• A load balance system distributes the load to active servers in the processing pool.

• The resources are grouped in clusters that include servers and local environmental control units (A/C, UPS, etc.). Each server can run multiple VMs that process the requests for one specific application. Resources can be fully active (servers and VM on), partially active (servers on and VMs off), or inactive (servers and resource off). The number of servers and their status configuration is defined based on historical analysis of the load demand.

• The management system can turn on/off machines overtime, but the question is when to activate resources on-demand? In other words, taking too much delay to activate resources in response to a surge of demand (too reactive) may result in the shortage of processing power for a while. This reflects directly on the quality of service, as it could deteriorate the service availability level (even if this is a short time). On the other hand, activating more unnecessary resources causes resources to be left idle and wastes energy consumption.

Green cloud with integrated management is a structure

that we see as a tendency of this area and seek like a goal. These aspects that are described below are the reference for what our model aims to fulfill. In comparison to green cloud, we infer the responsibility of consuming less energy in addition to ensuring the agreements predefined in the SLA.

• Flexibility: is state-aware of all equipment under its control, acting for when it will be necessary, not when it is needed, and plan their actions based in the information of the whole cloud. It is able to predict and execute necessary changes in hardware according to the demand of the cloud; such as slowing down an overheated CPU, turning on machines based on foreseen load coming, or triggering a remote backup in case of fire. It is able to interact automatically with public clouds [56], migrating or rising up new nodes on demand in remote clouds. It provides a better support for occasional workload peaks or DoS (Denial of Service) attacks.

• Availability: encompasses a new level by extending itself to the public clouds, allowing the creation of mirror clouds. It deals with context grouping automatically, being able to migrate these groups, or elements to public clouds.

• Cost reduction: by having an automated management based on previous experience and results, it can manage itself with minimal human intervention. It uses a 24/7 management system aiming to provide a better utilization of the resources. It will enlarge the equipments lifetime, decrease the downtime caused by human errors and reduce the expenses by adopting smart strategies for resource utilization. With inter- cloud communications it can adopt a minimalist configuration, ensuring local processing for most of their workload, and leaving the workload peaks to an external cloud.

• Sustainability: its structure has the ability to adopt goals for SLA, goals for energy consumption (average X kWh per day) or goals for heat emission (average Y BTU per day). The structure reacts with the environment events in order to fulfill the predefined goals. Events like UPS state down, temperature sensors accusing high degrees or fire alarms on. In parallel, adapts the environment dynamically in order to fulfill the internal goals; like decreasing the cooling system to reach consumption goals.

We propose that breaking the centralized management service in several little management services gives us the necessary elements to increase the “degree of freedom” of the cloud, creating the possibility to achieve a balanced situation between risk and consumption.

However, with several management services in the cloud we introduce a new problem: the management of these services becomes a complex job. For this, we use the principles of organization theory, to organize and classify such services, making them easier to control. Cloud management through the organization theory principles gives the possibility to auto configure the management system, since the addition of a new element (such as network device, VM, PM, UPS) is just a matter of including a new service in the management group.

Hence, we propose a proactive model for cloud management based on the distribution of responsibilities for roles. In this approach, the responsibility for managing the elements of the cloud is distributed among several agents, each one in one area. These agents will individually monitor the elements of the cloud of their responsibility. They act in an orchestrated way aiming for the fulfillment of the standards (norms).

Such orchestration is based on the fact that the knowledge about the state of the cloud (as a whole) be shared by all agents, the existence of planning rules, to guide the actions of the agents, and the development of beliefs about the inner workflow of the cloud, that are constantly revised.

Since the data center structure is scaled and used to provide services, this remains only a tool to provide such services. Generally, service level agreements are established in order to clarify the responsibilities of each part - client and provider. We emphasize that these agreements should be kept at their level (i.e. service), making them purely behavioral rules (e.g. delay, fault tolerance) for the service, excluding structural and physical requirements. Without the details of the environment configuration in the agreement, the cloud becomes flexible. With the independence and flexibility to change the configuration of the structure, it can become dynamic and extensible.

It can allow for covering external agreement factors still critical to the data center infrastructure (i.e., energy consumption, hardware wear, among others), but not related to the agreement. Just as we live under the laws of physics, the cloud should also exist in well-defined laws, which we call norms. These norms express the rules of the service behavior established in the SLA and the internal interests of the cloud, which need to be considered.

For the various elements of the cloud to work efficiently, seeking the enforcement of these standards, they should be coordinated by external agents to the services they audited; managing, for example: enabling and disabling VMs; enabling and disabling PMs; configuration changes in VMs; and enabling and disabling network devices.

Since there is a wide range of elements to manage, the complexity would grow proportionally with the size of the cloud. To avoid such complexity we infer a hierarchy to the existing agents. We can make an analogy to a large company where there is a hierarchy to be followed and responsibilities being delegated. Just as in a company, there must be a system manager (the boss) that controls the entire environment. Following the hierarchy we have the coordinators who split the operations between their teams [57] in order to facilitate the division of tasks and responsibilities among its teams.

Depending on the situation, decisions will generate system operations or service operations, or both. System operations can be divided into VM management, servers management, network management and environment management. The service operations can be divided into monitor element, service scheduler and service analyzer.

The action of each role is directly reflected in the configuration of the structure as a whole. The system operations will act over the structure and environment in which the services are being processed. The services

operations will act over the service layer and the environment, acquiring information from both.

The four roles that operations system may be classified as are:

• VM management: responsible for the actions implied the virtual machines. It has an interface between the model and the virtual machines. As an example, creating or destroying a VM, changing your settings and even moving it from one host to other host (either from local or remote data center).

• Servers management: responsible for the actions implied the physical machines. It has an interface between the physical machines and the model. As an example, turning off and on a physical machine, changing the settings of the host operating system (e.g. such as BIOS - Basic Input/Output System, SMART - Self- Monitoring, Analysis, and Reporting Technology), hardware configurations (e.g. cooler and accelerometer), and backend equipment (e.g. such as storage devices, switches and site backups).

• Network management: Responsible for actions implied the network devices. It uses SNMP tools gathering traffic data and computing the utilization of each port on all the switches, minimizing the active network components, while turning off unused switches, and disabling unused ports saving energy.

• Environment management: responsible for actions outside the structure. It has an interface between the environment and the model. As an example, temperature control of the data center, control of power backup systems (e.g. UPS and generator), control over the accessibility of the data center (e.g. physical security).

The three roles that service system may be classified as are:

• Monitor element: responsible for the collection of information structure in general, and your understanding. It has the responsibility to keep the model aware of the state of the cloud by monitoring the servers, VMs, network traffic and so on. It is based on specific parameters previously configured by the System Manager, such as the use of a resource and its threshold notification, the availability of network links (binary data) or idleness of some element of the structure.

• Service scheduler: responsible for the cloud agenda. It has a proactive role in the model, planning the actions to be taken before the scheduled events. In an exchange of physical machines, for example, it will generate the following list of steps to be followed: test secondary UPS; enabling secondary server; and VM's migration.

• Service analyzer: responsible for testing services and behavioral analysis. It has the role of auditing the service provided by the framework and understanding it. It makes sure that the service provided is in accordance with the norms to be followed, by inferring pre-established thresholds and alerting the system manager. It monitors the quality of service that is provided, and tries to relate it with the variations in the structure, finding patterns between the performance obtained and the variants elements.

Planning Rules and Beliefs:

• Planning rules: the basis of theoretical knowledge, which relates contexts and objectives. They are used at times when decisions must be made, during the planning of actions. They are pieces of primitive knowledge gleaned from the experience of managers. We can take as an example of Planning Rules the following notions: if a VM increases the use of page swap ping, to decrease it, we will increase memory RAM (Random Access Memory); if the physical machine presents a high load, to decrease the load, we will move the VM with more processing to another physical machine; if the data center presents a high load, to decrease the general load, we will turn on more physical machines.

• Beliefs: empirical knowledge used to improve the decisions to be taken. In this we have the empirical understanding above the functioning of the cloud. The beliefs express the junction of practical knowledge (the premises), coming from the norms and empirical knowledge, originating from the historical data and past experiences. The beliefs must be reviewed frequently by all elements of the model, as well as the sharing of these reviews. We can take as an example of beliefs the following notions: the activation of a server type X represents an increase of Y degrees in Z minutes; the activation of a VM type A increases the consumption in B kWh; the VM type A supports C requests per second.

VII. Provisioning and Resource Allocation for Green Clouds

A. Scope and Context

We are also proposing two strategies for allocation and provisioning of PMs (Physical Machines) and VMs (Virtual Machines) using DVFS (Dynamic Voltage and Frequency Scaling) as an improvement of private clouds sustainability, transforming the Cloud into Green Cloud [5]. Green Clouds crave for efficiency of its components, so, we adopted positive characteristics of multiple existing strategies, developing hybrid strategies that, in our scope, aim to address:

- A sustainable solution to mitigate peaks in unpredictable workload environments with rapid changes;

- An optimization of the data center infrastructure without compromising the availability of services during the workload peaks;

- Balance between the sustainability of the infrastructure and the services availability defined on SLAs (Service Level Agreements).

This work was based on actual data collected by the university data center, that has multiple services suffering often with unexpected workload peaks, whether from attacks on servers or overuse of services in short periods of time. First, we propose an allocation model for private Clouds that aims to reduce the costs (energy and SLA fines) while improving the resource optimization. Second, we propose a provisioning model for private Clouds, turning them into Green Clouds, allowing the reduction of energy consumption and resource optimization while maintaining the SLAs with the integration of public Cloud resources. Third, after we validate our hybrid provisioning strategy, we have the opportunity to apply the hybrid provisioning strategy in a

Cloud environment that uses DVFS (Dynamic Voltage and Frequency Scaling) in its physical machines. This way we achieve an improvement in energy consumption and resource optimization with no impacts on the Cloud SLAs.

The motivation for this work can be summarized in the following points:

- Energy saving: [58] says ”Energy saving is just one of the motivational topics within green IT environments.” We highlight the following points: the reduction of monthly data center OPEX (Operating Expenses); the reduction of carbon emissions into the atmosphere (depending on the country); and the extension of the lifespan of UPS (Uninterruptible Power Supply) [54].

- Availability of Services: Given the wave of products, components, and computing elements being delivered as services by the Cloud (*aaS), a series of pre-defined agreements or governing the behavior of the service that will be supplied / provided is needed [59]. According to Cloud Administrators, agreements that provide availability rates, usually 99.9% of the time (or more) are a concerning factor. Thus, the question is how to provide this availability rate while consuming little power.

- Variation Workload: In environments with multiple services, the workload prediction is complex work. Historical data is mostly used to predict future needs and behaviors. However, abrupt changes are unpredictable causing temporary unavailability of provided services. The need to find new ways to deal with these sudden changes in the workload is evident.

- Delayed Activation: Activation and deactivation of resources are a common technique for reducing power consumption, but the time required to complete this process can cause some unavailability of provided services, generating contractual fines.

- Public Clouds: Given the growing amount of public Clouds and the development of communication methods among Clouds, like Open Cloud Consortium [60], and Open Cloud Computing Interface [61], it became possible, for small or big companies, to easily use multiple public Clouds as extensions of a single private Cloud. We considered this as an alternative resource to implement new Green Cloud strategies. This is beneficial to those who need to expand their Cloud, and to the new clients of Cloud providers.

In a broad sense, this proposed model is for the Cloud provider that seeks the balance between energy saving and service providing (defined by the SLA).

We aim to propose an allocation strategy for private Clouds and a provisioning strategy for Green Clouds, which suits the oscillatory workload and unexpected peaks. We will focus on finding a solution that consumes low power and generates acceptable request losses, in comparison to other base strategies.

B. Proposals and Solutions

The concept of combining organization theory and complex distributed computing environments is not new. [62] already proposed the idea of virtual organizations (VOs) as a set of individuals and / or institutions defined by such sharing rules in grid computing environments. This work concludes that VOs have the potential to radically change the way we use computers to solve problems the same way as the Web has changed the way we consume and create information.

Following this analogy, we have a similar view: Management Systems based on the Organization Theory would provide means to describe why / how elements of the Cloud should behave to achieve global system objectives, which are (among others): optimum performance, reduced operating costs, appointment of dependence, service level agreements, and energy efficiency.

These organizational structures, proposed in [5], allow network managers to understand the interactions between the Cloud elements, how their behavior is influenced in the organization, the impact of actions on macro and micro structures, as the macro level processes allowing and restricting activities at the micro level. This way, it provides computational models to classify, predict, and understand the elements interactions and their influence on the whole environment.

Managing Cloud through the principles of the Organization Theory provides the possibility for an automatic configuration management system, since adding a new element (e.g., Virtual Machines, Physical Machines, Uninterrupted Power Supply, Air Conditioning) is just a matter of adding a new service on the Management Group.

The proposed strategies are based on a pro-active management of Clouds, which is based on the distribution of responsibilities in roles. The management responsibility of the Cloud elements is distributed among several agents; each agent controls individually a Cloud element that suits him.

[5] proposed a model based on the Organization Theory to manage a Cloud environment using decentralized management services. They proposed agents to manage the Cloud elements, each agent managing the elements that are in its area. These agents would individually monitor and manage the elements they are responsible for, orchestrating them to fulfill the norms that are imposed to the system.

Norms are the rules or agreements used as input into the system such as SLAs, energy consumption, resource optimization, air conditioning (data center temperature), etc. They are a primitive knowledge collected from experienced administrators and are used at times when decisions need to be made. In complement to Norms, [5] defined believes that are empirical knowledge used to improve the decisions at management. It is the junction of the practical knowledge from the norms and empirical knowledge from historical data, derived by the system, analyzing historical data traces and correlating them with the norms that have or have not been fulfilled.

[5] also defined roles that the agents would assume while monitoring/managing the Cloud environments or services. The roles defined for agents that act at Cloud environment level are: VM management, server management, network management and environment management. The roles defined for agents that act at service level are: monitor element, service scheduler and service analyzer.

Based on [5], we conclude that the Organization Theory model would be applicable for managing the entities of a Cloud computing environment in a decentralized way. So far, our models apply the Organization Theory ideas as describe by [5], using decentralized agents to monitor and manage the Cloud entities.

The DVFS (Dynamic Voltage and Frequency Scaling) was presented by [63]. It provides an alternative solution to decrease power consumption by giving the possibility to the PMs to independently decrease their power dissipation, by lowering the processor clock speed and supply voltage during the idle periods of time of the processor.

DVFS pros:

• Adaptive Consumption: lower energy consumption by adapting the processor frequency to the workload.

• Out-of-the-box: There is no need to adapt applications or services to use it.

• Management: The user (or application) is allowed to determine when to use (or not) the solution, giving the possibility to control the CPU temperature.

DVFS cons:

• Low Performance: decreasing the CPU frequency will reduce the system performance, which is expected [64].

• Inertia of Changes: The frequency takes some time to adapt to the system's needs. So, in scenarios with high load variations, DVFS could become a problem.

• Over Changes: The rapid and constant act of "overvolting" and "undervolting" the processor, trying to fulfill immediately the system needs, could decrease the equipment lifetime [65].

DVFS enhancements, as seen on the right side of Figure 1.8, also shows a deeper level of DVFS. The idea is to apply it at the core level, not at the processor level as a global unit. Another work is trying to decrease the gap between voltage and frequency changes. The idea is to optimize the processor and build a fast DVFS that adapts quickly to system needs.

For the conscious resource provisioning in Green Cloud environments, we propose a hybrid strategy that uses public Cloud as an external resource used to mitigate SLA breaches due to unexpected workload peaks. In parallel, for the optimal use of local resources, we propose a strategy of dynamic reconfiguration of the VMs attributes, allocated in the data center. Given the distributed model presented in the previous section, we used the Cloud simulation tool CloudSim simulate the university data center environment and workload.

VIII. Optimizing Green Clouds through Legacy Network Infrastructure Management

A. Scope and Context

Traditionally, computer systems have been developed focusing on performance and cost, without much concern for their energy efficiency. However, with the advent of mobile devices, this feature has become a priority because of the need to increase the autonomy of the batteries.

Recently, the large concentration of equipment in data centers brought to light the costs of inefficient energy management in IT infrastructure, both in economic and environmental terms, which led to the adaptation and application of technologies and concepts developed for mobile computing in all IT equipment.

The term Green IT was coined to refer to this concern about the sustainability of IT and includes efforts to reduce its environmental impact during manufacturing, use and final disposal.

Cloud computing appears as an alternative to improve the efficiency of business processes, since from the point of view of the user, it decreases energy costs through the resources sharing and efficient and flexible sizing of the systems. Nevertheless, from the standpoint of the service provider, the actual cloud approach needs to be seen from the perspective of Green IT, in order to reduce energy consumption of the data center without affecting the system's performance. This approach is known as Green Cloud Computing.

Considering only IT equipment, the main cause of inefficiency in the data center is the low average utilization rate of the resources, usually less than 50%, mainly caused by the variability of the workload, which obliges to build the infrastructure to handle work peaks that rarely happen, but that would decrease the quality of service if the application was running on a server fully occupied [66].

The strategy used to deal with this situation is the workload consolidation that consists of allocating the entire workload in the minimum possible amount of physical resources to keep them with the highest possible occupancy, and put the unused physical resources in a state of low energy consumption. The challenge is how to handle unanticipated load peaks and the cost of activation of inactive resources. Virtualization, widely used in the Cloud approach, and the ability to migrate virtual machines have helped to implement this strategy with greater efficiency.

Strategies to improve efficiency in data centers have been based mainly on the servers, cooling systems and power supply systems, while the interconnection network, which represents an important proportion of consumption, has not received much attention, and the proposed algorithms for load consolidation of servers, usually disregard the consolidation of network traffic.

The concepts of Green IT, albeit late, have also achieved design and configuration of network equipment, leading to Green Networking, which has to deal with a central problem: the energy consumption of traditional network equipment is virtually independent of the traffic workload. The Green Networking has as main strategies proportional computing that applies to adjust both the equipment processing speed such as the links speed to the workload, and the traffic consolidation, which is implemented considering traffic patterns and turning off components not needed. According to [67], traditionally the networking system design has followed two principles diametrically opposed to the aims of Green Networking, over-sizing to support demand peaks and redundancy for the single purpose of assuming the task when other equipment fail. This fact makes Green Networking technically challenging, with the primary objective of introducing the concept of energy-

aware design in networks without compromising performance or reliability.

While the techniques of Green Networking begin to be standardized and implemented in the new network equipment, a large amount of legacy equipment forms the infrastructure of current data centers. In the works to be presented, it is shown that it is possible to manage properly these devices to make the network consumption roughly proportional to the workload.

Thereby, there is the need and the possibility to add, to the Green Cloud management systems, means of interaction with the data center network management system, to synchronize the workload consolidation and servers shutdown, with the needs of the network traffic consolidation.

Taking into account that the more efficient becomes the management of virtual machines and physical servers, the greater becomes the network participation in the total consumption of the data center, the need to include network equipment in green cloud model is reinforced.

The principles suggested in recent papers by several authors for power management in legacy network equipment are presented, and their application to optimize our approach of green cloud is proposed.

B. Proposals and Solutions

The proposal considers the network topology of a typical data center shown in Figure 8, where the switches are arranged in a hierarchy of three layers: core layer, aggregation layer and access or edge layer. In this configuration, there is redundancy in the connections between layers so that the failure of a device does not affect the connectivity.

Consequently, we consider, in our model, that each rack accommodates forty servers and two access layer switches. Each of these switches has 48 Gigabit Ethernet ports and two 10 Gigabit Ethernet uplink ports, and each server has two Gigabit Ethernet NICs (Network Interface Controllers) each one connected to a different access switch.

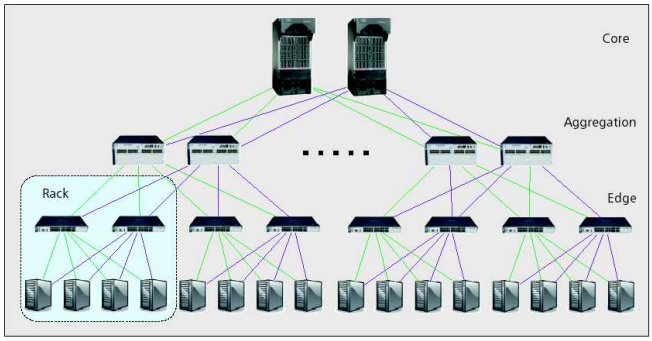

Figure 8. Typical network topology of a data center [68].

We also consider that if there is only one rack, aggregation layer switches are not required, and up to 12 racks can be attended by 2 aggregation layer switches with twenty four 10 Gigabit Ethernet and two 10 Gigabit Ethernet or 40 Gigabit Ethernet uplinks, with no need for core switches.

Finally, the model assumes that, with more than 12 racks two core switches with a 24 ports module for every 144 racks will be required. The module's port speed may be 10 Gigabit Ethernet or 40 Gigabit Ethernet, according to the aggregation switches uplinks.

In traditional facilities, the implementation and management of this redundancy is done by the Spanning Tree Protocol and in most recent configurations by the MultiChassis Links Aggregation Protocol (MC-LAG), which allows using redundant links simultaneously expanding its capacity, as described in [69].

Extensions to the Organization Theory Model:

To include the management of legacy network equipment in the model proposed by [5], such that the network consumption becomes relatively proportional to the traffic workload and the energy savings contribute to the overall efficiency of the system, it is proposed to add the following elements to its architecture:

1) Management Roles

Add to the "System Operations" components the "Network Equipment Management" role, which acts as an interface between the model and the network equipment being responsible for actions taken on these devices such as: enabling and disabling ports or equipment or change MC-LAG protocol settings.

The "Monitoring Management" role, responsible for collecting structure information and its understanding, should be augmented with elements for interaction with the network management system to provide data, from which decisions can be made about the port speed configuration, or turning on or off components and ports. These decisions will be guided by the rules and beliefs.

2) Planning Rules

These rules are used when decisions must be taken, and therefore, rules to configure the network equipment in accordance with the activation, deactivation and utilization of physical machines should be added.

To implement the settings pointed out in [68], already presented, the following rules are proposed:

- If a PM (Physical Machine) is switched off, the corresponding ports of access layer switches must be turned off.
- If the occupation of a PM is smaller than a preset value, network interfaces and corresponding access switches ports must be slowed down.
- If the aggregate bandwidth of the downlink ports of an access layer switch is smaller than a preset value, their uplink ports must have their speed reduced.
- If an access layer switch has all its ports off, it must be turned off.
- If an access layer switch is turned off, the corresponding ports of the aggregation layer switch must be turned off.
- If the aggregate bandwidth of the downlink ports of an aggregation layer switch is smaller than a preset value, their uplink ports must have their speed reduced.
- If an aggregation layer switch has all its ports off, it must be turned off.

 If an aggregation layer switch is turned off, the corresponding port of the core layer switch must be turned off.

 If a module of a core layer switch has all its ports off, it must be turned off.

 If a core layer switch has all its ports off, it must be turned off.

 All reversed rules must also be included.

The application of these rules does not affect the reliability of the network, since port and devices are only turned off when servers are turned off. The system performance will only be affected if the network equipment activation cost is bigger than the server activation cost.

For more efficiency in traffic consolidation, the model should consider the racks in virtual machines allocation and migration strategies, and rules that consolidate active physical machines in as fewer racks as possible are necessary.

3) Beliefs

They are a set of empirical knowledge used to improve decisions, and are linked to the used resources characteristics and to the type of services implemented in each specific case.

For each of the rules listed in the previous paragraph, a belief related to energy consumption should be stated. If we consider [70], examples include:

 Disconnecting a port on a switch access layer generates a saving of 500 mWh.

 Decreasing the speed of a port from 10 Gbps to 1 Gbps generates a saving of 4.5 Wh.

It will also be necessary to include beliefs about the time required for a deactivated port or device to become operational after the boot. These beliefs will be used to make decisions that must consider performance requirements.

Simulation Model:

The typical data center network topology, rules and beliefs proposed form the basis for building a simulation model to validate different strategies and rules in specific settings and with different workloads. As already done in previous works by [5], it is possible to expand the CloudSim [71] or work on some of its extensions as TeachCloud [72].

The simulator must create the network topology and calculate their initial consumption based on the amount of physical servers using the following rules:

 If the number of servers is smaller than 40, the topology will have only two access layer switches interconnected by their uplink ports. Turn off unused ports.

 If the number of servers is greater than 40 and smaller than 480 (12 Racks), put two access layer switches for every 40 servers or fraction and two aggregation layer switches interconnected by their uplink ports. Turn off unused ports of both layers switches.

 If the number of servers is greater than 480, apply the previous rule for each group of 480 servers or fraction, add two core layer switches and put on each switch a 24 ports module for each 5,760 servers (144 racks) or fraction. Turn off unused port.

The equation to calculate the consumption of the switches and modules is:

Power (W) = BP + no. P 10Giga x 5 + no. P Giga x 0,5 + no. P Fast x 0,3 (1)

In this expression, the power in Watts is calculated by summing the BP (Base Power), which is a fixed value specific to each device, and the consumption of every active port at each speed, which is the variable component. The consumption of each type of port is specific to each device, but the proposed values are the average values according to the works already cited.

In equation (1), if the switch is modular, the base power of the chassis must be added.

During the simulation, when servers are connected or disconnected, the simulator must apply the network management rules by turning on or off the corresponding ports or configuring its speed, and update the calculation of the total consumption of the network.

In order to analyze the system performance and SLA violations, the model must know the time needed to put into operation each type of equipment, and at the moment of the server's activation, compare the uptime of the server with the uptime of the network equipment and use the greatest.

IX. CONCLUSIONS

The work "Decision-Theoretic Planning for Cloud Computing" has presented a decision-theoretic modeling to decision making for CC management in an AC context, using the MDP as a mathematical framework, contributing to the state-of-the-art in CC research in the sense that it tackles the phase planning of an autonomic cycle with a mathematical model which takes into consideration the uncertainty of action resulting in complex systems such as a CC management systems. For future work, the following steps will be considered: A big data model to feed the transition function created with the monitoring data bases; Extend the CloudStack to implement the model on its resource manager and perform experiments to observe the performance of the model in taking decisions; Analyze a meta-management model to optimize the autonomic planner; and Research methods of action discovery and learning.

The work "An Architecture for Risk Analysis in Cloud" presented a model of shared responsibilities for risk analysis in cloud computing environments. In addition to the traditional CC and CSP agents the model adds the ISL agent, which is responsible for identifying and specifying the security requirements. The model presented in this paper is an initiative to allow the CC can perform the risk analysis on its current or future CSP, and this risk analysis is broad, current, unbiased and reliable. The characteristics presented in this article aim at generating a more reliable risk analysis for CC, so that it can choose its CSP based on more solid information. Several papers on cloud computing indicate the lack of trust from CC to CSP as a relevant factor in avoiding the purchase of cloud computing services. A risk analysis can act to reduce or eliminate this suspicion and boost the acquisition of cloud computing services. The presented model performs a free and reliable risk analysis, because the analysis is not centered in the CSP. The identification and quantification of threats and vulnerabilities are carried out collaboratively by several

laboratories. Safety and impact on information assets are quantified by the CC. The risk analysis of the proposed model is broad, because the security requirements are defined by specialized laboratories and the CC itself defines and quantifies their information assets. It is dynamic, because the various ISLs can modify their security requirements for considering new vulnerabilities in future risk analyzes. This work opens possibilities for the development of future research. There is a need for research on the reliability of the data reported between CSP and ISL during risk analysis. The RDL - Risk Definition Language can be further explored in specific jobs. Further research should be done on the inferences on the results of risk analysis. These inferences can help all stakeholders in understanding the causes of incidents and their solutions. Finally, there is the need to extend this work in order that the proposed model can also suggest the controls or countermeasures to the CSPs.

The work “Risk-based Dynamic Access Control for a Highly Scalable Cloud Federation” presented a risk-based dynamic access control system to enable cloud federations without the need, but allowing the possibility, of identity federations. By eliminating the need for identity federations our proposal eases the creation of cloud federations, since it doesn't depend on the establishment of agreements and circles of trust, also enhancing scalability, by avoiding the formation of “identity islands” [44]. The main contributions of this paper are the definition of a risk-based access control system for cloud federations and the proposed use of risk policies in the form of XML files to allow the use of different risk metrics and quantification methods that are not necessarily predefined. The proposal is flexible enough to handle the needs of a cloud federation and the performance evaluations indicate that it is scalable and that the risk estimation process is not a big hindrance in the process, especially if the quantification is performed locally. In comparison to the related works we first have to clarify that we have not implemented a whole cloud federation system, since it is a huge task and not our focus. We have, however, described and implemented a simple federation model that is sufficient for our access control research and we can highlight that our proposal is the only that uses risk-based access control. Also, we still allow the use of identity federations, but offer a choice of establishing the cloud federation without the need for Federated Identity Management. Compared to the works that deal with risk-based access control in cloud, our approach has the advantage of allowing the resource owner to choose different risk quantification and aggregation engines through a risk policy definition file, also the cloud that hosts the resource can define a baseline risk policy, to ensure its minimum security requirements are met. As future work we foresee the possibility of enlarging the federation used in our experiments and deploying it to real use. Also we want to explore further the use of the risk policies with different risk metrics and quantification methods.

The work “Challenges of Operationalizing PACS on Cloud Over Wireless Networks” presented the need for finding more precise diagnostics allowing effective treatment for patients pushes for a constant technological evolution in medical equipment, as well as smartphones, tablets and laptops that are used to access the images and the communication of these devices with the cloud. The presence of communication resources in daily life over Wi-Fi, 3G, 4G or WiMax reaching high data transmission rates, enables access to medical diagnostics at a distance using Internet-connected mobile devices, downloading DICOM images with an appropriate application. The use of a cloud-based PACS has the goal of showing the archiving of medical exam images from different locations in a centralized repository, lowering the investments on storage and processing infrastructure for hospitals and clinics. On the cloud, doctors and patients may visualize these images through any connected mobile device that provides Internet access. The test results were satisfactory, considering data transmission rates, showing that mobile devices and a cloud-based PACS present a viable solution for the practice of telemedicine.

The work “Environment, Services and Network Management for Green Clouds” proposed an integrated model of environment, services and network management for green clouds based on organization model of autonomous agent components. Concepts related to cloud computing and green cloud computing were presented. We demonstrated that the proposed solution delivers both reliability and sustainability, contributing to our goal to optimize energy utilization. Tests were realized to prove the validity of the system by utilizing the CloudSim simulator from the University of Melbourne in Australia. We have implemented improvements related to service-based interaction. We implemented migration policies and relocation of virtual machines by monitoring and controlling the system. We achieved the following results in the test environment: Dynamic physical orchestration and service orchestration led to 87,18% energy savings, when compared to static approaches; and Improvement in load balancing and high availability schemas provide up to 8,03% SLA error decrease. We are building a unified power management strategy for green cloud computing, minimizing the total power consumed by including network device power, server power, and cooling power.

The work “Provisioning and Resource Allocation for Green Clouds” presented strategies for allocation and provisioning, both aimed at optimizing the energy resource without sacrificing service availability. The allocation strategy in private clouds, compared to a normal cloud, demonstrated a 87% reduction in energy consumption. It was observed that this strategy is not effective in scenarios where the workload is oscillating. That’s because it ends up generating too much unnecessary reconfigurations and migrations. Despite this, it still shows a significant gain in energy savings when compared to a cloud without any strategy deployed. The hybrid strategy for provisioning in green clouds, demonstrated a 52% consumption reduction over the SR strategy, and a timeout rate 3% lower than the OD strategy. Thus, we conclude that the use of this strategy is recommended in situations where the activation time of the resource is expensive for the health of SLA. We also identified that using

this is not recommended when the public cloud should be used sparingly due to their course or other factors. As future work, we aim at adding the strategy of Dynamic Reconfiguration of VMs in public clouds. This procedure was not adopted because, during the development of this work, this feature was not a market reality. We also intend to invest in new simulations of the cloud extending the variables (such as DVFS and UPS) and, if possible, explore some artificial intelligence techniques such as Bayesian networks, the recalculation of beliefs. Our PCMONS (Private Cloud Monitoring System), open-source solutions for cloud monitoring and management, also will help to manage green clouds, by automating the instantiation of new resource usage. We foresee, in opposition to unexpected peaks scenarios, work with cloud management based on prior knowledge of the behavior of hosted services. It is believed to be necessary to develop a description language to represent the structure and behavior of a service, enabling the exchange of information between applications for planning, provisioning, and managing the cloud.

The work "Optimizing Green Clouds through Legacy Network Infrastructure Management" presented basic concepts related to Green IT were first presented, i.e., Green Cloud and Green Networking, demonstrating the need of considering the network equipment in strategies designed to make data centers more efficient, since the network represents a significant percentage of total consumption, and this participation will be more expressive when the other components become more efficient. A green cloud management model called OTM (Organization Theory Model) was presented, as well as network equipment management principles that, when properly applied, make the behavior of the total consumption of the network approximately proportional to the traffic load, even when legacy energy-agnostic equipment are used in. The proposal was to extend the OTM to manage the network traffic consolidation according to these management principles. Then, the elements that must be added to the architecture of the OTM were described, including the rules and beliefs required for the correct network configuration according to the load consolidation on servers. It was also proposed a model to determine the data center network topology based on the number of physical servers, the rules to manage and set the network devices according to the servers' state changes, and equations to calculate the switches consumption and the total network consumption. This model is the basis to create a simulator and perform simulations to test the viability and the impact of the proposal application in different configurations, with different performance requirements and with different rules and beliefs. The model was validated by its application in a case study, which allowed verifying that equations and rules are correct and enough to create the topology and to calculate the consumption of the network in each step of the simulation, as well as highlight the possible effects of the application of the proposal. It was also demonstrated, that in the described scenario it is possible to get a power saving of approximately 11% only by the proper initial configuration of the network and without any compromise of the performance. In a hypothetical situation of low utilization, a power saving of approximately 45% through proper workload consolidation is possible. It was thus demonstrated the possibility and desirability of extending the green cloud management model as proposed. As future research, it is proposed to continue this work by developing the necessary extensions to CloudSim to implement the model, and perform experiments to determine the most effective rules and virtual machine allocation policies, and the actual contribution of the model in scenarios with different configurations, real workloads and taking into account possible violations to the SLA. To evaluate the applicability of the model, it is also proposed to determine, through simulation, how many times a day a port or a device is turned on and off in real scenarios, and its possible impact in equipment failure rate. Finally, since system performance may be affected if the network devices activation cost is bigger than the server activation cost, it is also suggested to study the proper network configuration and technologies to avoid this situation, with special consideration to protocols that manage the links redundancy and aggregation, like the Spanning Tree Protocol, MC-LAG, and other new networking standards for data centers.

REFERENCES

[1] R. S. Mendes, R. Weingartner, G. A. Geronimo, G. B. Brascher, A. A. Flores, C. B. Westphall, C. M. Westphall. "Decision-Theoretic Planning for Cloud Computing," *in Thirteenth International Conference on Networks - ICN* 2014.

[2] P. F. Silva, C. B. Westphall, C. M. Westphall, M. M. Mattos. "An Architecture for Risk Analysis in Cloud," *in Tenth International Conference on Networking and Services - ICNS* 2014.

[3] D. R. Santos, C. M. Westphall, C. B. Westphall. "Risk-based Dynamic Access Control for a Highly Scalable Cloud Federation," *Seventh International Conference on Emerging Security Information, Systems and Technologies - SECURWARE* 2013.

[4] R. F. Souza, C. B. Westphall, D. R. Santos, C. M. Westphall. "Challenges of Operationalizing PACS on Cloud Over Wireless Networks," *Ninth International Conference on Wireless and Mobile Communications – ICWMC*. 2013.

[5] J. Werner, G. A. Geronimo, C. B. Westphall, F. L. Koch, R. R. Freitas, C. M. Westphall. "Environment, Services and Network Management for Green Clouds," i*n CLEI ELECTRONIC JOURNAL, Vol, 15, Number 2, Paper 2, Aug.* 2012.

[6] G. A. Geronimo, J. Werner, C. B. Westphall, C. M. Westphall, Leonardo Defenti. "Provisioning and Resource Allocation for Green Clouds," *in Twelfth International Conference on Networks – ICN*. 2013.

[7] S. R. Villarreal, C. B. Westphall, C. M. Westphall. "Optimizing Green Clouds through Legacy Network Infrastructure Management," *in Thirteenth International Conference on Networks – ICN*. 2014.

[8] W. Walsh, G. Tesauro, J. Kephart, R. Das. "Utility functions in autonomic systems", *in Autonomic Computing, 2004. Proceedings. International Conference on, May 2004, pp. 70–77.*

[9] S. Dobson, R. Sterritt, P. Nixon, and M. Hinchey. "Fulfilling the vision of autonomic computing," *in Computer, vol. 43, no. 1, Jan. 2010, pp. 35–41.*

[10] Apache Foundation. Apache CloudStack, 2013 retrieved in September 2013 from http://cloudstack.apache.org/

[11] P. Mell, T. Grance. The NIST Definition of Cloud Computing, Tech. rep., National Institute of Standards and Technology, Information Technology Laboratory, Jul. 2009.

[12] G. Aceto, A. Botta, W. de Donato, A. Pescape`, "Cloud monitoring: A survey", Computer Networks, vol. 57, issue 9, Jun. 2013, pp. 2093–2115.

[13] A. Viratanapanu, A. Hamid, Y. Kawahara, T. Asami, "On demand fine grain resource monitoring system for server consolidation",

Kaleidoscope: Beyond the Internet? - *Innovations for Future Networks and Services*, 2010 ITU-T, Dec. 2010, pp. 1–8.

[14] R. B. Bohn, J. Messina, F. Liu, J. Tong, J. Mao, "Nist cloud computing reference architecture", *in: Services (SERVICES), 2011 IEEE World Congress on*, July 2011, pp. 594–596.

[15] M. Litoiu, M. Woodside, J. Wong, J. Ng, G. Iszlai, "A business driven cloud optimization architecture," *In Proceedings of the 2010 ACM Symposium on Applied Computing (SAC '10)*, 2010, pp. 380–385.

[16] S. Leimeister, M. Bhm, C. Riedl, H. Krcmar, "The business perspective of cloud computing: Actors, roles and value networks," *In Proceedings of the 7th international conference on Economics of grids, clouds, systems, and services (GECON'10)*, 2010, pp. 129–140.

[17] S. Marston, Z. Li, S. Bandyopadhyay, J. Zhang, A. Ghalsasi, "Cloud computing - The business perspective," *Decis. Support Syst. vol. 51, no. 1, Apr. 2011, pp. 176–189.*

[18] A. B. Letaifa, A. Haji, M. Jebalia, S. Tabbane, "State of the Art and Research Challenges of new services architecture technologies: Virtualization, SOA and Cloud Computing," *International Journal of Grid & Distributed Computing, vol. 3, issue 4, Dec. 2010, pp. 69–88.*

[19] C. Tan, K. Liu, L. Sun, "A design of evaluation method for saas in cloud computing," *in Journal of Industrial Engineering and Management, vol. 6, no. 1, Feb. 2013, pp. 50–72.*

[20] J. Kephart, D. Chess, "The vision of autonomic computing," *in Computer, vol. 36, no. 1, Jan. 2003, pp. 41–50.*

[21] S. Gutirrez, J. Branch, "A comparison between expert systems and autonomic computing plus mobile agent approaches for fault management," *in DYNA, vol. 78, no. 168, Aug. 2011, pp 173–180.*

[22] M. Puterman, Markov Decision Processes: Discrete Stochastic Dynamic Programming, Wiley Series in Probability and Statistics, Wiley, 2009.

[23] B. Lindgren, Elements of decision theory, Macmillan, 1971.

[24] Wikipedia, Markov Decision Process, 2013 retrieved in September 2013. From http://en.wikipedia.org/wiki/Markov decision process.

[25] D. Durkee, "Why Cloud Computing Will Never Be Free," *Queue Journal vol. 8 no. 4, Apr. 2010, pp. 20-29.*

[26] U. Sharma, Elastic resource management in cloud computing platforms, Ph.D. thesis, University of Massachusetts, May 2013.

[27] M. K. Srinivasan et al., "State-of-the-art cloud computing security taxonomies: a classification of security challenges in the present cloud computing environment". ICACCI '12: Proceedings of the International Conference on Advances in Computing, Communications and Informatics. August 2012.

[28] H. Yu et al., "Cloud computing and security challenges". ACM- SE '12: Proceedings of the 50th Annual Southeast Regional Conference. March 2012.

[29] K. Ren, C. Wang and Q. Wang, "Security Challenges for the Public Cloud," Internet Computing, IEEE, vol.16, no.1, pp.69, 73, Jan.-Feb. 2012 doi: 10.1109/MIC.2012.14.

[30] B. Grobauer, T. Walloschek and E. Stocker, "Understanding Cloud Computing Vulnerabilities," Security & Privacy, IEEE , vol.9, no.2, pp.50,57, March-April 2011 doi: 10.1109/ MSP.2010.115.

[31] ISO/IEC 27005:2011, Information Security Risk Management. [Online]. Available: http://www.iso.org.

[32] J. Zhang, D. Sun and D. Zhai, "A research on the indicator system of Cloud Computing Security Risk Assessment," Quality, Reliability, Risk, Maintenance, and Safety Engineering (ICQR2MSE), 2012 International Conference on , vol., no., pp.121,123, 15-18 June 2012 doi: 10.1109/ ICQR2MSE. 2012.6246200.

[33] M. L. Hale and R. Gamble, "SecAgreement: Advancing Security Risk Calculations in Cloud Services," Services (SERVICES), 2012 IEEE Eighth World Congress on , vol., no., pp.133-140, 24-29 June 2012 doi: 10.1109/SERVICES.2012.31.

[34] J. Morin, J. Aubert and B. Gateau, "Towards Cloud Computing SLA Risk Management: Issues and Challenges," System Science (HICSS), 2012 45th Hawaii International Conference on , vol., no., pp.5509-5514, 4-7 Jan. 2012 doi: 10.1109/HICSS.2012.602.

[35] S. Ristov, M. Gusev and M. Kostoska, "A new methodology for security evaluation in cloud computing," MIPRO, 2012 Proceedings of the 35th International Convention , vol., no., pp.1484-1489, 21-25 May 2012.

[36] J. Chen, Y, Wang and X. Wang, "On-Demand Security Architecture for Cloud Computing," Computer, IEEE, vol.45, no.7, pp.73,78, July 2012 doi: 10.1109/MC.2012.120.

[37] P. Zech, M. Felderer and R. Breu, "Towards a Model Based Security Testing Approach of Cloud Computing Environments," Software Security and Reliability Companion (SERE-C), 2012 IEEE Sixth International Conference on , vol., no., pp.47,56, 20- 22 June 2012 doi: 10.1109/SERE-C.2012.11.

[38] P. Wang et al., "Threat risk analysis for cloud security based on Attack-Defense Trees," Computing Technology and Information Management (ICCM), 2012 8th International Conference on, vol.1, no., pp.106-111, 24-26 April 2012.

[39] P. Mell, T. Grance, "The NIST Definition of Cloud Computing", 2011. Available at: http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf.

[40] E. Carlini, M. Coppola, P. Dazzi, L. Ricci and G. Righetti, "Cloud Federations in Contrail", Proc. Euro-Par 2011: Parallel Processing Workshops, pp. 159-168, 2012.

[41] B. Rochwerger, D. Breitgand, E. Levy, A. Galis, K. Nagin, I. M. Llorente, R. Montero, Y. Wolfsthal, E. Elmroth, J. Cáceres, M. Ben-Yehuda, W. Emmerich and F. Galán, "The reservoir model and architecture for open federated cloud computing", IBM J. Res. Dev., vol. 53, no. 4, pp. 535-545, july 2009.

[42] T. Kurze, M. Klems, D. Bermbach, A. Lenk, S. Tai and M. Kunze, "Cloud Federation", The Second International Conference on Cloud Computing, GRIDs, and Virtualization, pp. 32-38, september 2011.

[43] A. Ferreira, D. Chadwick, P. Farinha, R. Correia, G. Zao, R. Chilro and L. Antunes, "How to Securely Break into RBAC: The BTG-RBAC Model", Proc. ACSAC '09, pp. 23-31, 2009.

[44] K. Lampropoulos, S. Denazis, "Identity management directions in future internet", IEEE Communications Magazine, vol. 49, no. 12, pp. 74-83, december 2012

[45] World Health Organization. http://www.who.org. [retrieved: December 2008].

[46] H. A. Franke, F. L. Koch, C. O. Rolim, C. B. Westphall and D. O. Balen, "Grid-M: Middleware to Integrate Mobile Devices, Sensors and Grid Computing," Third International Conference on Wireless and Mobile Communications, March 2007, pp. 19-25.

[47] E. M. B. Junior. "Teleradiology: Central Remote Diagnostic Imaging Digital Integrated Portal to a Distributed Medical Information. Application of public," Federal University of São Paulo, São Paulo, 2009.

[48] L. He, X. Ming, W. Ding and Q. Liu, "A Novel Approach To Remote Access Picture Archiving And Communication System On Mobile Devices Over Wireless Networks," Biomedical and Health Informatics (BHI), 2012 IEEE-EMBS International Conference on January 2012, pp. 581-583.

[49] Open Source Clinical Image and Object Management. http://www.dcm4chee.org. [retrived: May 2013].

[50] M. J. Warnock, C. Toland, D. Evans, B. Wallace and P. Nagy, "Benefits of Using the DCM4CHE DICOM Archive," Journal of Digital Imaging in November 2007, pp. 125-129.

[51] Valancius, V., Laoutaris, N., Massoulie, L., Diot, C., and Rodriguez, P. (2009) "Greening the internet with nano data centers," in CoNEXT '09: Proceedings of the 5th international conference on Emerging networking experiments and technologies. New York, NY, USA: ACM, pp. 37–48.

[52] Gruber, C. G. (2009) "Capex and opex in aggregation and core networks," in Optical Fiber Communication Conference. Optical Society of America, pp. 1–3.

[53] Lefevre L. and Orgerie, A.-C. (2010) "Designing and evaluating an energy efficient cloud," The Journal of Supercomputing, vol. 51, pp. 352–373.

[54] Buyya, R., Beloglazov, A. and Abawajy, J. (2010) "Energy-Efficient management of data center resources for cloud computing: A vision, architectural elements, and open challenges," in Proceedings of the 2010 International Conference on Parallel and Distributed Processing Techniques and Applications (PDPTA 2010), Las Vegas, USA, July 12, vol. 15.

[55] Liu, L., Wang, H., Liu, X., Jin, X., He, W. B., Wang, Q. B. and Chen, Y. (2009) "Greencloud: a new architecture for green data center," in ICAC-INDST '09: Proceedings of the 6th international conference industry session on Autonomic computing and communications industry session. New York, NY, USA: ACM, 2009, pp. 29–38.

[56] Buyya, R., Ranjan, R. and R. Calheiros, R. (2010) "Intercloud: Utility-oriented federation of cloud computing environments for scaling of application services," in Proceedings of the 10th International Conference on Algorithms and Architectures for Parallel Processing. LNCS, Springer.

[57] Dignum, F., Dignum, V., Padget, J. and Vazquez-Salceda, J. (2009) "Organizing web services to develop dynamic, flexible, distributed systems," in iiWAS'09: Proceedings of the 11th International

Conference on Information Integration and Web-based Applications & Services. New York, NY, USA: ACM, pp. 225–234.
[58] Murugesan, S. (2008) "Harnessing green it: Principles and practices," IT Professional, vol. 10, no. 1, pp. 24–33.
[59] Leandro, M. A. P., Nascimento, T. J., dos Santos, D. R., Westphall, C. M. and Westphall, C. B. (2012) "Multi-tenancy authorization system with federated identity for cloud-based environments using shibboleth," in ICN 2012, The Eleventh International Conference on Networks, 2012, pp. 88–93.
[60] OpenCC, "Open cloud consortium," 2012, "[Online; Last access: 2013-01-15]". [Online]. Available: http://opencloudconsortium.org/
[61] OCCI, "Open cloud computing interface," 2012, "[Online; Last access: 2013-01-15]". [Online]. Available: http://www.occi-wg.org
[62] Foster, I., Zhao, Y., Raicu, I. and Lu, S. (2008) "Cloud computing and grid computing 360-degree compared," in Grid Computing Environments Workshop. GCE 08, nov. 2008, pp. 1–10.
[63] Magklis, G., Semeraro, G., Albonesi, D., Dropsho, S., Dwarkadas, S. and Scott, M. (2003) "Dynamic frequency and voltage scaling for a multiple-clock- domain microprocessor," Micro, IEEE, vol. 23, no. 6, pp. 62–68.
[64] Wang, Q., Kanemasa, Y., Li, J., Lai, C. A., Matsubara, M. and Pu, C. (2013) "Impact of dvfs on n-tier application performance," in Conference on Timely Results in Operating Systems (TRIOS). ACM.
[65] Basoglu, M., Orshansky, M. and Erez, M. (2010) "Nbti-aware dvfs: A new approach to saving energy and increasing processor lifetime," in Low- Power Electronics and Design (ISLPED), 2010 ACM/IEEE International Symposium on, 2010, pp. 253–258.
[66] Beloglazov, A., Buyya, R., Lee, Y.C. and Zomaya, A. (2011) "A taxonomy and Survey of Energy-efficient Datacenters and Cloud Computing". Advances in Computers, vol 82, pp. 47-111, Elsevier, November.
[67] Bianzino, A., Chaudet, C., Rossi, D. and Rougier, J. (2012) "A survey of Green Networking research". IEEE Communications Surveys and Tutorials, vol 14, pp. 3-20, February.
[68] Mahadevan, P., Banerjee, S., Sharma, P., Shah, A., Ranganathan, P. (2011) "On Energy Efficiency for Enterprise and Data Center Networks," in IEEE Comunications Magazine. August.
[69] Sher Decusatis, C. J., Carranza, A. and Decusatis, C. M. (2012) "Comunication within clouds: open standards and proprietary protocols for data center networking", IEEE Communication Magazine. Vol. 50, pp. 26-33, September.
[70] Christensen, K., Reviriego, P., Nordman, B., Mostowfi, M. and J. Maestro, J. (2010) "IEEE 802.3az: The road to Energy Efficient Ethernet", IEEE Comunication Magazine, vol 48, pp. 50-56, November.
[71] Calheiros, R. N., Ranjan, R., Beloglazov, A., De Rose, C. A. F. and Buyya, R. (2011) "Cloudsim: A toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms," Software: Practice and Experience, vol. 41, pp. 25–50.
[72] Jararweh, Y., Kharbutli, M. and Alsaleh, M. (2013) "TeachCloud: A Cloud Computing Educational Toolkit". International Journal of Cloud Computing (IJCC), Vol. 2, No. 2/3, February 2013, pp. 237-257.

Carlos B. Westphall is a full professor in the Department of Informatics and Statistics at the Federal University of Santa Catarina, where he is the leader of the Networks and Management Laboratory. His research interests include network and service management, security, and cloud computing. He received his D.Sc. in computer science at Paul Sabatier University, France. He was the founder of LANOMS. In 2011 he was named an IARIA Fellow. He has served as Technical Program and/or Organizing Committee member (since 1994) of IFIP/IEEE IM, IEEE/IFIP NOMS, IEEE LANOMS, and IEEE APNOMS. He has been an Editorial Board member (since 2004) of the Computer Networks Journal of Elsevier. Since 1993 he has been a member of IFIP TC6 Working Group 6.6 (Management of Networks and Distributed Systems). Since 2008 he has been Latin America International Academy, Research, and Industry Association (IARIA) Liaison Board Chair. He was a member (2004–2005 and 2006–2007) of the IEEE ComSoc Membership Programs Development Board. From May 2000 to May 2005 he acted as Secretary of the IEEE Committee on Network Operation and Management (CNOM). From May 2005 to May 2009 he acted as Vice- Chair of IEEE CNOM. He has been a member of IEEE CNOM since 1994.

Carla M. Westphall is a professor in the Department of Informatics and Statistics at the Federal University of Santa Catarina, Brazil. Her research interests include distributed security, identity management, and grid and cloud security. Westphall received her PhD in electrical engineering from the Federal University of Santa Catarina. Contact her at carlamw@ inf.ufsc.br.

Fernando L. Koch is a Director R&D at SAMSUNG Research Institute Brazil. He was a Research Scientist at IBM Research Brazil (2011-2013) where he received the IBM Eminence and Excellence Award (2012) and the IBM Outstanding Contributor Award (2013) for leadership in research. He has over 20 years of IT Industry experience with practice in R&D, product development, and business development with companies in the Silicon Valley, Europe, Brazil, and Australia. He received the Ph.D in Computer Sciences (2009) from Utrecht University in collaboration with The University of Melbourne. He participated in the PostDoc (2010) in Computer Sciences at Utrecht University. He also holds M.Sc. (1997) and B.Sc. (1993) degrees in Computer Sciences by the Federal University of Santa Catarina, Brazil. He has over 50 papers published and more than 20 patents. He is IEEE Senior Member and ACM Distinguished Speaker. His research interests include Artificial Intelligence, Mobile Computing, Computational Social Sciences, and Cognitive Computing.

Guilheme A. Geronimo is doing his doctoral theses in Computer Science at Federal University of Santa Catarina. He splits his time between the university Data Center and his doctoral theses in Computer Science. He is studing new ways turn the cloud in a sustainable environment.

Jorge Werner is doing his PhD degree in Computer Science at Federal University of Santa Catarina. He has a Master degree in Computer Science at Federal University of Santa Catarina and graduated from at Estacio de Sa University of Santa Catarina in Computer Network Technology. His research interests include distributed security, identity management, and cloud security.

Rafael S. Mendes is doing his PhD degree in Computer Science at the Networks and Management Laboratory at the Federal University of Santa Catarina. He is currently researching about decision making to Cloud Computing in Autonomic and Big Data environments.

Paulo F. Silva has a Master degree in Computer Science (PPGCC / UFSC). His a Doctoral student in Computer Science (PPGCC / UFSC). Researcher of the Network and Management Laboratory (LRG / UFSC). Professor in the Department of Systems and Computing (DSC / FURB). Researcher in Software Quality (LQS / FURB) and Development and Technology Transfer Laboratories (LDTT / FURB).

Daniel R. Santos is a PhD student in Computer Science at the Università degli Studi di Trento. Bachelor and Master in Computer Science from the Federal University of Santa Catarina. Member of the Laboratory of Network and Management at the Department of Informatics and Statistics of UFSC and researcher focusing on Computer Security, Identity Management and Access Control and Cloud Computing.

Mauro M. Mattos has Doctoral degree in Production Engineering from the Federal University of Santa Catarina (2003), Masters in Computer Science from the Federal University of Rio Grande do Sul (1988) and a degree in Data Processing Technologist in the Universidade do Vale do Rio dos Sinos (1984). He is associate professor of the Regional University of Blumenau since 1995.

Ricardo F. Souza is an associated professor at University Center of Brusque and Uniasselvi Group. Has over six years working as IT manager and administration of RIS / PACS systems. He holds the titles of M.Sc. (2013) from the Federal University of Santa Catarina and B.Sc. (2005) in Computer Science from Regional University of Blumenau. His research interests include computer science applied to health and education, security and computer networks.

Sergio R. Villarreal is a Computer Engineer graduated at the Superior Technical Institute (IST), Argentina, and is doing his master degree in Computer Science at the Networks and Management Laboratory (LRG) of the Federal University of Santa Catarina (UFSC), Brazil. He is currently an associated professor at the University of the State of Santa Catarina (UDESC)

Rafael Weingärtner obtained his bachelor degree in Information System at UNISUL University (2012). He is currently enrolled at the graduate program in Computer Science at Federal University of Santa Catarina (UFSC), in which he is pursuing a master degree in Computer Science.

Leonardo Defenti is doing his master degree in Computer Science at the

Networks and Management Laboratory (LRG) in Federal University of Santa Catarina. He is curently employed at Petrobras as an Infraestructure Analyst.

Alexandre A. Flores is doing his master degree in Computer Science at the Networks and Management Laboratory (LRG) in Federal University of Santa Catarina. He is curently employed at Eletrosul.

Rafael R. Freitas has received his Bachelor in Computer Science from Federal University of Santa Catarina.

Gabriel B. Bräscher is doing his Bachelor in Computer Science from Federal University of Santa Catarina.