2012 IEEE 4th International Conference on Cloud Computing Technology and Science

Business-Driven IT Management for Cloud Computing Providers

J. Oriol Fitó*†, Mario Macías*†, Ferran Julià† and Jordi Guitart*†
*Barcelona Supercomputing Center and †Universitat Politècnica de Catalunya - Barcelona Tech (UPC)
Jordi Girona 31, 08034 Barcelona, Spain
{fito, mario, fjulia, jguitart}@ac.upc.edu

***Abstract*—Nowadays, enterprises have high expectations on Cloud systems to achieve their Business-Level Objectives (BLOs). However, those systems are becoming very complex, thus leading to critical management issues. For these reasons, new self-management strategies of virtualized entities driven by business interests have to be explored.**

In this paper, we present a business-driven self-management optimization loop, which can firmly contribute to fulfill business strategies of Cloud providers. In this sense, a Business-Driven IT Management (BDIM) model is aimed to assess the impact of IT-related events on business-level metrics. Thereupon those high-level impacts are consumed by a policy management framework able to autonomously determine the most suitable IT-level management actions in terms of provider's BLOs compliance. Furthermore, we show the benefits achieved by a Platform as a Service (PaaS) provider when adopting such business-driven management. The results obtained through experimentation demonstrate that it is able to maximize, individually and simultaneously, two high-level objectives, in this case economical profit and ecological efficiency.

***Keywords*-Cloud computing providers, Business-Driven IT Management (BDIM), Business-Level Objective (BLO)**

I. INTRODUCTION

More than ever, enterprises are focusing their attention on new and emerging IT solutions to efficiently achieve their high-level goals, also known as Business-Level Objectives (BLOs). Instead of just looking at optimizations of managerial appearance, the adoption of groundbreaking ways to improve business interests, strategies, and goals must be put into consideration. In this regard, Cloud computing is widely recognized as the most promising computing paradigm of nowadays [6]. It is based in a model in which computing requirements are offered as a service through the Internet. The use of Cloud systems leads, among others, to promising business models. Actually, benefits for both stakeholders, i.e. providers and end-users, are very clear [4].

Besides, world's data centers need to become greener due to their unacceptable impact on the natural environment. Therefore, an optimized use of virtualized infrastructures is becoming crucial. Apart from the typical aim of maximizing the profit earned, Cloud providers are beginning to have disparate high-level objectives to be accomplished, such as minimizing its ecological impact (e.g. carbon footprint) and maximizing its customers' satisfaction.

For these reasons, an innovative self-management of Cloud environments driven by disparate BLOs is undoubtedly necessary. The motivations for developing such autonomous management are: (1) virtualized infrastructures, used to support deployment and operation of services, are becoming very complex in terms of dimension and management; (2) there are several aspects to deal with in order to efficiently manage those Cloud-related entities during their whole life cycle; (3) for businesses in general, adopting a business-driven IT management (BDIM) model is a significant requirement for an effective IT governance; and, as stated before, (4) the rising expectation from enterprises that IT systems should assist in achieving their BLOs.

In this regard, the usefulness of the BDIM discipline for addressing all these challenging issues is widely recognized [15]. It aims ensuring successful alignment between business and IT systems through a complete understanding of the impact that such systems have on business results, and vice versa. To that end, proper linkage models between IT-level metrics and business indicators are mandatory. Through this linkage and an autonomic monitoring of metrics of interest, Cloud providers will be able to efficiently manage virtualized systems according to well-defined BLOs. Heretofore, there have been some research contributions in designing and applying BDIM models to IT management (see Section II for further information). However, its applicability in the Cloud computing realm is still to be fully dealt with.

As a result, in this paper we firstly contribute with a BDIM model that fits with requirements of Cloud providers relying on virtualized entities. Its main goal is to gauge impacts of IT-level events on the business. Thus, they can be used to determine the most beneficial IT management decision(s) from the business point of view. Secondly, an autonomous business-driven IT management optimization loop, which relies on the outcomes of such model, is presented. Lastly, we aim to prove its applicability in a self-managed PaaS provider which offers execution environments for hosting Web-based services. We present how it improves its BLOs by using such business-driven management when deploying services.

The remainder of this paper is organized as follows: Section II presents some related work. Section III details the proposed approach toward self-managed Cloud providers driven by business aspects. Section IV presents three use cases of a PaaS provider. Section V details the experimental environment and the evaluation of the presented approach. Finally, Section VI exposes paper conclusions and future work.

978-1-4673-4510-1/12/$31.00 ©2012 IEEE

II. Related Work

BDIM is defined in [17] as a new, evolutionary and comprehensive IT management approach aimed to improve IT infrastructure, Quality of Service (QoS) and business results at the same time. It involves a new culture, tools and decision-making processes that help businesses in achieving their BLOs. In this sense, the Cloud computing itself opens a wide range of opportunities for extending traditional BDIM approaches, which mainly lack automation capability.

A main research agenda for BDIM is presented in [15]. Moreover, a vision of a business-driven adaptive IT infrastructure is described in [12]. Almost all ideas presented in these papers serve as the basis of BDIM challenges in the Cloud.

Up to now, autonomous IT management processes driven by business aspects have been proposed in several articles, such as [19]. Aiber et al. [2] present a set of technologies and methodologies enabling such self-optimization according to BLOs. Unfortunately, they only consider the goal of maximizing the income of a single e-commerce site. The ability to deal with changes in IT infrastructures is explained. Even so, they avoid to present a validation test about how to recognize and react to those changes in order to constantly keep the IT infrastructure optimized with regard to BLOs. Hence, approaches like this must be extended in order to be used by autonomous Cloud providers driven by disparate BLOs. Furthermore, some research efforts have also used BDIM methodologies to increase the business value of e-commerce applications, such as [13]. Notably, these approaches do not provide dynamism when allocating resources to services, which is extremely required to deal with typical changes in the environment (e.g. demand variations).

Besides, Sauvé et al. [18] pose the problem of planning and scheduling changes in business-driven IT management. Unfortunately, the evaluation is based on a numerical illustration, and they only address the BLO of minimizing business loss. Moreover, Abi et al. [1] introduce a business-driven framework for managing utility computing environments. However, these last two approaches need human intervention when applying management actions.

The virtualization technology, which is the basis of Cloud environments, allows to overcome all these limitations of traditional BDIM approaches. In fact, the whole Cloud entities life cycle can be self-managed through virtualization-level actions aligned with provider's business objectives. Summarizing, the BDIM discipline represents a promising way for an efficient business-driven IT self-management in Cloud providers, and in several disparate Cloud scenarios. Nevertheless, we have just identified some research challenges, complementary to those detailed in [15], to be addressed to ensure the success of BDIM methods working on top of virtualized environments.

III. Business-driven Cloud Providers

In this section, we present the BDIM model for Cloud providers, as well as the business-driven IT self-management optimization procedure.

A. Adopting a BDIM model for the Cloud

As shown in Figure 1, we contemplate the widely used hierarchical three-layer modeling architecture for building BDIM models (see for instance [2]) as the most suitable one due to requirements of business-driven Cloud providers. However, some goals and responsibilities of its layers have been reinterpreted with the aim to adapt such model to be used by business-driven Cloud providers. In particular, solving a BDIM model needs the definition of general linkage models aimed to correlate low-level metrics with high-level ones. In accordance with the architecture proposed below, two linkage models have to be designed: the *Lower* ($Link_{low}$) and the *Upper* ($Link_{up}$). The former performs the matching between IT-related metrics and business-level parameters (BLPs), while the latter maps those parameters with final business metrics (i.e. BDIM metrics) that quantify and represent BLOs. Moreover, the proposed model is largely based on the Cloud environment described below.

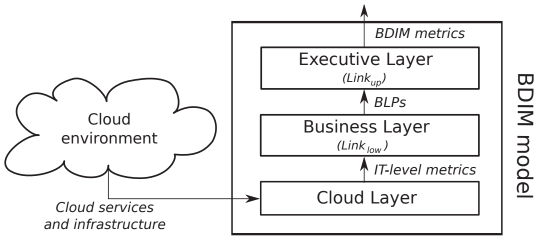

Fig. 1. BDIM model architecture.

Cloud environment. The model considers a set of Cloud services $CS = \{cs_1, ..., cs_{|CS|}\}$ deployed on the underlying virtualized infrastructure. Each one is composed of a cluster of virtual machines $VM(cs_i) = \{vm_1, ..., vm_{|VM(cs_i)|}\}$. All VMs run on a set of physical hosts $H = \{h_1, ..., h_{|H|}\}$. In addition, we consider the involvement of providers in federated Cloud environments. Hence, virtual machines can be deployed and running on in-house resources (vm_k^L) or outsourced (vm_k^O) ones offered by third-party IaaS providers $IaaS = \{iaas_1, ..., iaas_{|IaaS|}\}$. Finally, each physical host can be powered by either the common grid (brown) energy (h_j^B) or renewable (green) energies (h_j^G).

Cloud Layer. It extracts, from such Cloud environment and by means of a monitoring system, IT-level metrics about the underlying Cloud infrastructure and services deployed on it. Some examples of these metrics are: Web-based service's response time (seconds) and availability (percentage of annual uptime); batch jobs deadline (date); and private Cloud's capacity, utilization and power consumption.

Business Layer. It models how IT-related metrics or events (new services deployment, SLA violations, resource failures, ...) affect changes in Business-Level Parameters (BLPs), such as revenue, ecological efficiency, risk, and trust. As mentioned, this matching is performed by the $Link_{low}$. Consequently, it provides to the upper layer the assessment of significant

978-1-4673-4510-1/12/$31.00 ©2012 IEEE

business parameters $BLPs(blo_i) = \{blp_1, ..., blp_n\}$ for each BLO being considered. Moreover, SLAs have an important role in this in-between layer, because it is mainly based on parameters specified in those agreements (e.g. the price paid by the customer).

Executive Layer. It captures variations in BDIM metrics $BDIM = \{bdim_1, ..., bdim_{|BDIM|}\}$, e.g. profit and client's satisfaction, caused by fluctuations on BLPs relevant to each business objective considered $BLOs = \{blo_1, ..., blo_{|BLOs|}\}$. This high-level impact is assessed by the $Link_{up}$. Noteworthy, BDIM metrics are actually the result of the whole BDIM model and must be understood by business executives. In fact, BLOs are directly represented and quantified by such metrics. Moreover, the complexity of this layer increases when multiple BLOs are pursued, since different BDIM metrics are assessed and, therefore, taken into account in decision making processes. In these complex cases, suitable business-driven policies must solve any arising tradeoff. This possibility to define multiple BLOs is a key property for contributing with a general and extensible BDIM model.

B. Policy Management Framework

This framework is the central point of the proposed self-optimization for business-driven Cloud providers. Basically, we contemplate a policy-based management of Cloud entities –i.e. virtualized resources, Cloud services, physical hosts, etc.– which has the potential to free those providers from a huge amount of administrative costs. In addition to this, it allows to fulfill the needs of business on demand. This framework is internally composed of a set of management policies useful to be applied under certain circumstances and events. Actually, the application of one policy or another clearly depends on the status of the provider in a given point in time. Regardless of how those policies are defined and implemented, the framework proposed allows the automated configuration of software actuators aimed to apply IT-level operations.Apart from that, and given the complexity of cloud infrastructures and their entities to be managed, there is also a need of heuristics and greedy algorithms in order to deal with complex decision models. In any case, the next subsection details how such kind of policy management framework can be incorporated, together with BDIM models and low-level actuators, in a closed-loop system aimed to self-manage Cloud providers with the goal to meet their business-level interests.

C. Business-driven IT Self-Management Optimization Process

Going beyond the capabilities and outcomes of the BDIM model just described, we consider its involvement in a closed-loop system (see Figure 2 and the order of its internal interactions). Essentially, it is aimed to perform an efficient IT self-management in terms of BLOs fulfillment. When IT-related events take place, its main goal is to foresee consequences on business results of the possible solution space $S = \{s_1, .., s_{|S|}\}$ that seems appropriate to ameliorate provider's BLOs. Considering the set of services (i.e. virtual machines) and physical hosts present in the provider, the time complexity is $O(|H| \times \sum_{i=0}^{|CS|} |VM(cs_i)|)$. In any case, each instantaneous BLO fulfillment can be determined by the *Policy Management Framework* as the expected difference in the corresponding BDIM metric due to a particular set of management actions s_r, and with regard to the metric value given the current status ($bdim_i(s_o)$):

$$\delta Ful(blo_i) = \frac{bdim_i(s_r) - bdim_i(s_o)}{bdim_i(s_o)} \text{ (where } i \text{ and } r \geq 1)$$

If $bdim_i(s_o) = 0$, $\delta Ful(blo_i)$ is considered maximum. The result of this function is a percentage of improvement in accomplishing a given BLO. Thus, the highest positive value indicates the best set of management actions. Then the overall instantaneous BLOs fulfillment score is equal to the sum of each individual one multiplied by a weight of importance: $\delta Ful(BLOs) = \sum_{i=1}^{|BLOs|} \delta Ful(blo_i) * w(blo_i)$. In this sense, different priorities to BLOs will be properly reflected in decisions taken by management policies and, therefore, in the BLOs fulfillment value. Noteworthy that such weights are applied to $\delta Ful(blo_i)$, which are expressed as a percentage, regardless of the BLOs units.

It is worth noting that BLOs, business-driven policies and the BDIM model must be defined a priori by the business *(0)*. In fact, this is the minimum and inevitable human intervention in the whole BDIM optimization process described below, which operates completely autonomously thereafter. Actually, the BDIM model and policies can be modified during the process without affecting its proper operation.

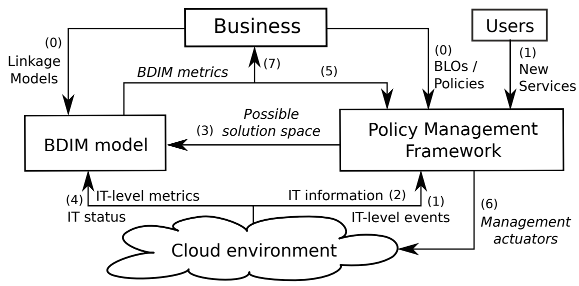

Fig. 2. Business-driven IT self-management optimization closed-loop.

While new BLOs are not defined, the optimization process is as follows: when either an IT-level event takes place or an end-user asks for a new service to be deployed (1), an adaptation of the Cloud environment is required. For this purpose, the policy management framework, based on the current status of the Cloud environment *(2)*, sends to the BDIM model the possible solutions space to face this adaptation *(3)*. Afterward, such model, which also considers IT-level metrics *(4)* representing the Cloud environment status, provides to business-driven policies the foreseen impacts (i.e. BDIM metrics) that potential solutions will have on provider's BLOs if they are carried out *(5)*. The determination of those predictions plays a unique role in the decision making process done by management policies. In fact, those are directly consumed by the latter, which assesses the instantaneous BLOs fulfillment of each possible solution: $I(S, BLOs) = \{< s_1, \delta Ful(BLOs) >$

978-1-4673-4510-1/12/$31.00 ©2012 IEEE

, ..., $< s_{|S|}, \delta Ful(BLOs) >\}$. Based on these values, the instantaneous BLOs fulfillment of each case can be calculated by comparing them with the achieved by the current solution s_0. Then, it decides to enact IT-level management actions that make up the most efficient solution in terms of business results and, moreover, solves all possible tradeoffs among conflicting objectives *(6)*. As stated before, such decision-making process is managed by policies able to properly configure management components responsible of applying IT-level actions. Anyway, and once an efficient solution is implemented, BDIM metrics are recalculated by the model and sent to the business *(7)*. That governing entity may decide (based on new business results) to modify BLOs, business-driven policies and/or linkage models *(0)*. Finally, only to emphasize that the awareness of future business consequences at the time of decision making is the fact that allows providers to better achieve their BLOs. For clarity purposes, we present the pseudocode (Algorithm 1) of the business-driven IT self-management loop along with the communication steps illustrated in previous Figure 2.

Algorithm 1 Business-driven IT Self-management Loop

Require: $BLOs, policies\ and\ linkage\ models\ defined$ {(0)}
Ensure: $BLO\text{-}driven\ optimization$
while !$(new\ BLO(s))$ **do**
  if $(IT\text{-}level\ event\ \|\ new\ service)$ **then** { {(1)}
    $Management\ Policy \Leftarrow updateITInformation()$ {(2)}
    for all $s_i \in S$ **do**
      $BDIM\ model.simulateBusinessImpacts(s_i)$ {(3)}
      $BDIM\ model \Leftarrow getITMetrics()$ {(4)}
      $Management\ Policy \Leftarrow < s_i, BDIM\ metrics >$ {(5)}
    end for
    $Cloud\ env. \Leftarrow efficientActions(I(S, BLOs))$ {(6)}
    $BDIM\ model.assessNewBusinessImpacts()$
    $Business \Leftarrow BDIM\ metrics$ {(7)}
  end if
end while

IV. Use Cases

This section exemplifies the suitability of the business-driven IT self-management approach, explained just above in Section III-C, in Cloud providers. We present three disparate use cases of a PaaS provider –which offers execution environments to host Web-based services– aimed to improve the achievement of its BLOs at service deployment time. The complexity to cope with this phase of services' lifecycle is $O(N)$ –where N is the number of physical hosts–, which allows us to compute the best case.

A. Use Case I: Profit Maximization

The maximization of the profit is traditionally the most common BLO for executives. We propose the following linkage models and business-driven policy to address this case.

1) Lower linkage model: It contains three low-level linkage functions needed to calculate all the financial metrics of interest (expressed in €/hour): (1) the *revenue* is the sum of incomes obtained for each Cloud service deployed:

$$\delta R(s_r) = \sum_{i=1}^{|CS|} \sum_{j=1}^{|VM(cs_i)|} Inc(vm_j);$$

(2) the amount of *costs* is equal to the price of local and external infrastructures used. The former includes the amortization costs of local hosts (in 4 years), the space (in 10 years) required to deploy them using a price of 2000€$/m^2$, and the electricity consumed as to the annual billing; while the latter comprises only the costs of outsourcing VMs in terms of usage per hour (i.e. pay-as-you-go):

$$\delta C(s_r) = Prc_{local} + Prc_{external} = \sum_{i=1}^{|H|} C_{host}(H_i) + \sum_{j=1}^{|VM^O|} C_{vm}(vm_j^O);$$

and, (3) the *loss* captures the economic income missed attributable to violation of services SLAs. Certainly, it is equal to the sum of penalties of all VMs used by services:

$$\delta L(s_r) = \sum_{i=1}^{|CS|} \sum_{j=1}^{|VM^L(cs_i)|} Pen(vm_j^L)$$

Note that $Pen(vm_j^L)$ is the amount of penalty that the provider has to pay to its clients, which is equal to a percentage of the price paid by them (see Section V-A3 for further information on how determine this percentage):

$$Pen(vm_j^L) = \frac{\%SLA\ penalty(vm_j)}{100} \cdot Inc(vm_j)$$

Generally speaking, an SLA is considered as violated if the agreed QoS, e.g. availability level, is not met.

All these linkage functions are used to calculate the instantaneous value of such economic parameters, either the currently achieved (by s_0) or the foreseen ones (by s_r, where $|S| \geq r \geq 1$). Actually, this is true for the foreseen revenue and costs, which can be assessed directly by simulating new scenarios. However, foreseeable loss can only be estimated. As stated by [17] and [15], in most cases estimations are enough for a lot of decision making. To this end, we define the *tenancy ratio* as follows:

$$T(h_k) = \frac{CPU_{used}(h_k) + CPU_{demand}(vm_{new})}{CPU_{cap}(h_k)}$$

As virtual machines are provided based on the CPU specified by users, this ratio is calculated by comparing the CPU used of a given host (plus the one demanded by the new VM) with the total CPU capacity of that host. Supported by the widely proven fact that servers' response time increases sharply when they are overloaded, the linkage model is able to approximate the expected loss as follows (where vm_j is on h_k):

$$\delta \widehat{L(s_r)} = \sum_{i=1}^{|CS|} \sum_{j=1}^{|VM^L(cs_i)|} Pen(vm_j^L)^{T(h_k)}$$

At the end, those functions provide reasonable estimation accuracy. In general, they are targeted to assist in decision-making processes by quantifying the effect of IT events, like the failure of a given VM, on business-level parameters and, consequently, on BDIM metrics.

2) Upper linkage model: Taking into account the profit maximization BLO *(PMax)*, the linkage function $Link_{up}^f$ is quite simple: it must consider all the financial BLPs (achieved with a particular management solution s_r) detailed above: instantaneous *revenue*, *costs*, and revenue *loss* of the provider. Then, the BDIM metric, i.e. instantaneous *profit*, is defined as: $\delta P(s_r) = \delta R(s_r) - (\delta C(s_r) + \delta L(s_r))$, where $|S| \geq r \geq 0$.

3) Business-driven management policy: In this first use case, we propose the *Profit-efficient (PE)* policy, which is aware of the instantaneous profit maximization achieved when deploying new Cloud services. Basically, the solutions space

978-1-4673-4510-1/12/$31.00 ©2012 IEEE

is composed of each possible deployment option, i.e. in each local node or outsourced provider. This is very important since placing a new service in a given node can affect the performance of other services already running on it and, therefore, increase the loss due to SLA breaches. Anyway, the steps (2) to (7) of the Algorithm 1 take place: the BDIM model provides to the policy management framework the foreseen profit of the possible solutions space; afterward, the *PE* policy decides to apply the one with the highest positive value. This means the option with less cost and loss is chosen, because the revenue obtained per service is fixed.

B. Use Case II: Maximizing Eco-Efficiency

The energy consumption is clearly becoming one of the major concerns for today's data centers. Therefore, an energy-efficient management of Cloud environments is imperative.

1) Lower linkage model: It aims to quantify the BLPs (expressed in KWh) needed to assess the provider's overall eco-efficiency: (1) the *green energy* is the instantaneous amount of energy consumed by nodes powered by renewable energy:

$$\delta E_{green}(s_r) = \sum_{j=1}^{|H|} Pwr(h_j^G) \cdot PUE;$$

(2) the *brown energy* is defined as the previous one, but considering the nodes powered by the common grid energy:

$$\delta E_{brown}(s_r) = \sum_{j=1}^{|H|} Pwr(h_j^B) \cdot PUE;$$

and, (3) the *energy base* E_o is the energy that would be consumed without using any energy-related optimization. Actually, it is calculated as if services had been distributed equally among nodes, and if all of these were powered. It also comprises the energy consumed by outsourced services.

Those BLPs are predicted by considering the future power of each host, which is obtained from the power modeling presented in [8]. In addition, the *Power Usage Effectiveness* (PUE) metric [5] is used to calculate the real energy consumption. It can range from 1.0, which means 100% of efficiency (i.e. all power is used by IT equipment), to infinity. For instance, a PUE of 2.0 indicates that IT equipment uses 50% of the power consumed by the whole data center.

2) Upper linkage model: We define the eco-efficiency as:

$$\delta EcoEff(s_r) = \frac{E_{green}(s_r)/(E_{green}(s_r)+E_{brown}(s_r))}{(E_{green}(s_r)+E_{brown}(s_r))/E_o}$$

The numerator is the degree of renewability ($DoRnw(s_r)$), i.e. the ratio between the green energy used and the total energy consumed; whereas the denominator is the energy efficiency ($EnEff(s_r)$) achieved by comparing such total consumption with the energy base. After simplifying, it can be written as:

$$\delta EcoEff(s_r) = \frac{E_o \cdot E_{green}(s_r)}{(E_{green}(s_r)+E_{brown}(s_r))^2}$$

We believe that a proper eco-efficiency metric has to give a great potential to the use of renewable energies. In this sense, note that the proposed metric is zero if there is no consumption of green energy, regardless of the energy efficiency achieved. Moreover, and considering a private Cloud fully powered by green energy, $\delta EcoEff(s_r) = 1$ means that any optimization to save energy is currently applied; otherwise, the metric's value is greater than one.

3) Business-driven management policy: Typically, energy management has only been aware of technical parameters, such as the power consumption [11]. Thanks to virtualization, the consolidation of virtual machines on the same physical host is a habitual practice to save energy [22]. There are other techniques, e.g. power on/off nodes dynamically or put them in stand-by mode, which also aid in energy efficiency.

However, all these energy-aware management actuators have been used so far by considering only IT-level metrics. We go beyond by proposing the *Eco-efficient (EE)* policy. It uses those same management actuators, but its decisions are guided by high-level (business) facets. To that end, whenever the provider receives a new service to be deployed, it contacts the BDIM model to predict the eco-efficiency of each set of actions conforming the possible solutions space. Hence, it not only maximizes the energy efficiency, but also the use of renewable energies. The former is achieved by selecting deployment options, either in in-house nodes or external providers, that allow consolidating services and powering off underused nodes. The latter is ensured by solutions containing the action of turning off nodes of the data center powered by non-renewable energy. Altogether, the cluster of management actions that contribute most to improve provider's eco-efficiency is chosen by the policy.

C. Use Case III: Multiple BLOs fulfillment

Now we aim to meet, at the same time, the couple of BLOs addressed before: the maximization of provider's profit and eco-efficiency. In cases with multiple BLOs, we suggest to assign weights of importance to each one as a way to weigh the utility for the business.

1) Lower linkage model: In accordance with subsections IV-A and IV-B, it has to assess instantaneous revenue, costs, losses, green and brown energy consumed, and energy base.

2) Upper linkage model: Each individual BDIM metric, i.e. instantaneous profit $\delta P(s_r)$ and eco-efficiency $\delta EcoEff(s_r)$, is assessed as formulated before in subsections IV-A2 and IV-B2, respectively. However, in this case it is important to highlight an extra consequence of the energy used in terms of costs: the price of power used, which depends on both the type of energy used and the country where the nodes of the private Cloud are. The price of renewable energy is actually greater than the conventional one (see next subsection V-A2). This fact needs to be carefully considered since using renewable energies leads to improve *(EEMax)* while diminishing *(PMax)*.

3) Business-driven management policy: In the same way as with the previous policies, the procedure followed is the one detailed in Algorithm 1. Whenever the provider receives a new service, management policies contact the BDIM model to predict the instantaneous BLO fulfillment of each set of management actions conforming the possible solutions space. In any case, those BLO fulfillments are calculated as explained before in the corresponding use case; however, in this case the *Profit and Eco-efficient (PEE)* policy also takes into account the impact of the kind of energy used on the profit. In fact, the foreseen cost is evaluated by considering the different

978-1-4673-4510-1/12/$31.00 ©2012 IEEE

prices between brown and green energies, as well as between themselves depending on the country where private Cloud hosts are located. Afterward, the *PEE* policy determines the best group of IT-level management actions. In this way, the policy is able to solve tradeoffs between multiple BLOs, such as the improvement of the energy efficiency achieved by means of consolidating services may lead to lower provider's profit due to higher SLA violation penalties.

V. EXPERIMENTATION

In this section we present the experimentation carried out with a business-driven PaaS provider. The evaluation of the results obtained in the three use cases, previously detailed in subsections IV-A, IV-B and IV-C, is presented. The improvements in BLOs fulfillment of the business-driven management policies proposed are compared against the most efficient traditional policies. For reasons of fairness, all policies are capable of configuring the same low-level management actuators. In this sense, we can truly evaluate the contributions to provider's BLOs achieved by the fact of being aware of business-level metrics. Contrariwise, decisions taken by traditional (IT-driven) policies are only based on IT-level metrics.

A. Experimental Environment

The experimental scenario is composed of three different stakeholders: (1) *SaaS providers* or *end-users*, who want to deploy their services on the Cloud; (2) the *PaaS provider* (on which we focus) offering execution environments to host Web-based services, i.e. a Cloud hosting provider with its own private Cloud (from now on denoted as PaaSP); and (3) three third-party *IaaS providers* to which the PaaS provider can outsource resources to their public Clouds.

1) EMOTIVE and the EEFSim simulator: EMOTIVE [7] is a middleware aimed to the management of Cloud providers according to different policies. It allows executing Cloud services and provides virtualized environments to them. Moreover, it makes possible to implement and test management policies by using either a real Cloud environment or a simulated one. This experimentation is centered on the latter case, where the EEFSim simulator [10] reproduces the behavior of a real virtualized data center. Note that the validation of this simulator can be found in this last citation.

2) Experimentation Parameters: We simulated a PaaS provider operating over two different private Clouds: a PV solar-powered data center composed of 40 physical nodes and placed in Spain, and another one at Sweden composed of 60 physical nodes powered by common grid power. Thus, we use the corresponding power pricing when assessing operational costs: 0.33€/KWh for the green data center located in Spain, and 0.063€/KWh for the swedish one. Furthermore, the PaaSP has three options for outsourcing resources: *Rackspace Hosting* [16], at a cost of 0.4€/hour; and *Amazon EC2* [3], at a cost of 0.337€/h or 0.162€/hour for on-demand and spot large standard instances, respectively. All of them comprise the costs of computation (2 CPUs), storage (2GB), and bandwidth (5GB of data transfer in/out). With this amount of resources, outsourced services' SLA will never be violated.

Moreover, the energy consumed by the private Cloud is provided by the simulator. It follows a Gaussian distribution $\mathcal{N}(PC,1)$, where PC is the real power consumption obtained with a power meter (see [8] for further information). This fact also implies variability in the energy billing (i.e. provider's costs). Therefore, we present the results with the 95% confidence interval if significant (i.e. equal or greater than 0.01). In order to be fair and somehow evaluate the eco-efficiency achieved when outsourcing to third-party providers, we estimated their energy consumption by using measurements from [20]. We also assume that public Clouds' nodes have an average utilization closer to the maximum. On the one hand, the power consumption of Rackspace's nodes is 302Wh (which corresponds to the AMD Opteron 2356, 2.30 GHz). Actually, the consumption of each outsourced resource was estimated as a quarter of this value, since four of them fit in one physical node. On the other hand, large instances bought at Amazon EC2 Europe run on Xeon E5430, 2.66 GHz, which has a power consumption of 320W. In this case, knowing that each large instance has 4 Elastic Compute Units (ECUs), and a total of 11 ECUs fit in a single host, thus the power consumption per resource is estimated as 116.36W. In any case, conventional data centers run at a PUE of 2.0 [9], and this is the value used for calculating the real energy consumption of data centers. In addition, there are three different prices that clients can pay for the hosting of services in the PaaSP, depending on the level of QoS desired: 0.5€/hour, 0.35€/hour and 0.2€/hour for *Gold*, *Silver* and *Bronze* quality, respectively.

3) SLA violation penalties: The provider in question is subjected to suffer losses if SLAs agreed with its clients are not met. In particular, an SLA i is considered as violated if the response time offered by VMs composing the service, $VM(cs_i)$, is greater than a threshold R_i^{TH} —we used a value of 8 seconds, since this is the limit time a user is supposed to wait for a Web page according to [14]—; as well as if VMs' availabilities are below a minimum value of four nines (99.99%). In this sense, we define two functions to assess those amount of penalties for SLA violations, *Magnitude of Violation* (MoV_j) and *Time of Violation* (ToV_j) respectively. As shown in Figure 3, each QoS level has a different increment ratio (λ, α, β) in terms of percentage of SLA penalty. Moreover, note that the maximum amount of penalty is equal to 125% of the price paid by clients. We guarantee enough compensation to customers if the service offered is far below of what was agreed in the SLA. However, we apply this upper bound in order to avoid malicious users, who can overload services themselves with the aim to receive large amounts of monetary recompenses. Finally, the percentage of SLA penalty for each VM is equal to the sum of both functions: $\%\ SLA\ penalty(vm_j) = ToV_j + MoV_j$

4) Workload and Cloud Service: On one hand, we use a workload obtained from an anonymous European ISP (collected during 2009), which indeed is composed of requests to several services. The workload pattern used is of a whole

978-1-4673-4510-1/12/$31.00 ©2012 IEEE

2012 IEEE 4th International Conference on Cloud Computing Technology and Science

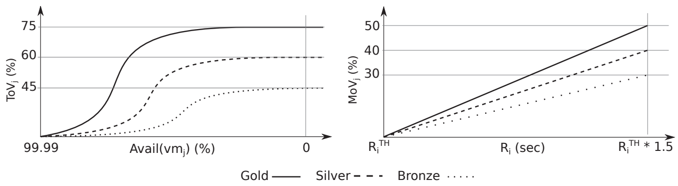

Fig. 3. Categorization of SLA violation penalties.

week, thus representing the typical time-varying users demand over this period. The tests carried out simulate one week. On the other hand, the web-based application used to simulate Cloud services is the SPECweb2009 e-Commerce [21]. Its performance model has been incorporated into the simulator. It was obtained by performing several real tests, on a Tomcat server, in front of different input loads and with different processing units (from 0.25 to 4) available to it (note that such application is CPU-bounded in this environment). Actually, this modeling is focused on the response time metric. A performance pattern which relates this metric with both incoming users' load and CPU usage of the server has been detected.

B. Profit Maximization

This first experiment evaluates the PaaSP's profit earned when using the business-driven and the most efficient IT-level policy (*IT-Out*). We use a workload that deploys a total of 2880 services: 640 lasting one week and 320 lasting one day arriving every day of the week. Table I shows the policy used and the significant economic parameters (in €).

Policy	*Revenue*	*Cost*	*Loss*	*Profit*
IT-Out	68601.10	35610.30±10.69	9145.63	23845.17 ±10.69
PE		24803.52±10.69	0	43797.58 ±10.69

TABLE I
PROFIT MAXIMIZATION OF THE *PE* POLICY.

As shown in the table, the provider's profit when using the *PE* policy is nearly 84% higher than the obtained with the *IT-Out* policy. The former is aware of services for which their owners pay more and, consequently, it ensures a better compliance of their respective SLA. On the contrary, *IT-Out* is only aware of low-level aspects. Commonly, this kind of policies seek for outsourcing deals only when the private Cloud is full, regardless of economic variables associated with services operation (e.g. the revenue obtained). Therefore, the loss paid by this policy is much higher.

C. Eco-Efficiency Maximization

Now we want to evaluate the eco-efficiency achieved by the *Eco-efficient (EE)* policy and the most efficient IT-level policy, i.e. *IT-Backfilling (IT-Bfill)*. The tests performed submit a total of 840 services to the PaaSP during the simulated week: 280 lasting one week and 80 lasting one day arriving every day of the week. With the federated Cloud configuration explained above and this workload, E_o is equal to 9405.10±0.61 KWh. Table II shows the policy used, the total brown and green energy consumption (in KWh), and the weekly eco-efficiency.

Policy	E_{brown}	E_{green}	*EcoEff*
IT-Bfill	4121.53 ±0.61	2037.47 ±0.61	0.51
EE	2036.49 ±0.61	4168.35 ±0.61	1.02

TABLE II
ECO-EFFICIENCY MAXIMIZATION ACHIEVED.

The eco-efficiency obtained by the *EE* policy is about 100% better than the achieved by the *IT-Bfill*. Both try to consolidate services on in-house nodes in order to reduce the total energy consumption. However, the *EE* policy is also aware of filling, as much as possible, the nodes powered by renewable energy; while the *IT-Bfill* policy is unaware of eco-efficiency aspects.

D. Profit and Eco-Efficiency Maximization

Now our goal is to show how the business-driven IT self-management process is able to simultaneously improve the PaaSP's profit and eco-efficiency. We compare both IT-level policies with the proposed *Profit and Eco-efficient (PEE)* policy. The latter looks for the best deployment options (local or external) in terms of fulfilling both BLOs, by considering profit and eco-efficiency in a synergistic way. We show the results depending on the provider's interests, specified by weights of importance. Moreover, we experimented with both workloads used in the previous use cases. E_o is equal to 32172.98 and 9405.10 (±0.61) KWh for the more and less intensive workload (from the use case I and II), respectively.

Table III shows the final BLOs fulfillment, *Ful(PMax, EEMax)*, when different policies are used. It is worth noting that all the results depend highly on the workload received. Nevertheless, in all cases the *PEE* policy achieves better BLOs. It solves the significant tradeoff in their consecution, which is demonstrated by the fact that when one of them is maximized the other one is diminished greatly.

On the one hand, if the private Cloud capacity is not enough for supporting the provider's workload, the IT-level policies outsource the rest to Amazon EC2 since it is three times cheaper than Rackspace. However, the energy consumed by Amazon's servers is higher than Rackspace ones. For these reasons, both profit and eco-efficiency can be greatly improved by the *PEE* policy which, based on the weights of importance, balances both costs and energy savings. On the other hand, if the workload is not heavy, the IT-level policies consolidate services in physical hosts and power off underused ones. They achieve a good enough eco-efficiency since nodes powered by renewable energy are used unconsciously. Contrariwise, the *PEE* policy only uses 'green' nodes as far as it can.

In both cases, the profit can be maximized because IT-level policies are unaware of SLA violation penalties. In this sense, the *PEE* policy outsources services, energy issues apart, when the price of this operation is less than the costs and losses to be paid if they run locally.

Summarizing, this experiment has demonstrated how: (1) business-driven management policies consider successfully the

978-1-4673-4510-1/12/$31.00 ©2012 IEEE

Wkld	Policy	*w(PMax)*	*w(EEMax)*	*Profit* (±*10.69*)	*EcoEff*	*Ful(PMax)*	*Ful(EEMax)*	*Ful(PMax, EEMax)*
Use case I	IT-*	-	-	23849.45	0.13	-	-	-
	PEE	1	0	43797.60	0.07	87.9%	-40.2%	87.9%
		0.75	0.25	43124.75	0.09	80.8%	-35.4%	51.8%
		0.5	0.5	28108.25	0.12	17.9%	-9.4%	4.2%
		0.25	0.75	7801.65	0.22	-67.3%	74%	38.7%
		0	1					74%
Use case II	IT-*	-	-	6088.35	0.50	-	-	-
	PEE	1	0	10318.15	0.28	69.7%	-45.5%	69.7%
		0.75	0.25	10047.40	0.32	65.2%	-36.4%	39.8%
		0.5	0.5	9255.50	0.45	52.2%	-18%	17.1%
		0.25	0.75	342.17	1.02	-94.4%	101.6%	52.6%
		0	1					101.6%

TABLE III
PROFIT AND ECOLOGICAL EFFICIENCY MAXIMIZATION OF DIFFERENT MANAGEMENT POLICIES.

weights of importance of each BLO, thus solving tradeoffs among the achievement of them; and (2) the provider is able to efficiently meet its governing BLOs by relying on the business-driven IT self-management optimization loop.

VI. CONCLUSIONS

In this paper, we have presented a business-driven IT self-management optimization loop suitable to be embedded in Cloud providers driven by Business-Level Objectives (BLOs). It is mainly composed of (1) a Business-Driven IT Management (BDIM) model aimed to assess the impacts of IT-level events on business results, and (2) a policy management framework which, based on the outcomes of such model, determines the most appropriate group of IT management actions from the business point of view, and not only considering IT-level metrics as has been done so far. Moreover, we have demonstrated that a Cloud hosting PaaS provider is able to efficiently improve its BLOs by means of embodying such business-driven IT self-management optimization loop. We have presented three significant and disparate use cases of such provider. The results obtained from the experimentation evidence the success of the management approach presented in aligning business objectives with IT management decisions.

As future work, firstly we plan to extend the suggested BDIM model by incorporating a two-way feedback among its layers, which will be very useful for cause-effect relationships. Secondly, we plan to use heuristics and/or greedy algorithms to deal with the quadratic time complexity of the general model. Thirdly, we aim to deal with typical IT events and risks during the operation of Cloud providers, such as SLA violations. Lastly, we will incorporate other factors, e.g. carbon emissions, for assessing providers eco-efficiency.

ACKNOWLEDGMENT

This work is supported by the Ministry of Science and Technology of Spain and the European Union (FEDER funds) under contract TIN2007-60625, by the Generalitat de Catalunya under contract 2009-SGR-980, and by the European Commission under FP7-ICT-2009-5 contract 257115 (OPTIMIS).

REFERENCES

[1] I. Abi et al. A business driven management framework for utility computing environments. In *9th IFIP/IEEE International Symposium on Integrated Network Management*, pages 16–19, Nice, France, 2005.

[2] S. Aiber, D. Gilat, A. Landau, et al. Autonomic self-optimization according to business objectives. In *1st International Conference on Autonomic Computing*, pages 206–213, New York, USA, 2004.

[3] Amazon EC2. Website, 2011. *aws.amazon.com/ec2*.

[4] M. Armbrust, A. Fox, et al. Above the clouds: A berkeley view of cloud computing. *EECS Department, University of California, Berkeley, Tech. Rep. UCB/EECS-2009-28*, 2009.

[5] C. Belady, A. Rawson, et al. Green Grid Data Center Power Efficiency Metrics: PUE and DCIE. *Consortium Green Grid*, 2008.

[6] R. Buyya et al. Market-oriented cloud computing: Vision, hype, and reality for delivering it services as computing utilities. In *10th IEEE International Conference on High Performance Computing and Communications*, pages 5–13, Dalian, China, 2008.

[7] EMOTIVE Cloud. Website, 2011. *www.emotivecloud.net*.

[8] I. Goiri, F. Julià, R. Nou, et al. Energy-aware Scheduling in Virtualized Datacenters. In *12th IEEE International Conference on Cluster Computing*, pages 58–67, Heraklion, Crete, Greece, 2010.

[9] S. Greenberg et al. Best practices for data centers: Lessons learned from benchmarking 22 data centers. In *ACEEE Summer Study on Energy Efficiency*, volume 3, pages 76–87, Pacific Grove, CA, USA, 2006.

[10] F. Julià, J. Roldán, R. Nou, J. O. Fitó, A. Vaqué, I. Goiri, and J. L. Berral. EEFSim: Energy Efficency Simulator. Technical report, Technical University of Catalonia (UPC), 2010. UPC-DAC-RR-2010-19.

[11] L. Liu, H. Wang, X. Liu, et al. GreenCloud: a New Architecture for Green Data Center. In *6th International Conference on Autonomic Computing and Communications*, pages 29–38, Barcelona, Spain, 2009.

[12] V. Machiraju, C. Bartolini, and F. Casati. Technologies for business-driven IT management. *Extending Web Services Technologies: the Use of Multi-Agent Approaches, Kluwer Academic Publishers*, 2004.

[13] F. Marques et al. Business-oriented capacity planning of IT infrastructure to handle load surges. In *10th IEEE/IFIP Network Operations and Management Symposium*, pages 1–4, Vancouver, Canada, 2006.

[14] D. Menascé, V. Almeida, L. Dowdy, and L. Dowdy. *Performance by design: computer capacity planning by example*. Prentice Hall, 2004.

[15] A. Moura, J. Sauve, and C. Bartolini. Research challenges of Business-Driven IT management. In *2nd IEEE/IFIP International Workshop on Business-Driven IT Management*, pages 19–28, Munich, Germany, 2007.

[16] Rackspace Hosting. Website, 2011. *www.rackspace.com*.

[17] Sauvé et al. An Introductory Overview and Survey of Business-driven IT Management. In *1st IEEE/IFIP International Workshop on Business-Driven IT Management*, pages 1–10, Vancouver, Canada, 2006.

[18] J. Sauvé et al. Business-driven decision support for change management: planning and scheduling of changes. In *17th IFIP/IEEE International Workshop on Distributed Systems: Operations and Management*, pages 173–184, Dublin, Ireland, 2006. Springer.

[19] M. Sedaghat, F. Hernández, and E. Elmroth. Unifying cloud management: Towards overall governance of business level objectives. In *2011 11th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing*, pages 591–597, 2011.

[20] SPECpower_ssj2008. Website, 2008. *www.spec.org/power_ssj2008/results/res2009q1/power_ssj2008-20081212-00104.html*.

[21] SPECweb2009 E-commerce. Website, 2011. *www.spec.org/web2009/docs/design/EcommerceDesign.html*.

[22] S. Srikantaiah, A. Kansal, and F. Zhao. Energy aware consolidation for cloud computing. In *Proceedings of the 2008 conference on Power aware computing and systems*, pages 10–10, Berkeley, CA, USA, 2008.

978-1-4673-4510-1/12/$31.00 ©2012 IEEE