Sustainable Expert Virtual Machine Migration in Dynamic Clouds

Doraid Seddiki[a], Sebastián García Galán[a,*], J. Enrique Muñoz Expósito[a], Manuel Valverde Ibañez[b], Tomasz Marciniak[c], Rocío J.- Pérez de Prado[a]

[a] Telecommunications Engineering Department, University of Jaén. Spain.
[b] Electrical Engineering Department, Univertsity of Jaén. Spain.
[c] Institute of Telecommunications and Computers Sciences. University of Science and Technology in Bydgoszcz. Poland.
*Corresponding Author.Tel +34 953 648556. E-mail: sgalan@ujaen.es.

***Abstract*— Operation on demand flexibility in cloud computing services has resulted in great popularity and wide adoption. These services integrate thousands of computers, storage and communication networks, which implies a high consumption of electrical energy. Therefore, renewable energy-based cloud data centers are replacing traditional energy power grids. In this regard, the workload could be transferred to different nodes among different cloud data centers geographically distributed regarding renewable energy availability. In this sense, this paper presents a framework based on Cloudsim for virtual machine migrations among cloud data centers regarding sustainability optimization. Moreover, an approach for migrations among datacenters based on an expert system has been tested in several scenarios with renewable energy dynamically available. Experimental results show the improvements of the proposed framework and how expert systems can take advantage of renewable energy availability in terms of sustainability while preserving the QoS in terms of execution time.**

Keywords—Cloud computing, migration of virtual machines, scheduling, follows the renewable, expert systems.

I. INTRODUCTION

Currently, digital activities such as social media activity, search, file sharing and streaming services are creating a large amount of data. Every bit of data created needs to be processed, stored and transmitted, increasing energy costs and environmental impact in the form of polluting emissions [1]. In this context, Cloud Data Centers (CDC) are increasingly being used by Information Technology (IT) Service Providers such as Google, Amazon and Microsoft [2]. Thus, the CDC struggles to provide all the required processing capabilities and an efficient infrastructure to store a large amount of data, so that different service level agreements (SLAs) require computing and storage facilities to be replicated redundantly to provide minimal fault tolerance and service delays. As a consequence of the abovementioned idea, IT service providers run data centers 24/7 with thousands of servers and storage devices and network services to provide 99.99% availability of cloud services [3]. Therefore, the size and number of data centers has increased exponentially in parallel with the growth in the number of applications and users of the cloud so that the CDC is estimated to consume more than 2.5% of the world's electricity, with a global economic impact of $30 trillion annually [4]. In addition, it is also estimated that data centers have been responsible for the emission of 2% of polluting emissions all over the world in recent decades and that it has increased today [2].

According to these figures, the application of innovative elements and disruptive measures in CDC to achieve greater efficiency in the use of energy and reduce polluting emissions needs to be addressed. First, modular processing centers or MDCs are strategically located to allow the use of renewable energy, and then dynamic load migration techniques allow the execution of the load in that CDC that make up the MDC based on their possibility of consuming renewable energies or, as they are more commonly known, *follow the renewables* [2]. To be precise, this is the case for certain IT service providers, such as Google and Facebook, which are betting on migration techniques that allow the use of renewable energies in strategically geographically distributed data centers and have coined the term sustainable CDC for those CDCs that allow the application of strategies and technologies based on the reduction of energy consumption and energy reduction [5].

Schedulers are the software components aimed at task allocation to the different nodes of the CDC and MDC under the consideration of working conditions. In fact, Schedulers interacts with other elements regarding local resource management systems and network maintenance systems to carry out their function. Specifically, MDC schedulers, also called intercloud schedulers, are responsible for coordinating CDC schedulers to achieve efficient global scheduling of the entire cloud [2]. The interaction and coordination of these schedulers is critical for the efficient execution of the tasks. To be precise, many efforts need to be devoted to designing scheduling strategies with the capability of working with the inherent uncertainty and dynamism of cloud networks to reduce the energy consumption of the MDC and perform the processing in those nodes or CDCs where a consumption of energy from renewable sources is allowed. Thus, new planning proposals are being sought capable of tolerating and managing the uncertainty in the state of the network so that sustainable processing can be offered based on an efficient description of the real conditions of the cloud [6, 7, 8, 9]. In this respect, many static and dynamic scheduling heuristics have been proposed to schedul a set of virtual machines and cloudlets on datacenters. Braun et al. [10] compare a wide range of static scheduling strategies extensively used in heterogeneous computing environments. FRBSs have proven a high accuracy in their decisions on the basis of the features describing the controlled system and their acquired knowledge. Many applications use fuzzy rules-Based systems (FRBSs) in diverse areas, such as the modeling of traffic flow [11], connection admission in ATM networks [12], speech/music discrimination [13] and scheduling [14]. These systems are proposed in this work for improving the scheduling process in cloud computing datacenters.

Regarding these ideas, Cloudsim is a framework for modeling and simulating cloud computing infrastructures and services [15], which is widely used all over the world to simulate and analyze the behavior of different scheduling policies, including virtual machine (VM) migration. Recently, this platform has been extended with the appropriate mechanisms to simulate VM migration among data centers and as a consequence, among geographically distributed cloud computing infrastructures, which is the scenario where different CDCs can be supplied with different amounts of renewable energy [16].

This work takes advantage of these Cloudsims' new capability of performing simulations scenarios and follows the renewable policies regarding VM's migrations among geographically distributed data centers. Therefore, this work proposes the use of an expert system for following the renewable strategy for the improvement of intercloud schedulers. Specifically, it focuses on the available renewable energy for the design of a meta-scheduler, which makes decisions on the virtual machine's migration among data centers, and whose objective is to select, at each stage of the scheduling, the most suitable CDC to process the different computational cloudlets, trying to optimize the use of renewable energies and the energy consumption by performing the necessary load migration operations [2]. Therefore, the main contribution of this work is focused on taking advantage of the new capabilities with which Cloudsim has been provided, and the proposal of an expert "follows the renewable" approach for VM migration among data centers based on an FRBS, which considers sustainability to increase the use of MDCs supplied with renewable energy to test the aforementioned capabilities.

The remainder of the paper is organized as follows. Section 2 presents a background regarding the different VM placement algorithms and the functionalities of Cloudsim. The proposed FRBS for intelligent VM migration among data centers is introduced in Section 3. Section 4 focuses on the experimental results of the proposed algorithm and provides comparisons with traditional schemas. Finally, Section 5 concludes the paper.

II. BACKGROUND

In this day and age, cloud computing is well-known to have emerged as the leading technology for delivering computational services in terms of reliability, security, fault tolerance, and scalability by providing software, infrastructure, or platform as services (SaaS, IaaS, PaaS). In fact, all these services, provided by the cloud computing paradigm, may be offered in private data centers through private clouds, may be commercially offered for clients through public clouds, or even it is possible that both public and private clouds may be used in a combined way through hybrid clouds. In this respect, virtualization is playing an increasing role as the key technology for the efficient operation of cloud data centers. In this context, virtual machine placement is the process of selecting the most suitable physical machine for a given virtual machine. Therefore, a VM placement algorithm aims at determining the best VM for mapping physical machines, regardless of whether it is an initial VM placement or a VM migration for placement reoptimization. Taking into account these ideas, many efforts have been made to design efficient virtual machine placement algorithms, which can be broken down into several taxonomies:

a) *Depending on the goal of placement:*

- *Power-based approach:* This approach aims to obtain VM-PM mapping, which results in a system that is energy-efficient with the utmost resource utilization. Examples can be found in [17].
- *QoS-Based approach:* In this approach, VM-PM mapping is obtained to ensure maximal fulfillment of quality-of-service requirements. Some examples can be found in [18].

b) *Depending on the type of principal approach used to attain a desirable VM-PM mapping,* the most interesting principal approaches are presented as follows:

- *Constraint programming:* In this case, instead of mathematical approaches, constraint programming is aimed at solveing the complex combinatorial problem of optimal VM placement through the use of a set of constraints that can easily be extended further to involve more aspects. Examples can be found in [19].
- *Bin packing:* To model this problem as a resource allocation algorithm, every single VM has to be tightly packed in the minimum number of bins, and each VM is considered a Physical Machine. Examples can be found in [20].
- *Stochastic integer programming:* This is a mathematical optimization methodology for prediction regarding the uncertainty of future demands. Hence, it makes use of estimation models using probability distributions of the concerned data, where the future demand of a VM or an application is unknown, and therefore, some VM placement techniques use this approach to predict suitable VM-PM mapping. Some examples can be found in [21, 22].
- *Soft computing:* Soft computing [23], addresses approximate models and provides solutions to complex real-life problems. Soft computing is tolerant of imprecision and uncertainty and is based on techniques such as fuzzy logic, evolutionary computation, artificial neural networks, approximate reasoning and its hybridations [22, 24, 25]. It is important to point out that the methodology by authors in this paper fit flawlessly in this category given the use of fuzzy logic.

In this context, where timely, repeatable, and controllable methodologies for performance evaluation of new cloud applications and policies before their actual development are needed, the need for powerful simulation tools with the appropriate simulation capabilities to simulate real scenarios is remarkable. Cloudsim is a framework for modeling and simulating cloud computing infrastructures and services that answer this need. Furthermore, the Cloudsim goal is to provide a generalized and extensible simulation framework that enables modeling, simulation, and experimentation of emerging Cloud computing infrastructures and application services, allowing its users to focus on specific system design issues that they want to investigate, without getting concerned about the low-level details related to Cloud-based infrastructures and services.

Among the main Cloudsim functionalities, followings are worth mentioning [16, 26].

- Support for modeling and simulation of large-scale Cloud computing data centers.
- Support for modeling and simulation of virtualized server hosts, with customizable policies for provisioning host resources to virtual machines.
- Support for modeling and simulation of application containers.
- Support for modeling and simulation of energy-aware computational resources.
- Support for modeling and simulation of data center network topologies and message-passing applications.
- Support for modeling and simulation of federated clouds.
- Support for dynamic insertion of simulation elements, stop and resume simulation.
- Support for user-defined policies for allocation of hosts to virtual machines and policies for allocation of host resources to virtual machines.

To provide these functionalities, Cloudsim presents the architecture depicted in Fig. 1. [27]. The layered architecture of Cloudsim consists of the simulation engine, cloud services and source code, so Cloudsim is an extensible simulation toolkit that enables modeling and simulation of cloud systems and an application provisioning environment.

As can be observed in the Cloudsim architecture, there is room for improvement in terms of VM migration among different data centers.

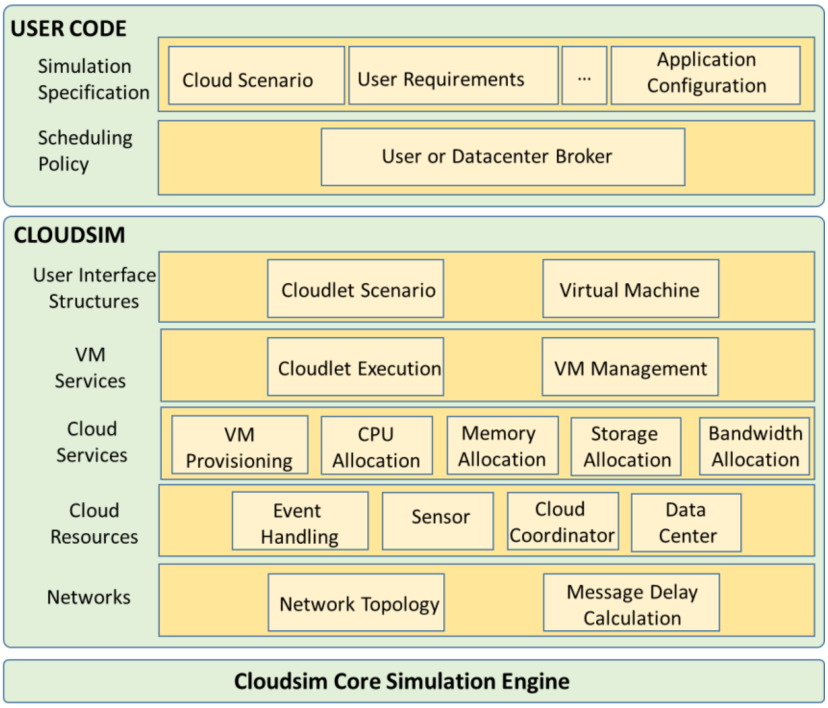

Fig. 1. Cloudsim Architecture.

Previously, Cloudsim could only perform “Intra” Cloud Data Center VM migrations, which means that if a CDC is powered to some extent with a renewable source, the optimization related to the VM migration policies will be applied to all kinds of energies present in the data center and not particularized to the renewable ones. In [16], the authors implemented some changes in Cloudsim so that it can provide VM migrations “inter” data centers, and, to test this new functionality, a sustainability-based algorithm for VM migration inter data centers was presented as well. To provide Cloudsim with the aforementioned functionality, several entities were introduced. These can be divided into two main categories: conceptual entities and simulation entities.

a) Conceptual entities: These are the new logical entities that model the structure and the behavior of

- Renewable Cloud Data Center (RCDC): a CDC that is powered by a real-time hybrid power supply composed of traditional and renewable sources of energy. This new entity can host a new type of host entity that we named “Renewable Power Hosts”
- Renewable Power Host: This entity extends the PowerHostUtilizationHistory and adds the necessary structure and behavior to deal with renewable state and power consumption, among other functionalities.

- Meta-Scheduler: This is an entity that addresses the dispatching of the workload among the different Geo-distributed RCDCs and can implement different strategies of VM migrations. Since it extends the PowerDataCenter entity of Cloudsim, we reuse some migration strategies in this work to compare against our proposed algorithm and test and validate our new simulator capacities.

b) Simulation entities: These are "RUNTIME" related entities that, altough they are not included in the core of entities that extend the Cloudsim functionality, enable the Cloudsim to create the new type of renewable infrastructure, generate the workload, apply the different migration strategies to the meta-scheduler, and finally, calculate the statistical results.

Both conceptual and simulation entities are presented in Fig.2.

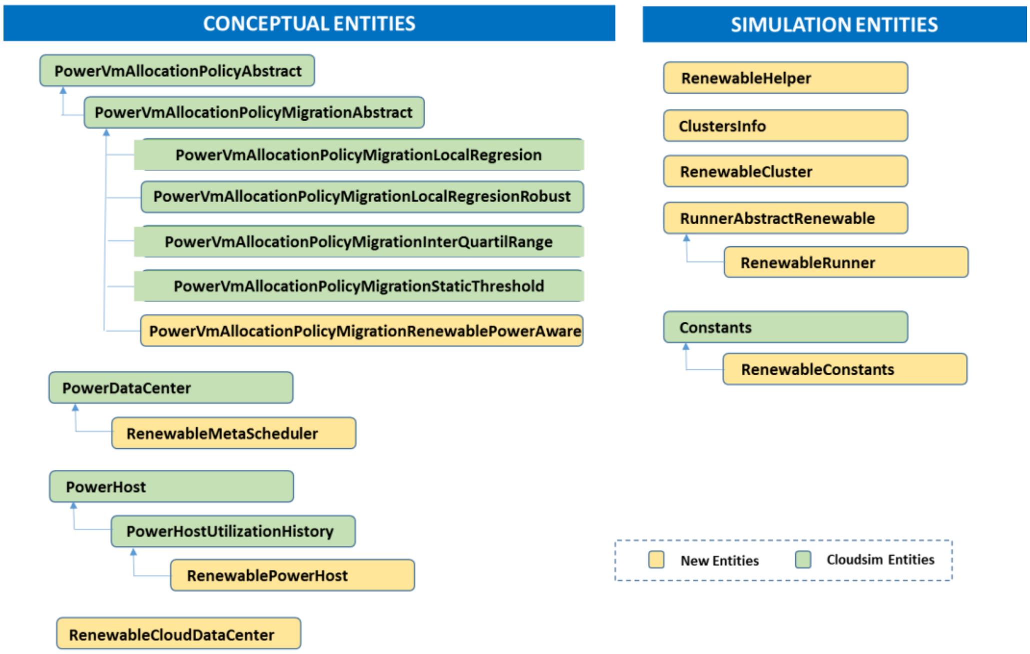

Fig.2 Cloudsim Conceptual and Simulation entities.

It is worth mentioning that these new entities have been addressed to achieve a significant improvement in the overall power consumption across several hybrid energy CDCs powered with a mix of traditional and renewable energy sources by means of VM migrations among data centers. To test these new functionalities, the authors proposed in [16] an algorithm that works as a meta-scheduler devoted to the VM migration bearing in mind renewable energy available in every single data center. This is based upon a Queued-based family of techniques and works as it has already been mentioned, in a specific context where there are several CDCs that are supplied by different percentages of renewable energy, which is available every time. Once this meta-scheduler has received the workload, it requests all CDCs to know the available renewable energy in every single one of them. It is important to point out that the available renewable energy will change dynamically, which is why this lookup is needed every time. Now that the information of the renewable energy is available for the meta-scheduler, it will order the CDCs by this availability to create an ordered list that reflects the preferences for the dispatching of the workloads to the CDCs, from them to the hosts and finally from the hosts to the VMs.

Again, and with the previous information, the meta-scheduler will try one by one every CDC (starting with the one with the largest availability of renewable energy and drilling down the one with the least renewable energy) to assign the workload to them. Then, every single CDC will check its capacity for hosting the workload asking its different hosts to simulate the creation of the corresponding VMs and check if these hosts will remain optimally used and not to switch to an “Overused” State.

Depending on the capacity of the CDC, it will keep the workload for itself, or it will reject the request of the meta-scheduler. If the CDC has enough capacity, it will compute the workload and inform the meta-scheduler after finishing, and the meta-scheduler will inform the broker. On the other hand, if the CDC will not have enough capacity to deal with the workload, it will answer the meta-scheduler of the VM creation failure. These will restart the process again, but this time with a different CDC. More precisely, the next one in the ordered list of CDCs by renewable energy until finding the CDC who could take in charge the workload. If by any means the workload could not find a VM to process it, which means that a VM creation failure will occur, the simulation will stop.

Fig. 3 depicts this algorithm by means of a sequence diagram that includes the interactions with the different CDC through its brokers.

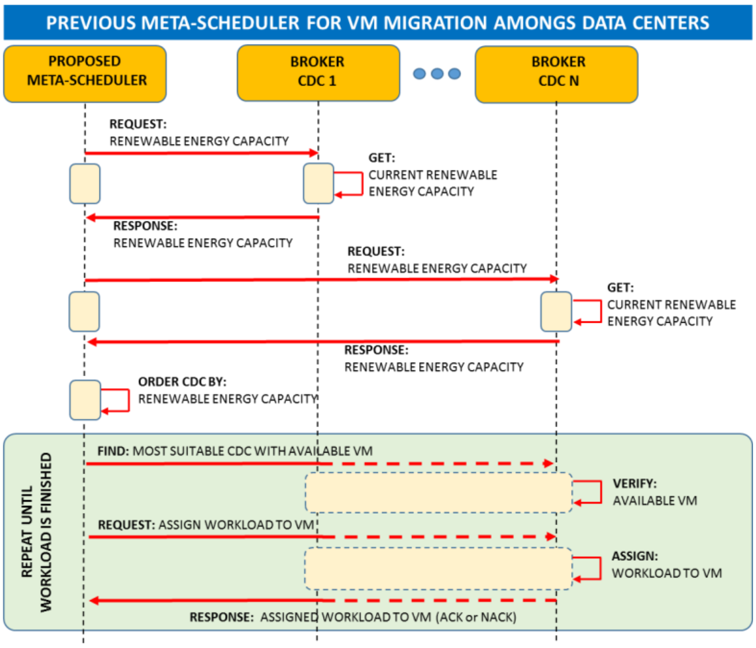

Fig.3. Sequence diagram of the previous algorithm considering VM migration among cloud data centers.

As an important contribution, in this work, the authors propose a new meta-scheduler based on a fuzzy rules-based system for VM migrations among data centers.

III. Expert System for VM Migrations Among Data Centers Regarding Sustainability

To check the performance of the new capabilities with Cloudsim in terms of VM migrations, in this paper, a new approach using a fuzzy rules-based system (FRBS) as the central algorithm for the meta-scheduler is proposed. The organization of the fuzzy rule-based meta-scheduler has been inspired by the typical modular structure of fuzzy logic systems (FLSs) [23], so that the fuzzy rule-based meta-scheduler general structure consists of four well-known systems for FRBSs: fuzzification interface, knowledge base, inference engine and defuzzification interface. The specifications for these systems have been extensively studied in the implementation of fuzzy systems [28]. As stated in [23], the specification of a set of rules or knowledge representation and the determination of some concepts associated to the reasoning strategy (i.e., features) describe an FRBS. According to Mamdani [29], a given fuzzy rule R_i consists of two differentiated parts, namely, antecedent and consequent parts, related to fuzzy concepts. Hence, on the one hand, rule activation conditions are reflected in the antecedent part of the rule. On the other hand, the consequent part depicts the scheduler output or response. In this way, a rule within this approach is represented by the following expression:

$$R_i: IF\ w_1\ is\ A_{1n} \frac{and}{or} \ldots w_m is\ A_{mn}\ THEN\ y\ is\ B_n \tag{1}$$

where w_m represents a system feature, y depicts the output variable and A_{mn} and B_n correspond to the fuzzy sets associated with feature m and output, respectively. Hence, a high-quality RB (rule base) for the fuzzy meta-scheduler must consist of a set of rules able to efficiently face a wide number of cloud conditions given by the dynamic inputs or cloud system features. Once the general structure of rules within our fuzzy meta-scheduler has been presented, the system features will be introduced. Features are those variables describing the state of the cloud environment and then constituting the inputs for the fuzzy system. In this work, the inputs for the fuzzy system are focused both on achieving a good performance in time and good performance in the use of renewable energy. Features are summarized in Table 1.

Table 1. Features for Cloud System Description. Input Variables.

Feature	**Description**
Cloud Data Center Renewable Availability (CDC-RA)	Renewable energy percentage supplied to the CDC

Host Computational Capacity (HCC)	Maximum computational capacity in MIPS
Host Computational Availability (HCA)	Remaining computational capacity in the MIPS after holding other VMs
VM Maximum Computational Needs (VM-MCN)	Maximum needs of the Virtual Machine in MIPS
VM Current Computational Needs (VM-CCN)	Remaining computational needs in MIPS for the VM

Despite the simplicity and efficiency of triangle-shaped membership functions, Gaussian ones have been selected since they are able to provide a whole characterization of the universe of discourse of the variables. According to this idea, every single feature is modeled regarding the combination of three membership functions (low, medium and high) with the following mathematical form:

$$g_i^w(x) = \frac{1}{\sigma_i^{(w)}\sqrt{2\pi}} exp\left\{\frac{-\left(x-\mu_i^{(w)}\right)^2}{2{\sigma_i^{(w)}}^2}\right\}\{x \in \mathbb{R}^+ | x \leq 1\} \tag{2}$$

where μ_i^w (mean) and $\sigma_i^{(w)}$ (deviation) describe each fuzzy set for every feature w and rule Ri. Hence, regarding a given rule Ri as a combination of feature description both in its antecedent and consequent part, it may be depicted by the following expression:

$$R_i = \left\{g_i^{(1)}(x), g_i^{(2)}(x), \ldots g_i^{(N_F)}(x), \Omega_i\right\} \tag{3}$$

where N_F represents the number of features considered, and the recommendation vector Ω_i indicates the consequent part of the rule. For simplicity, in this work, different weights for the rules are not considered, and as a consequence, every single rule within the RB has the same influence. Additionally, the suitability of a given host for a VM that has to migrate is the output of the FRBS, and it is represented by five different fuzzy sets, namely very not suitable, not suitable, suitable, very suitable and extremely suitable. Normalized values for inputs and output are illustrated in Fig. 4.

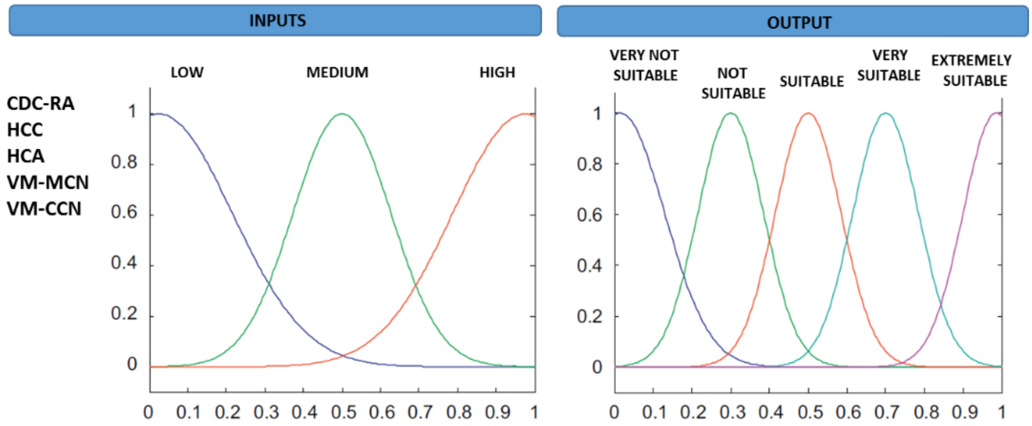

Fig. 4. Inputs and outputs membership functions.

It is worth mentioning that the considered knowledge consists of a set of fuzzy rules that relate the input and output variables of the system by a linguistic representation to obtain a final response for the system. Therefore, the rules of an FRBS constitute the main part of the knowledge of the expert system, and thus, its efficiency is strongly related to its quality. In this work, the rules used were obtained from the experience of authors and are presented in Table 2. To be precise, some of these, rules (R1, R2, R3, and R4) are related to computational issues, while the rest of the rules (R5, R6 and R7) are related to the management of renewable energy. As can be observed, the authors considered a reduced set of simple rules to favor their good understanding and avoided the use of more than two input variables in every single rule to maintain the interpretability of the rules.

Table 2. Rules Base of the Expert System

RULE	CDC-RA	HCC	HCA	VM-MCN	VM-CCN	AND/OR	OUTPUT
1		High		Low		AND	Extremely Suitable
2		Low		High		AND	Very Not suitable
3			High		Low	AND	Extremely Suitable
4			Low		High	AND	Very Not suitable
5	Low						Very Not Suitable
6	Medium						Suitable
7	High						Extremely Suitable

Finally, taking into account the abovementioned ideas, the structure of the meta-scheduler is depicted in Fig. 5.

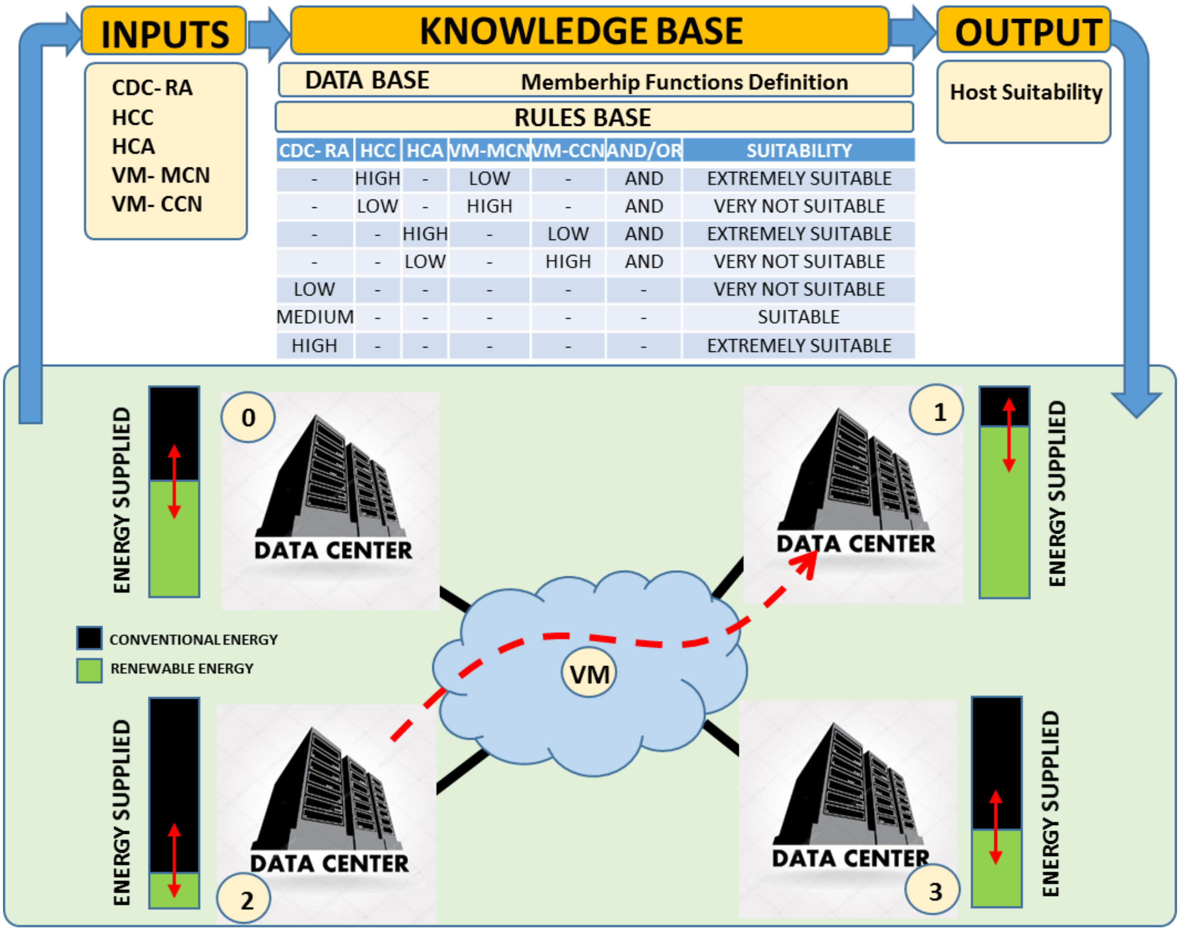

Fig. 5. General Structure of the meta-Schedulers.

IV. EXPERMENTAL RESULTS

Cloudsim is a simulator for cloud computing infrastructures and modeling that provides different server configurations to be used in power-aware simulation. Hence, several real-world computer model can be selected in simulations. The servers that have been considered in this work have the following configuration:

- HP ProLiant ML110 G4 (Xeon 3040, dual core, 1,8 GHz, 4 GB RAM, and 1Gbps of Bandwidth).
- Hp ProLiant MLl110 G5 (Xeon3075, dual core, 2,6 Ghz, 4Gb RAM, and 1Gbps of Bandwidth),

The power consumption data for the different real machines were obtained from the *Standard Performance Evaluation Corporation* (SPEC, https://www.spec.org), which is a nonprofit corporation formed to establish, maintain and endorse standardized benchmarks and tools to evaluate performance and energy efficiency for the newest generation of computing systems. The power consumption associated with a specific CPU utilization, provided by SPEC, ranges from 0% to 100% with a 10% step. If the real utilization is not exactly one of the previous values, then a linear interpolation method that approximates the real value of the power consumed is used.

Additionally, it is important to point out that to obtain the most realistic results, Cloudsim provides real traces of CPU utilization by means of 288 entries reflecting a real utilization of the CPU every 5 minutes, which gives a total of 24 hours of simulation potential data.

Regarding the workload, this work considers cloudlets using the following configuration:

- *Cloudlet Length*: 2500*Simulation time (MI)
- *Cloudlet File Size (Program+data):* 300 (bytes)
- *Cloudlet Output Size:* 300 (bytes)
- *Pes Number:*1

Related to the features of VM, four types of VM have been used to compute the different workloads. Their configurations are as follows:

- *Number of process elements*: 1
- *MIPS*: 2500, 2000, 1000, 500
- *RAM*: 870, 1740, 613
- *Bandwidth*: 100Mbps
- *Size*: 2,5 Gb

To test how the performance of the expert virtual machine migration policy, several comparisons were carried out taking into account the meta-scheduler proposed in [16] when extended capabilities included Cloudsim were checked. This was summarized in this work, and the previous migration policies provided by Cloudsim [30] are as follows:

- Inter Quartil Range (IQR): This technique isused in machine learning algorithms to identify outliers but in this case helps to calculate the median of an ordered dataset (n this case, the dataset is represented by some host utilization data) and tries to determine the difference between the first half and the second half (a measure of spread) using two medians called Q1 and Q2. The first one cuts the dataset in half, and the two others cuts on half the left and right sides. This strategy calculates an upper threshold that identifies the overused hosts
- Local Regression (LR): is a technique used in machine learning algorithms that tries to fit the best line to a dataset of samples that minimizes the least square distances to predict the new behavior for samples

outside the dataset. In this case this VM migration policy tries to predict host overload in different situations.

- Local Regression Robust (LRR): This is a more suitable model than LR in contexts where there are too many outlier samples (the LR Model in this case could give biased results that cannot be interpreted) and where the spread of the data is not linear. It is a technique used in machine learning algorithms that tries to fit a curve to a dataset approximating different subsets of it using a specific parameter of a window size "n" that assigns waits to its data samples and approximate it using two types of regression curves (lines or parabolas). It is used to predict host overutilization.
- Static Threshold: This is a VM allocation policy that uses a static CPU utilization threshold (THR) to detect host overutilization.

Finally, to determine the VMs that will be used in the migration process, a VM selection policy is needed. In this sense, this work uses one of the policies considered in Cloudsim, which is based on RAM utilization and called the *minimum migration time (MMT)* strategy.

A. Simulation scenario description

In this work, three main simulation scenarios have been considered and are described in Table 3. These scenarios have been selected to provide a wide variety of situations regarding different sizes of data centers and different workloads. To be precise, bearing in mind the capability of Cloudsim, which allows up to 800 hosts, three scenarios have been considered (small, medium and large). These scenarios represent different infrastructure sizes, because they allow the conclusive results to be obtained without being exhaustive and without making it difficult to present the obtained results.

Table 3. Different Scenarios Used in Simulations

	Scenario 1	**Scenario 2**	**Scenario 3**
Hosts	265	530	800
VM	350	695	1052
Workload (Cloudlets)	500 - 1,500 - 3,000	1,000 - 2,000 - 5,000	1,500 - 5,000 - 10,000

For each and every scenario, the hosts and the virtual machines were distributed randomly over four different CDC's. The first had no renewable energy source, the second had 33% of renewable energy, the third had 66% and the last had 100%. Nevertheless, in this work, as a relevant contribution, a dynamic change in available renewable energy has been considered. Therefore, the percentage of renewable available energy will rotate around the four cloud data centers every hour, which means that at the end of the simulations (4 hours), every one of them will have switched three times the renewable energy percentage available in it see Fig. 6. Note that this work considered only renewable energy availability regardless of its origin. Therefore, there is no dependency in terms of renewable energy type.

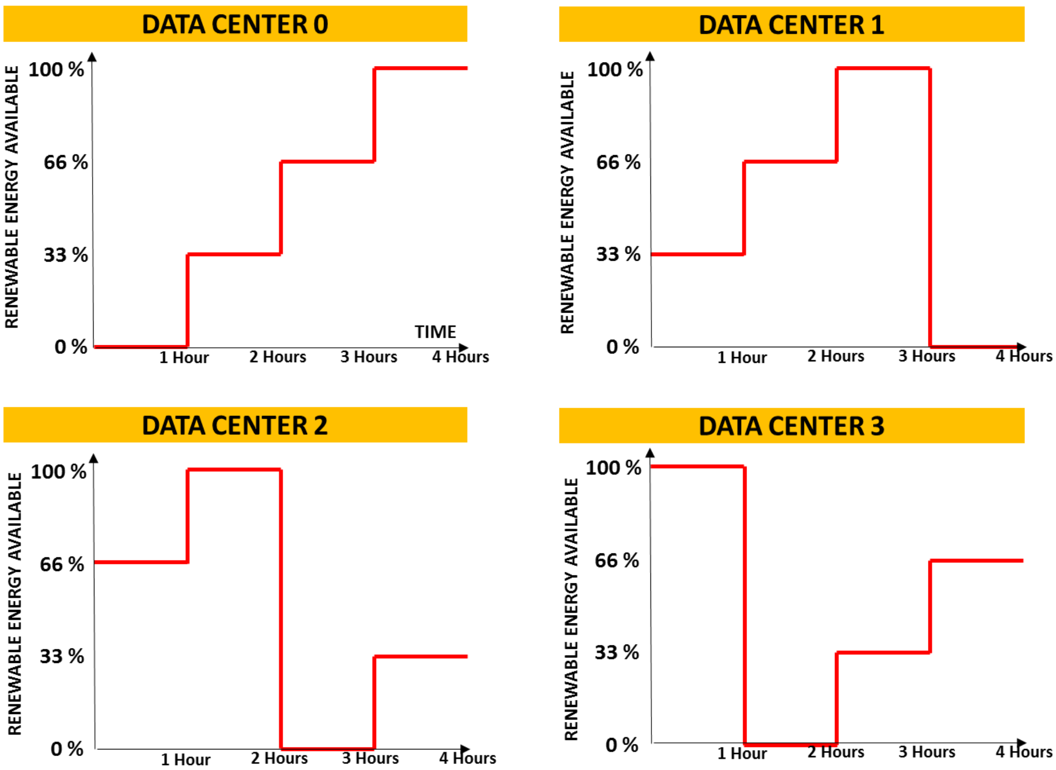

Fig. 6. Dynamic Availability of Renewable Energy in Data Centers.

B. *Experimental Results*

The simulation scenarios described in the previous section were performed to give the results shown in Tables 4, 5 and 6. These indicate the average of 30 experiments, which correspond to a rule-of-thumb threshold based on the Central Limit Theorem where the sample distribution will approach a normal distribution as the sample size increases. In Tables 4, 5 and 6, it can be observed that the algorithms THR, IQR, IR and IRR consistently lead to a usage of 53-54% of renewable energy regardless of the scenarios used. Moreover, the previous algorithm presented in [16] to check Cloudsim's new capabilities has reached an average of 63-64% of renewable energy use (average over the three scenarios), resulting in a clear improvement of a proximately 9-10% regarding the precedent algorithms. It can also be observed that in the context of the conditions related to the dynamic environment, the FRBS-based meta-scheduler proposed in this work manages to reach 63-64% of the usage of renewable energy, which represents a similar behavior to the best results of the best algorithms used in the previous work [16]. Nevertheless, the improvement reached by the proposed algorithm in comparison to the previous algorithm can be observed when the cloud infrastructure sizing and complexity are increasing. This means that the proposed meta-scheduler presents good performance when optimizing the sustainability of cloud systems in terms of the use of renewable energy despite the unoptimized expert knowledge used in its knowledge base. Finally, in relation to the use of renewable energy expressed in Tables 4,5 and 6, a confidence interval of 99% was also included. On the other hand, it is important to point out that the more VM migrations there are among the cloud data centers, the more information must be transmitted through the communications networks, which can imply a deterioration in the user experience. However, regarding the SLA expressed in Tables 4, 5 and 6, this situation is not significant.

Table 4. Obtained results in Scenario 1

SCENARIO 1							
HOSTS: 265, VM: 350		**ALGORITHM**					
Cloudlets	**Performance**	**THR**	**IQR**	**IRR**	**LR**	**Previous**	**FRBS**
500	Renewable used (%)	54,09±0,09	53,89±0,09	54,08±0,11	54,40±0,13	**63,21**±0,11	**63,31**±0,13
	SLA Deg. (%)	< 0,0001	< 0,0001	< 0,0001	< 0,0001	< 0,0001	< 0,0001
1500	Renewable used (%)	54,44±0,09	53,58±0,08	53,80±0,10	53,73±0,14	**63,31**±0,12	**62,75**±0,11
	SLA Deg. (%)	< 0,0001	< 0,0001	< 0,0001	< 0,0001	< 0,0001	< 0,0001
3000	Renewable used (%)	54,29±0,07	54,08±0,11	54,17±0,13	53.62±0,14	**63,91**±0,10	62,68±0,15
	SLA Deg. (%)	< 0,0001	< 0,0001	< 0,0001	< 0,0001	< 0,0001	< 0,0001

Table 5. Obtained results in Scenario 2

SCENARIO 2							
HOSTS: 530, VM: 695		**ALGORITHM**					
Cloudlets	**Performance**	**THR**	**IQR**	**IRR**	**LR**	**Previous**	**FRBS**
1000	Renewable used (%)	53,81±0,07	53,65±0,07	53,59±0,07	53,49±0,10	63,37±0,07	**63,71**±0,07
	SLA Deg. (%)	< 0,0001	< 0,0001	< 0,0001	< 0,0001	< 0,0001	< 0,0001
2000	Renewable used (%)	53,61±0,07	53,66±0,09	53,91±0,08	53,35±0,07	**63,19**±0,10	**63,40**±0,08
	SLA Deg. (%)	< 0,0001	< 0,0001	< 0,0001	< 0,0001	< 0,0001	< 0,0001
5000	Renewable used (%)	53,83±0,09	53,67±0,08	53,24±0,07	53,56±0,09	**63,57**±0,09	**63,65**±0,09
	SLA Deg. (%)	< 0,0001	< 0,0001	< 0,0001	< 0,0001	< 0,0001	< 0,0001

Table 6. Obtained results in Scenario 3

SCENARIO 3								
HOSTS: 800 **VM: 1052**		**ALGORITHM**						
Cloudlets	**Performance**	**THR**	**IQR**	**IRR**	**LR**	**Previous**	**FRBS**
1500	Renewable used (%)	53,78±0,06	53,67±0,05	53,69±0,06	53,62±0,07	63,54±0,07	**64,36**±0,05
	SLA Deg. (%)	< 0,0001	< 0,0001	< 0,0001	< 0,0001	< 0,0001	< 0,0001
5000	Renewable used (%)	53,86±0,06	53,88±0,07	53,75±0,07	53,58±0,06	63,55±0,06	**63,85**±0,07
	SLA Deg. (%)	< 0,0001	< 0,0001	< 0,0001	< 0,0001	< 0,0001	< 0,0001
10000	Renewable used (%)	54,12±0,06	53,90±0,07	53,39±0,07	53,70±0,06	63,56±0,09	**64,28**±0,09
	SLA Deg. (%)	< 0,0001	< 0,0001	< 0,0001	< 0,0001	< 0,0001	< 0,0001

Next, Figures 7, 8, 9 and 10 depict the behavior of the fuzzy rule-based system and the previously proposed algorithm in [16] in the context of the second scenario with 5000 cloudlets for both energy consumption and VM migrations. To be precise, these bar charts in Figures 7 and 9 represent the total and the renewable energy consumed in each and every data center, while Figures 8 and 10 show how both strategies carry out VM migration following the renewable approach.

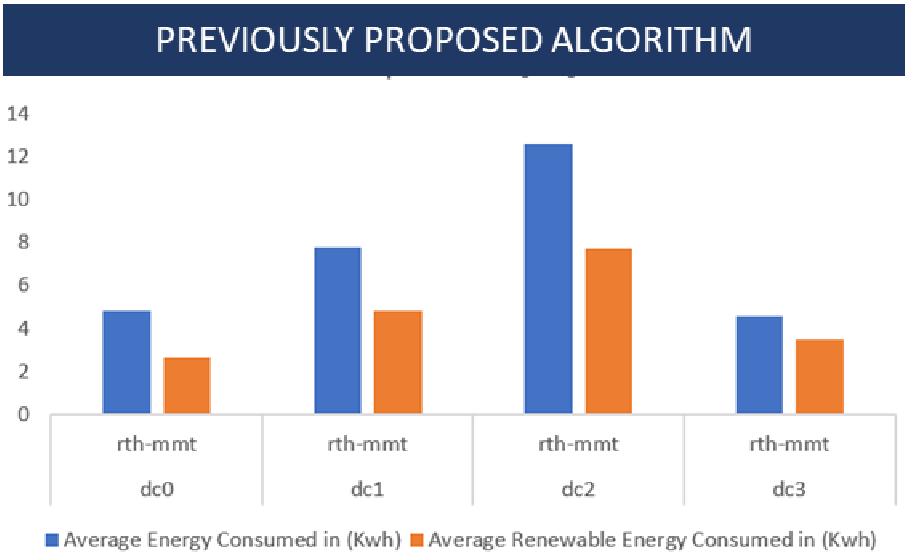

Fig. 7. Previously proposed algorithm’s performance in scenario 2 with 5000 Cloudlets.

Related to Figure 7, it can be observed how most of the consumed energy (total and renewable) concentrates on cloud data centers DC2 and DC1, followed by DC3 and DC0 in this specific order. In this respect, Figure 6 presents the dynamic change behavior of the renewable energy in the different cloud data centers. During the first 2 hours of simulation, the data centers DC2 and DC1 have the largest average percentages of renewable energy available of the four data centers. In addition, the cloudlet load in the network is the largest in the first hour and decreases with time, which makes the most part of the processing of this load in data centers DC2 and DC1 and the remaining load is dealt with in the other two data centers in the third and fourth intervals of simulation.

This behavior can clearly be explained by Figure 8 where the first interval (from 0 to 1 hour) shows that the largest migrations are toward data center DC3, which has 100% of renewable energy available, followed by data center DC2 with 66%, data center DC1 with 33% and finally data center DC0 with the lowest number of VM migrations.

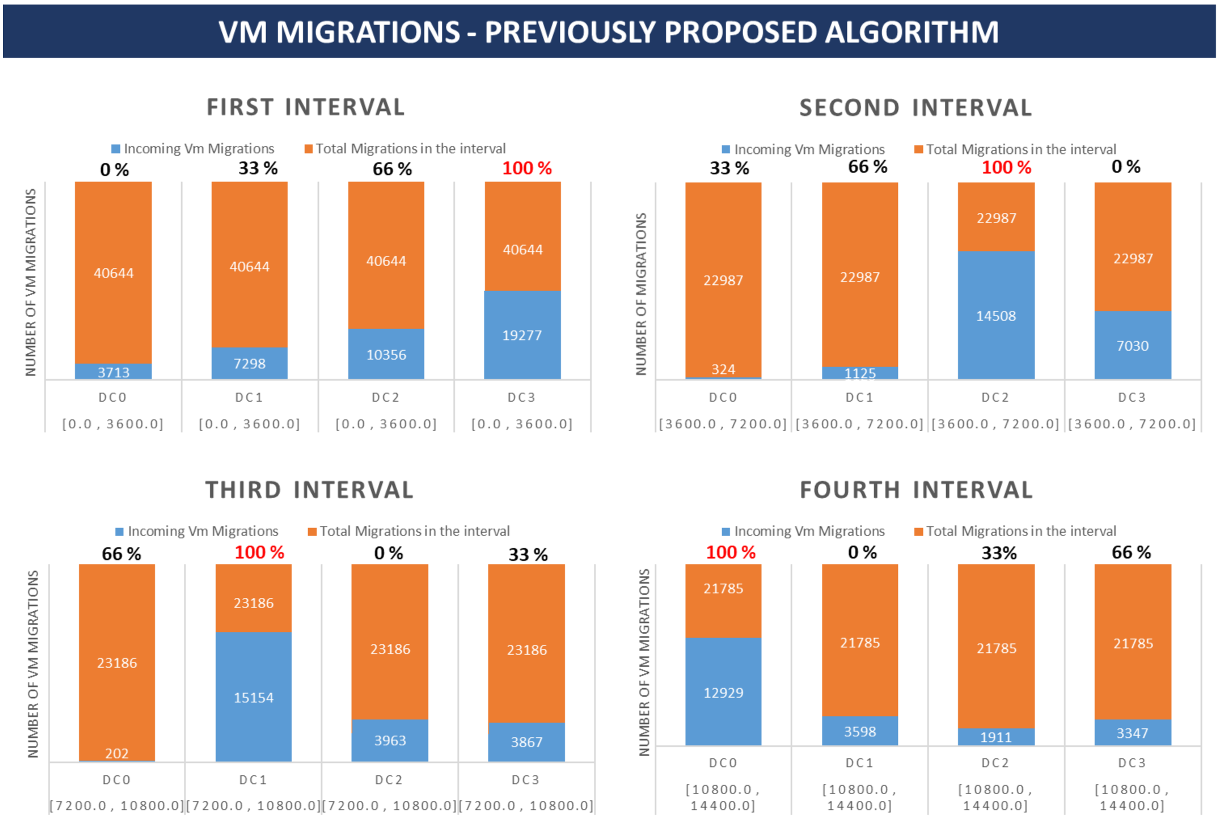

Fig. 8. VM Migrations for the previously proposed algorithm in scenario 2.

During the second interval, a large spike of migrations can be observed, in this case toward data center DC2 (which now has 100% of renewable energy) and a large decrease in the migrations toward the other three datacenters. Related to the third and fourth intervals, the algorithm follows the same pattern as before, so that the data center with the largest renewable energy available receives the largest number of migrations. This is clearly seen through the spike in the data center DC1 in the third interval and the same behavior for DC0 in the last interval. Bearing in mind the aforementioned ideas, it can clearly be concluded that this algorithm follows the renewable approach.

The FRBS-based scheduler proposed in this work, which uses expert knowledge, has reached the same performance as the previously proposed algorithm in [16] averaging 63-64% of the use of renewable energy over the four hours of simulation and over the four data centers. In this sense, Figure 9 represents the consumption of total and renewable energy in the different data centers.

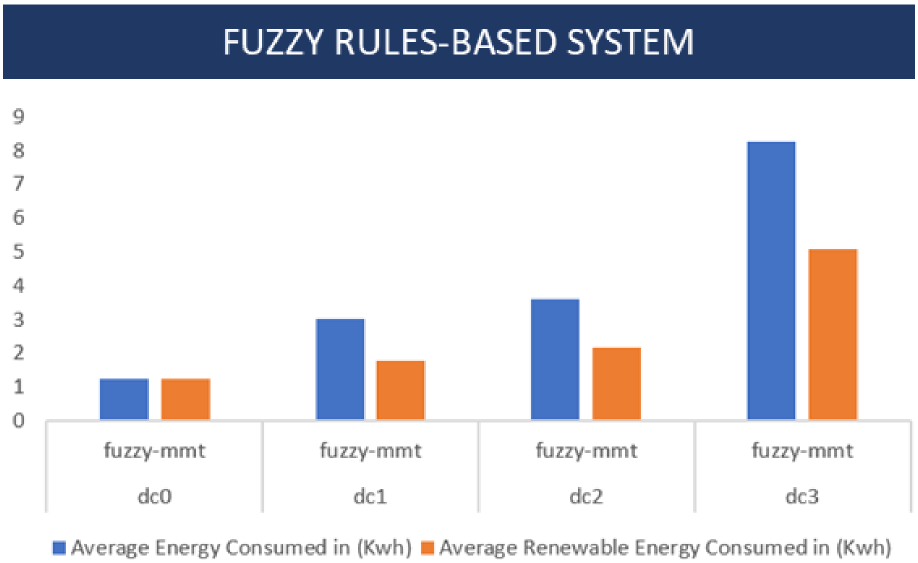

Fig. 9. Behavior of FRBS Algorithm in scenario 2.

In Figure 10, it can be observed how clearly the FRBS Algorithm is "following the renewables". Hence, in the first interval, the data center with the most renewable energy of 100% has virtually all incoming migrations of the virtual machines (52.815) and leaves only 66 for the datacenter DC2, which has 66% of renewable energy available and 30 migrations for the DC1 with 33% of renewable energy available.

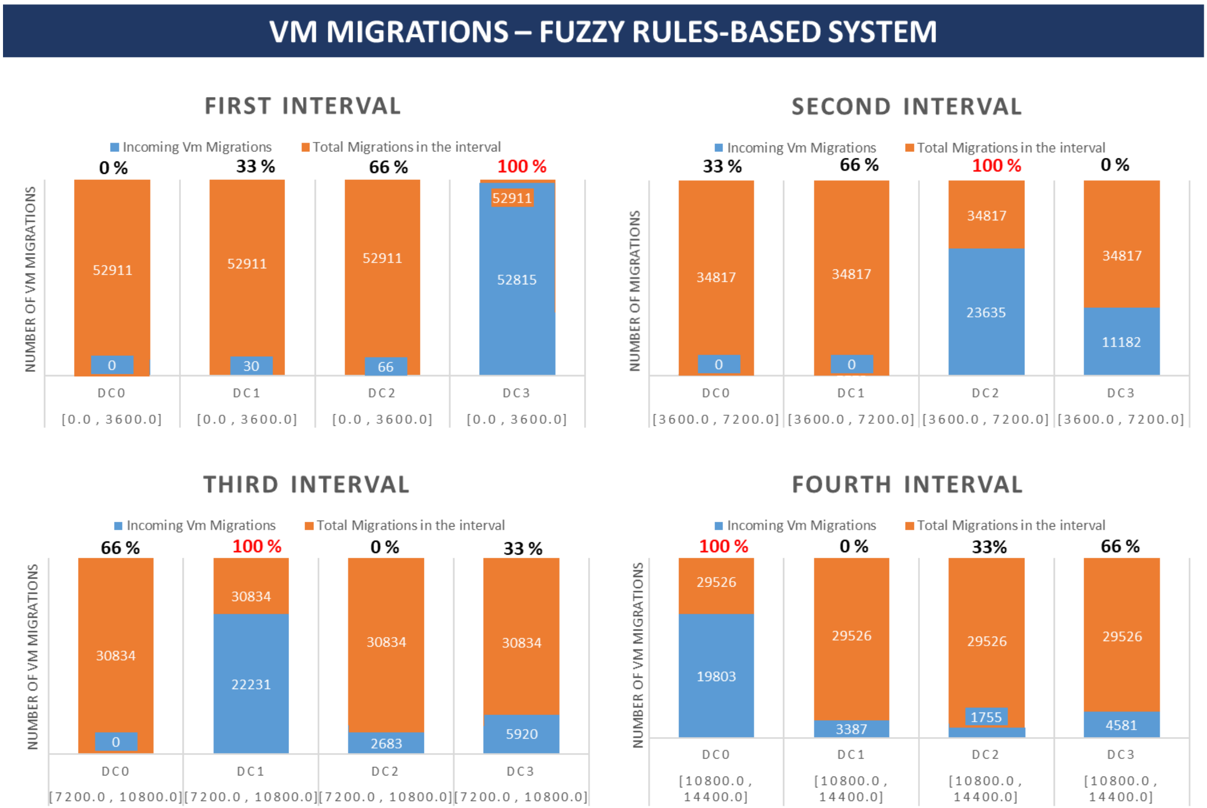

Fig. 10. VM Migrations for FRBS in scenario 2.

During the second interval, data center DC2 has 100% of renewable energy available, so that the number of received migrations have changed from 66 to 23.635. Data center DC3, which has 100% of renewable energy available in the previous interval and now has 0%, has decreased from 52.815 to 11.182. In the third and fourth intervals the trend is the same, recording the largest migrations toward the data centers with the largest percentage of renewable energy available. However, due to the general-purpose nature of the rules that have been used and the conditions in terms of computational resource availability, there are some residual migrations to other data centers. Nevertheless, this behavior can be improved by introducing some knowledge acquisition techniques [31, 32].

V. Conclusions and Future Actions

In this work, a new approach based on a fuzzy rules-based system for VM migration among cloud data centers was introduced. This expert meta-scheduler takes advantage of the Cloudsim's new capabilities regarding the "follow the renewable" strategy, which allows interclouds VM migrations. Taking into account the different experiments that have been carried out, the behavior of the proposed meta-scheduler was tested bearing in mind several simulation scenarios in a dynamic environment in terms of renewable energy availability in cloud data centers. The FRBS approach showed good behavior, improving upon the previously proposed algorithm when the cloud infrastructure sizing and complexity are increasing and obtaining 63-64% of the renewable energy usage across the four cloud data centers even if the FRBS is using a general-purpose set of rules in the knowledge base. In this sense, regarding future actions, the authors are considering the use of knowledge acquisition techniques to optimize the rules base and consequently to improve the behavior of the meta-scheduler in terms of renewable energy utilization and total execution time.

In this respect, it is necessary to indicate that the interpretability of FRBS-based intercloud meta-schedulers, even with the involvement of different knowledge acquisition techniques, represents an important feature in order to understand the internal mechanism that rules the appropriate work, which could result in an improvement in knowledge about the follow the renewable approach.

Acknowledgment

This work has been supported by the Research Project P18-RT-4046, funded by the Andalusian Government in Spain.

References

[1] M. Romero, H. Hasselqvist, G. Svensson. Supercomputers keeping peoplewarm in the winter. Proceedings of the 2nd International Conference on ICT for Sustainability (ICTS), August 24–27, 2014. Stockholm, Sweden, pp. 324–32, 2014.

[2] J. Shuja, A. Gani, S. Shamshirband, R. Wasim Ahmad, K. Bilal. Sustainable Cloud Data Centers: A survey of enabling techniques and technologies. Renewable and Sustainable Energy Reviews, vol. 62, pp. 195-214, 2016. DOI: 10.1016/j.rser.2016.04.034.

[3] A. Ali, Li Lu, Y. Zhu, and J. Yu. An Energy Efficient Algorithm for Virtual. Machine Allocation in Cloud Datacenters. CCIS 626, pp. 61–72, 2016. DOI: 10.1007/978-981-10-2209-8 6

[4] G. Meijer. Cooling energy-hungry data centers. Science; vol. 328(5976), pp. 318–9, 2010.

[5] Google, Google's green data centers. Network pop case study, 2012. URL ⟨https://www.google.com/about/datacenters/efficiency/external⟩.

[6] Muhammad Sohaib Ajmal, Zeshan Iqbal, Farrukh Zeeshan Khan, Muneer Ahmad, Iftikhar Ahmad, Brij B. Gupta. Hybrid ant genetic algorithm for efficient task scheduling in cloud data centers. Computers & Electrical Engineering. Volume 95, 2021, 107419, ISSN 0045-7906, https://doi.org/10.1016/j.compeleceng.2021.107419.

[7] Jingchen Yan, Yifeng Huang, Aditya Gupta, Anubhav Gupta, Cong Liu, Jianbin Li, Long Cheng. Energy-aware systems for real-time job scheduling in cloud data centers: A deep reinforcement learning approach. Computers & Electrical Engineering. Volume 99, 2022, 107688, ISSN 0045-7906, https://doi.org/10.1016/j.compeleceng.2022.107688.

[8] Juan Carlos Salinas-Hilburg, Marina Zapater, José M. Moya, José L. Ayala. Energy-aware task scheduling in data centers using an application signature. Computers & Electrical Engineering. Volume 97, 2022, 107630, ISSN 0045-7906, https://doi.org/10.1016/j.compeleceng.2021.107630.

[9] M. Seddiki, R.P. Prado, J. E. Munoz Exposito, and S. García Galán. Fuzzy rule-based systems for optimizing power consumption in data centers. 5th International Conference on Image Processing & Communications. Bydgoszcz, Poland, 2013.

[10] Braun, T.D., Siegel, H.J., Beck, N., Bölöni, L.L., Maheswaran, M., Reuther, A.I., Freund, R.F.: A comparison of eleven static heuristics for mapping a class of independent tasks onto heterogeneous distributed computing systems. Journal of Parallel and Distributed Computing 61(6), 810–837 (2001)

[11] Quek, C., Pasquier, M., Lim, B.: A novel self-organizing fuzzy rule-based system for modelling traffic flow behaviour. Expert Syst. Appl. 36(10), 12167–12178 (2009)

[12] Cheong, F., Lai, R.: Connection admission control of mpeg streams in atm network using hierarchical fuzzy logic controller. Eng. Appl. Artif. Intell. 22(1), 117–128 (2009)

[13] Munoz-Exposito, J.E., García-Galán, S., Ruiz-Reyes, N., Vera-Candeas, P. Adaptive network-based fuzzy inference system vs. other classification algorithms for warped lpc-based speech/music discrimination. Eng. Appl. Artif. Intell. 20(6), 783–793 (2007)

[14] Prado, R.P., García Galán, S., Yuste, A.J., Muñoz Expósito, J.E., Sánchez Santiago, A.J., Bruque, S.: Evolutionary Fuzzy Scheduler for Grid Computing. In: Cabestany, J., Sandoval, F., Prieto, A., Corchado, J.M. (eds.) IWANN 2009, Part I. LNCS, vol. 5517, pp. 286–293. Springer, Heidelberg (2009

[15] R. N. Calheiros, R. Ranjan, A. Beloglazov, C. A. F. De Rose, and R. Buyya, "CloudSim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms." Software: Practice and Experience (SPE), Volume 41, Number 1, Pages: 23-50, ISSN: 0038-0644, Wiley Press, New York, USA, January, 2011.

[16] D. Seddiki, S. G. Galán, E. M. Expósito, M. V. Ibañez, T. Marciniak and R. J. P. De Prado, "Sustainability-based Framework for Virtual Machines Migration Among Cloud Data Centers," *2021 15th International Conference on Signal Processing and Communication Systems (ICSPCS)*, 2021, pp. 1-8, doi: 10.1109/ICSPCS53099.2021.9660237.

[17] Song, W., Xiao, Z., Chen, Q. and Luo, H., 2014. Adaptive resource provisioning for the cloud using online bin packing. Computers, IEEE Transactions on, 63(11), pp.2647-2660.

[18] Beloglazov, A. and Buyya, R., 2013. Managing overloaded hosts for dynamic consolidation of virtual machines in cloud data centers underquality of service constraints. Parallel and Distributed Systems, IEEE Transactions on, 24(7), pp.1366-1379.

[19] Y. Zhai, G. Zhao, H. Xu, Y. Zhao, J. Liu and X. Fan, "Towards Robust Multi-Tenant Clouds Through Multi-Constrained VM Placement," 2021 IEEE/ACM 29th International Symposium on Quality of Service (IWQOS), 2021, pp. 1-6, doi: 10.1109/IWQOS52092.2021.9521344.

[20] Nurşen Aydın, İbrahim Muter, Ş. İlker Birbil, Multi-objective temporal bin packing problem: An application in cloud computing, Computers & Operations Research, Volume 121, 2020, 104959, ISSN 0305-0548, https://doi.org/10.1016/j.cor.2020.104959.

[21] Chen, M., Zhang, H., Su, Y.Y., Wang, X., Jiang, G. and Yoshihira, K., 2011, May. Effective VM sizing in virtualized data centers.In Integrated Network Management (IM), 2011 IFIP/IEEE International Symposium on (pp. 594-601). IEEE.

[22] C. Pandiselvi and S. Sivakumar 2021 IOP Conf. Ser.: Mater. Sci. Eng. 1110 012007. Performance of particle swarm optimization bin packing algorithm for dynamic virtual machine placement for the consolidation of cloud server. IOP Conference Series: Materials Science and Engineering, Volume 1110, Annual International Conference on Data Science, Machine Learning and Blockchain Technology (AICDMB 2 021) 15th-16th February 2021, Mysuru, India

[23] O. Cordón, F. Herrera, F. Hoffmann, L. Magdalena, Genetic FuzzySystems: Evolutionary Tuning and Learning of Fuzzy Knowledge Bases, World Scientific, 2001.

[24] Binbin Zhang, Xiao Wang, Hao Wang, Virtual machine placement strategy using cluster-based genetic algorithm, Neurocomputing, Volume 428, 2021, Pages 310-316, ISSN 0925-2312, https://doi.org/10.1016/j.neucom.2020.06.120.

[25] Tabrizchi, H., Kuchaki Rafsanjani, M. Energy Refining Balance with Ant Colony System for Cloud Placement Machines. J Grid Computing 19, 7 (2021). https://doi.org/10.1007/s10723-021-09547-1

[26] F. Ahmed1 and A. Al Nejam,." Cloud Computing: Technical Challenges and CloudSim Functionalities", International Journal of Science and Research, vol. 2, 2013.

[27] Tarun Goyal, Ajit Singh, Aakanksha Agrawal, Cloudsim: simulator for cloud computing infrastructure and modeling, Procedia Engineering, Volume 38, 2012, Pages 3566-3572, ISSN 1877-7058, https://doi.org/10.1016/j.proeng.2012.06.412.

[28] Franke, C., Lepping, J., Schwiegelshohn, U., 2007a. Genetic fuzzy systems applied to online job scheduling. IEEE International on Fuzzy Systems, July, pp. 1–6, doi:10.1109/FUZZY.2007.4295601.

[29] Mamdani, E.H. and Others, 1974. Application of fuzzy algorithms for control of simple dynamic plant. Proceedings of IEEE 121(12), 1585–1588.

[30] Anton Beloglazov, and Rajkumar Buyya, "Optimal Online Deterministic Algorithms and Adaptive Heuristics for Energy and Performance Efficient Dynamic Consolidation of Virtual Machines in Cloud Data Centers", Concurrency and Computation: Practice and Experience (CCPE), Volume 24, Issue 13, Pages: 1397-1420, John Wiley & Sons, Ltd, New York, USA, 2012

[31] R.P. Prado, J.E. Muñoz-Expósito, S. García-Galán Flexible Fuzzy Rule Bases Evolution with Swarm Intelligence for Meta-Scheduling in Grid Computing. Computing and Informatics, vol. 3, no. 4, pp. 810-830, 2014.

[32] S. García-Galán, R.P. Prado, J.E. Muñoz-Expósito Swarm Fuzzy Systems: Knowledge Acquisition in Fuzzy Systems and its Applications in Grid Computing IEEE-Transactions on Knowledge and Data Engineering, vol. 26, no. 7, pp. 1791-1804, 2014 DOI: 10.1109/TKDE.2013.118

Doraid Seddiki (M.Sc) received his M.Sc. degree in telecommunication engineering from Cartagena University. Currently he is a Ph. D student at University of Jaén. His current research interests include artificial intelligence, cloud computing and scheduling. In addition, he also has a wide experience working in the private sector for more than 20 years.

Sebastián García Galán (Ph.D) is an associate professor at the Telecommunication Engineering Department of the University of Jaén. His research interests include artificial intelligence, IoT technologies, biomedical signals and cloud computing. He is the author of more than 100 research publications. He is an advisor of the Spanish Assessment and Prospective Agency in the field of Electronic Engineering and Communications.

José Enrique Muñoz-Expósito (Ph.D) received an Ph.D. in Telecommunication Engineering from Jaén University. He is an associate professor in the Telecommunication Engineering Department of Jaén University. His research interests include speech and audio analysis, artificial intelligence and cloud computing. He is involved in research projects and also works for the editorial reviewer board of the Annual Telematics Engineering Conferences.

Manuel Valverde Ibañez (Ph.D) received an M.S. degree in Industrial Management Engineering from the University of Jaén, Spain, in 2002, and his Ph.D. degree in Industrial Engineering from the UNED in 2006. He is an associate professor in the Electrical Engineering Department of the University of Jaén, Spain. His research

interests include renewable energy power systems, power quality, modeling and control of power converters and smart grids. He is the author of almost 30 research publications.

Tomasz Marciniak (Ph.D) is an associate professor in the Telecommunication, Computer Science and Electrical Engineering Department of the Bydgoszcz University of Science and Technology. His research interests include networking, artificial intelligence, IoT technologies, and bioinformatics. He is the author of more than 60 research publications.

Rocio Pérez de Prado (Ph.D), Senior Member IEEE, she is an associate professor at the Telecommunication Engineering Department at the University of Jaén. Her research interests include soft-computing and cloud computing. She has authored more than 45 research publications. In addition, she has served as an external evaluator of the European Commission and the National Committee of Science and Technology and Technology Innovation in Pérú.