A Novel Green Service Level Agreement for Cloud Computing using Fuzzy Logic

Awatif Ragmani, Amina El Omri, Noreddine Abghour, Khalid Moussaid and Mohamed Rida
Department of Mathematics and Computer Science, Faculty of Sciences Ain-Chock, Hassan II University of Casablanca, B. P. 5366 Maarif, Casablanca, Morocco

Keywords: Cloud Computing, Service Level Agreement, Energy Efficiency, Virtual Machine Consolidation, Fuzzy Logic.

Abstract: Cloud computing has several features including elasticity and economy of scale that have allowed it to find several uses in scientific, economic and industrial fields. However, the Cloud model relies heavily on the architecture of the data centers that host the Cloud services. To date, these data centers consume huge quantities of electrical energy in order to ensure the operation of both computer equipment and auxiliary equipment such as cooling systems. In order to reduce the environmental and economic impact of this energy consumption, several initiatives have emerged especially in the context of Green computing. Through this article, we propose a Cloud architecture that includes a service level agreement negotiation module based on the concept of fuzzy logic. The proposed solution aims to introduce a virtual machine consolidation policy in order to complete a global three-tier architecture. The final solution includes different modules of load balancing and scheduling based on fuzzy logic, metaheuristic and Map reduce algorithms in order to optimize both energy efficiency, response time and cost of Cloud services.

1 INTRODUCTION

Cloud computing is an innovative concept that has quickly gained an important place in the scientific and industrial fields. The Cloud services model is based on the use of a legal and business agreement which manage the relationship between the customer and the service provider (Jamshidi et al., 2014). His contract is commonly referred to as service level agreement (SLA). This agreement sets, among other things, the pricing terms and the level of quality of service (SLA). This agreement sets, among other things, the pricing terms and the level of quality of service (Wieder et al., 2011). The main principle of the Cloud is based on the use of virtual resources hosted on the different physical servers that are grouped within data centers. The administration of resources within the data centers relies on different tasks including the creation of virtual machines (VM), the allocation of virtual machines within physical servers, the migration of VMs and the removal of VMs at the end of users’ requirements. The process of migrating a virtual machine from one node to another could be motivated by energy efficiency reason through the consolidation of multiple virtual machines within a single server and turning off the other servers that are no longer used. The virtual machines consolidation cloud have other motivations such as enhancing the performance or reliability of the system (Furht and Escalante, 2010; Josyula, 2012).

Despite the multiple advantages of Cloud services, the expansion of data centers which host those services has induced a negative impact on energy consumption. In order to limit this energetic effect on the environment and to reduce the operational cost, several research studies have tackled the issue of energy efficiency at different levels. Those research studies could be classified according to three themes that include the improvement of the hardware used, the research papers that propose scheduling and load balancing strategies and the works that propose hybrid solutions combining the hardware and software aspect (Banerjee et al., 2013; Horri et al., 2014; Marotta et al., 2018). One of the commonly used methods to improve the energy efficiency within data centers is the dynamic voltage and frequency scaling (DVFS) method. This concept is based on the use of the cubic relationship between CPU operating frequency and power consumption. This technique is applied both at the level of computer...
servers as well as network switches. In order to improve the DVFS technique, a second method called dynamic power management (DPM) has been proposed. This technique allows switching off or putting into standby mode one or more components. Particularly, the authors of (Kliazovich et al., 2013) highlight the strong positive impact on energy efficiency of shutting down an inactive computer and network hardware servers. However, they issue a reservation as to the inability to wake up a sleeping server in a reasonable time which could degrade the quality of service. In other words, the establishment of a well-organized strategy for energy efficiency does not rely merely on efficient physical machines. The overall policy must take into account all the mechanisms used within the Cloud environment. Several research papers have in turn addressed one aspect of energy efficiency within the Cloud environment. The best approaches remain the one that combines the use of energy-aware servers and smart virtual machines placement. The present paper aims to introduce a green service level agreement model taking into account the energy efficiency aspect. The main contribution of this article is the definition of a Cloud architecture able to track SLA violations in terms of performance and energy consumption in order to guarantee the energy efficiency of the Cloud data center while respecting the quality of service levels. The rest of the paper include section 2 which summarizes the main research papers that have oriented the present article. Section 3 introduces the key concept behind the suggested solution. Section 4 highlights the proposed solution and section 5 concludes the present paper.

2 RELATED WORK

Optimizing energy consumption in the Cloud cannot be accomplished without the active and responsible participation of customers and providers. Particularly, several research studies have tackled the implementation of a green service level agreement able to guarantee a better respect for sustainable development and environmental standards within data centers. The following section introduces some related work of SLA implementation, including green SLA. The article (Goyal et al., 2016) tackles the problem of minimizing energy consumption by relying on the definition of a service level agreement (SLA) oriented green computing. The authors proposed a resource allocation algorithm via a Green SLA agreement model. The article includes Cloud service negotiation policies that provide the best energy efficiency for users. In addition, the authors of (Moreno and Xu, 2011) defined a resource allocation model in a real-time Cloud environment. Thanks to a better estimation of the requested resources, the system minimizes energy waste. In addition, the system limits the negative effects of energetic optimization policies on the SLA requirements through dynamic remuneration. The tests carried out by the authors underline the major role of the policy of virtual machine migration on the performance of the system.

The authors of the research paper (Djemame et al., 2011) emphasized the role of the services level agreement in expanding the commercial uses of the Grid concept. In order to improve the efficiency of SLAs by minimizing the probability of failure, the authors introduced an SLA trading mechanism taking into account the risk assessment aspect. The proposed model makes it possible to rate an SLA based on the probability of failure which has increased the reliability of the system.

In the article (Verma, 2004), the authors present a general view of the specificities of service level agreements in IP networks. Their approach includes a study of the specific components of a service level agreement by specifying all three and identify three axes common to all SLAs in IP networks. Finally, the authors introduce the possibilities of dynamic SLA implementation that are more adapted to the IP network context than static SLAs. The authors of the research paper (Carlsson and Fullér, 2013) have defined a probabilistic and possibilistic risk model that is used to measure the risk of an SLA in a grid or Cloud computing environment. The authors relied on a predictive probabilistic approach and the definition of an increase in the number of failures for the management of computing resources. Moreover, the proposed model makes it possible to estimate the possibility of future failure of a node based on the concept of fuzzy non-parametric regression.

Another study (De Marco et al., 2015) emphasized the role of the service level agreement as a guarantor of the commitments negotiated by both parties, including the customer and the supplier, during the period of validity of the service. The authors proposed addressing the problem of breaches of the SLA by cybercriminals without both parties realizing it. The solution proposed by this paper is to guarantee a better control of the respect of the SLAs by analyzing the log files. The final model relies on a mechanism for automatic detection of SLAs violations. Regarding the effectiveness of fuzzy logic, the authors of the research papers (Jamshidi et al., 2015, 2016) have proposed an automatic scaling
approach that allows cloud applications to maintain an adequate level of resources in order to minimize costs and monitor performance during phases of high demand. The authors introduced a hybrid cloud controller using both fuzzy logic concepts and Q-Learning. This combination has proved its efficiency thanks to the quality of the update of the control rules.

In summary, several studies have been interested in the evaluation of energy efficiency in data centers and have proposed different techniques that are applicable at the level of hardware, scheduling, and load balancing. On the other hand, different research papers have examined the strength that an SLA could play in optimizing the performance of Cloud services and improving the security and reliability of services. Through the present paper, we will assess the possibility of using a SLA negotiation module for reducing energy consumption.

3 PRELIMINARIES

This section introduces the main concepts taken into account when developing the Cloud computing architecture presented in this article, including the notion of three-tier architecture and service level agreement.

3.1 Service Level Agreement

The establishment of a service level agreement is a complex process due to the number of stakeholders and the multitude of constraints to be included. The quality of service levels offered in a SLA must not exceed the actual need of users or be below their expectations. Service level agreements are geared more towards the accuracy of the expected service levels rather than specifying the mode of service operation. The common description of a SLA is a support which includes the details of the defined services, the levels of quality of service, the measurement indicators and the commitments of the supplier and the user on the clauses to be respected. Any taking over by one party or the other is considered as a violation of the SLA. The service level agreement acts as an arbiter in the relationship between the Cloud services provider and other parties who may be a service user or negotiator.

In short, the SLA provides an objective and measurable translation of service level objectives such as performance, availability and pricing plan. In addition, the SLA describes the remedies and penalties in the event of litigation or contravention.

The service level agreement has been widely used in the field of IT applications and services including web services in order to describe the evaluation criteria and key performance indicators. The SLA allows each stakeholder to monitor compliance with commitments which contributes to the establishment of an effective partnership between user and service provider (Aljoumah et al., 2015). In particular, SLAs have become very important in terms of Cloud computing because of the obligation to measure the quality of Cloud services and to draw up periodic performance reports. However, the complexity and scope of Cloud models make it difficult to identify causes of service disruption (Verma, 2004). Finally, as described in Figure 1, a typical SLA follows six major phases to be accomplished that include:

- the creation of a service model and its attachment to a suitable SLA;
- negotiating terms of the SLA defined;
- provisioning and implementation of the requested service;
- the execution of the service;
- the monitoring and application of the necessary corrections during the execution phase;
- the closing and decommissioning of the achieved or terminated service.

3.2 Data Center Topology

In recent years, the three-tier architecture (see Figure 2) including servers and switches has become the most applied layout within a data center. This architecture consists of three levels, the main level which corresponds to the root of the tree, the level of aggregation that takes care of the routing, and the level of access that regroups all the servers in the data center. This new architecture is an evolution of the previous version which included only two levels, without the level of aggregation. But this two-tier configuration was soon overtaken by the bandwidth constraint that limited the capacity of hosts to 5,000 units depending on the category of switches used. Considering the size of the server pool in today's data center, which are of the order of 100,000 hosts, it becomes essential to switch to the three-level architecture. Note that the three-tier architecture organizes the servers in racks that are interconnected using a 1 Gbit Ethernet link (Kliazovich et al., 2013).
3.3 Fuzzy Logic Controller

The proposed architecture is based on the use of the different module in charge of the various process including the load balancing, scheduling, and the SLA violation. In this article, we tackled the SLA violation detection module which uses the concept of fuzzy logic. The fuzzy logic is defined as being an extension of Boolean logic. This concept was introduced by Lotfi Zadeh in 1965 based on the theory of fuzzy sets, which is a generalization of classical set theory. This theory uses the notion of the degree of truth of a condition which allows a condition to be situated in another state other than false or true (see Figure 3). This reasoning is based on the fact that it is closer to human reasoning and can cover a broader set of options without mentioning the fact that uncertainties are taken into account (Chen and Pham, 2001). The definition of the fuzzy is represented by a function which describes the degree of the membership. This mathematical function can get different values between 0 and 1. In brief, a fuzzy set $A$ is defined as pairs $(x, u_A(x))$ and $A = \{(x, u_A(x)) | x \in A, u_A(x) \in [0,1]\}$. One of the essential components in fuzzy logic is the defuzzification stage. Several methods have been propped, but the most applied remain Mamdani and Larsen (Chen and Pham, 2001). As depicted in Figure 4, the proposed Cloud model is based on a three-tier architecture. The authors of (Jamshidi et al., 2014, 2015) have applied the fuzzy logic in order to define a very interesting self-learning Cloud controllers. They introduced that one main aspect of fuzzy logic is the fuzzy inference which is the procedure of mapping a list of control inputs to a list of control outputs by applying fuzzy logic rules. This mapping is used to describe the controlled output. The significant advantage of fuzzy controllers still the definition of the control system for problems which cannot be characterized by precise mathematical formulas because of the non-linearity of the studied system. Hence the interest of fuzzy logic which offers an alternative solution to this problem by defining an approximation based on the use of the knowledge in a comparable manner to the human reasoning.

4 PROPOSED ARCHITECTURE

As noted in previous sections, the environmental aspect is becoming increasingly important in the design and operation of data centers. In this section, we introduce a Cloud architecture that is based on the concept of fuzzy logic which has been applied to the service level agreement module in order to detect the energy consumption level violations. The proposed module aims to implement a green SLA in order to ensure energy-aware in a data center by consolidating virtual machines on a minimal number of servers. Figure 4 schematizes the proposed architecture for Cloud environment. The proposed architecture is composed of several multi-objective modules that aim at the overall purpose to guarantee the best response time, the optimal energy efficiency and the lowest service cost. Particularly, this article focuses on the description of the module relating to the detection of SLA violations concerning the level of energy consumed. This module has been defined using fuzzy logic for two main reasons. Firstly, the fuzzy logic allows us to have an intelligent monitoring of the level of energy consumption and secondly this additional control does not increase considerably the response time of the Cloud services because of the simplicity of the fuzzy controller. The first level includes the main controller, whose mission is to centralize the management of the data center. After receiving requests from users accompanied by their co-signed SLA specifying the commitments of
each party, the main controller generates virtual machines in charge of processing the users’ requests according to a process that includes the steps of scheduling, mapping, load balancing and in case of necessity virtual machine migration. This process has already been detailed in our previous work (Ragmani et al., 2016). The main controller is supported by the secondary controller, whose function is, on the one hand, to reduce the workload of the main controller by monitoring the state of the network and maintaining a database describing the status of the hosts and their workloads at intervals T. The secondary controller informs the primary controller only in the event of a change of state. Lastly, the secondary controller provides the backup and the relay of the main controller in case of failure. The second level includes the regional load balancer. Following earlier studies, we have shown that the response time is greatly improved when assigning a user to a node close to him geographically. Each load balancer is connected to a set of nodes which constitutes the third level of the proposed architecture. Each node has a technical capability and can create a set of virtual machines which could respond to users’ requests. In addition to the components of the three-tier architecture, the proposed model includes a module for detecting SLA violations in terms of performance and energy consumption. This module is based on the concept of fuzzy logic. The choice of control technique focused on fuzzy logic for two main reasons. First, the fuzzy logic is fast and simple to implement which contributes to maintaining the performance of the data center and provides an additional level of control. Secondly, fuzzy logic makes it possible to take into account uncertainties and complex cases where an arbitrary choice must be made between response time (RT) and power consumption (PC) based on a SLA violations detection indicator (SLAVDI).

In other words, the fuzzy logic module control rules rely on logical rules in the form of If-Then clauses. These rules are defined by numerical measurements or knowledge base. In short, the control performed by the module of the fuzzy logic follows four steps:

- the description of the energy efficiency and performance control rules based on the co-signed SLA content;
- the definition of membership functions on the knowledge base recorded by the system;
- the application of the logical rules defined by the system administrator via the main controller;
- the application of the defuzzification module.

Thus, at the time of the co-signature of the SLA, a set of rules is generated based on the commitments of both parties. These rules, described below as fuzzy controls, will be used to assess compliance with SLA clauses:

- If (power consumption is low) and (response time is very fast) then (SLAVDI is very good).
- If (power consumption is medium) and (response time is very fast) then (SLAVDI is good).
- If (power consumption is high) and (response time is fast) then (SLAVDI is medium).
- If (power consumption is very high) and (response time is medium) then (SLAVDI is bad).
- If (power consumption is very high) and (response time is very fast) then (SLAVDI is bad).
- If (power consumption is very high) and (response time is slow) then (SLAVDI is very bad).
- If (power consumption is low) and (response time is slow) then (SLAVDI is bad).

Based on the value of the SLAVDI, the main controller could activate the virtual machine migration process in order to decrease the overall power consumption of the data center. As depicted in Figure 5 and 6, the total power consumption of a data center increases considerably if we apply a virtual machines migration policy. These figures correspond to the energy consumption recorded during two simulations carried out within the CloudReport (Teixeira Sá et al., 2014) and CloudSim platform using one data center and 5 consumers.
5 SIMULATIONS AND RESULTS

In order to achieve the various simulations (see Table 1), we use the CloudReport simulator (Teixeira Sá et al., 2014). This toolkit is a graphic simulator based on the concept of Cloud. CloudReport offers the opportunity to disconnect the programming and experimentation roles. This simulator allows the accomplishment of repeatable trials by changing simulation parameters and modelling Cloud architecture by defining one or more data center which includes multiple hosts. Each server may host numerous virtual machines. CloudReport parameters could be adapted in order to perform different simulation in economic and easy manner. Those parameters include the number of VM, number of hosts, customers, VM scheduling, broker policy, data size, power model and CPU utilization model. Particularly, the service broker policies parameter allows the control of the data center which achieves the user’s request. The CloudReport use by default a load balancing policy based on round-robin algorithm. The accomplishment of the predefined scenarios (see Table 1) has been planned in two stages. For the first time, we achieve different experimentation on CloudReport. As depicted in Figures 7 and 8, the power consumption is highly influenced by CPU utilization and scheduling model. Indeed, in the case of applying the stochastic configuration for CPU, RAM, and bandwidth utilization model, the achieved result is better than the case of applying the full configuration. All these conclusions are used to update the knowledge base used by the fuzzy controller in order to reduce the value of SLAVDI (see Figure 9). In a second time, we apply the fuzzy logic toolbox of Matlab to calculate and predict the SLAVDI indicator per scenario (see Figure 10). In summary, we apply 3 scenarios by modifying the CloudReport parameters. Each scenario has produced a report and raw data which has been used to finalize the fuzzy module for service level agreement detection.
Table 1: CloudReport simulations.

<table>
<thead>
<tr>
<th>Scenarios</th>
<th>Number VM</th>
<th>RAM/VM</th>
<th>Bandwidth</th>
<th>CPU utilization model</th>
<th>RAM utilization model</th>
<th>Bandwidth utilization model</th>
<th>File size</th>
<th>VM migration</th>
<th>Power model</th>
<th>Maximum Resources</th>
<th>Power consumption</th>
<th>Average finish Execution time (ms)</th>
<th>Resource utilization Max CPU (MIPS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>10</td>
<td>512</td>
<td>10</td>
<td>Full</td>
<td>Full</td>
<td>Full</td>
<td>500</td>
<td>4 Enabled</td>
<td>Linear</td>
<td>250</td>
<td>25</td>
<td>25</td>
<td>1700</td>
</tr>
<tr>
<td>S2</td>
<td>10</td>
<td>512</td>
<td>10</td>
<td>Stochastic</td>
<td>Stochastic</td>
<td>Stochastic</td>
<td>500</td>
<td>4 Enabled</td>
<td>Linear</td>
<td>250</td>
<td>22</td>
<td>17</td>
<td>1700</td>
</tr>
<tr>
<td>S3</td>
<td>14</td>
<td>512</td>
<td>10</td>
<td>Stochastic</td>
<td>Stochastic</td>
<td>Stochastic</td>
<td>500</td>
<td>3 Enabled</td>
<td>Linear</td>
<td>250</td>
<td>28</td>
<td>32</td>
<td>1700</td>
</tr>
</tbody>
</table>

Figure 7: Provider resources utilization during scenario 1.

Figure 8: Provider resources utilization during scenario 2.

Figure 9: The fuzzy logic rules.

Figure 10: Response surface view.

6 CONCLUSIONS

This article has proposed a three-level architecture that includes a control module based on the concept of fuzzy logic that can detect SLA violations in terms of energy consumption. The main role of the introduced module and the deceleration of the virtual machine migration procedure in order to reduce the energy consumption while maintaining a level of quality of service. Thus, the role of fuzzy logic is based on the calculation of an SLAVDI performance indicator that takes into account both the level of energy consumption and the response time of the system. CloudSim and CloudReport simulations demonstrate the positive impact of an efficient and fast migration policy on the energy consumption of the data center. In a second step, we plan to finalize the implementation of the proposed solution within a Cloud model and to upgrade the fuzzy controller in order to be autonomous by using machine learning algorithms. The choice of fuzzy logic was justified by the efficiency and rapidity factors. Finally, we aim to finalize a comparative study of the proposed method regarding other concepts such as (Arabnejad et al., 2016; Buyya et al., 2011; Jamshidi et al., 2015).
ACKNOWLEDGEMENTS

The authors would like to present their gratefulness to the anonymous referees for their valuable suggestions which have greatly contributed to improving the content of this paper.

REFERENCES


Chen, G., Pham, T.T., 2001. Introduction to fuzzy sets, fuzzy logic, and fuzzy control systems. CRC Press, Boca Raton, FL.


