International Journal of Grid and Distributed Computing
Vol. 9, No. 8, (2016), pp.319-334
http://dx.doi.org/10.14257/ijgdc.2016.9.8.28

Energy Consumption Management using Fuzzy Logic in Distributed Systems Environment

Elham Haririan[1] and Amir Masoud Rahmani[2]

[1]Department of Computer Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran
[2]Department of Computer Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran
[1]elhamharirian@yahoo.com , [2]rahmani@srbiau.ac.ir

Abstract

Cloud computing has based on on-demand access to potentially significant amounts of remote datacenter capabilities in distributed systems environment. By propagation data centers in clouds, computers should have a lower energy consumption. Hence, a good prediction about the amount of needed resources is very helpful for correct decision on energy consumption management. In this paper the Fuzzy Logic approach has been used for decision making about the performance of processors to obtain the estimated amounts of energy consumed by them and have been applied Auto Regressive and Neural Network approaches to predict the future workload. In prediction operation, three modes On, Off and Idle are considered for each server. At times of peak demand all servers must be turned on but when consumption is low, an estimated number of servers that must be turned on is obtained, 10% of remaining servers are idle and others are off. In this case a significant amount of energy is saved compared to the case that all servers are clear. The result show that Neural Network approach has more accurate prediction and better performance than Auto Regressive model.

***Keywords**: cloud computing, on-demand access, energy consumption management, Fuzzy Logic, Auto Regressive, Neural Network*

1. Introduction

Cloud computing industry provides a large ecosystem including many models and service providers and be configured set of computing resources (like networks, servers, storage space, applications and services) that they can quickly provide with minimal work and effortlessly or the need for the involvement of the service provider. Because of growing requests and connection of new customers to the world of cloud computing, attention to power consumption of data centers and available energy reservoirs and designing efficient strategies for energy storage is very important and essential.

In order to avoid wasting resources, a new load balancing technique called cloud light weight (CLW) has been suggested that not only balances the workload of virtual machines in Cloud Computing datacenters but also provides the high level QoS for users. This technique prevent SLA violations by load balancing system, based on the requirements of tasks and decreases the number of VM migration process and migration time for running application [1].

Obviously in addition to balancing the workload of system we can make significant reduction in energy consumption by switching servers for example turn off the server when not used. Therefor holding information from the past always is enhanced the quality of the decision-making.

Due to the complexity of the clouds there is a pressing need to configure the real test platform which can do directly monitoring and control operation and registers relevant

ISSN: 2005-4262 IJGDC

information. For this purpose, In this paper is used a data set named Grid5000 that is a real platform in a distributed environment and supervision operation and so Grid5000 is designed to provide a scientific tool for simulation of computer science problem. Considering the complexity of information an exact description and definition for them should be made; for this purpose, approximate or Fuzzy description can be accepTable because here is required the hypothesis that can be formulated human knowledge in a systematic way. It along with a mathematical model from a time series for predict the future to deliver suiTable method. In this regard, for Multi Criteria Decision Making the design of a fuzzy system is the best option.

Considering the extent of use of fuzzy logic in many researches, to decide on the issue of Grid market-based resource allocation are designed two new Fuzzy Grid Market Pressure Determination Systems (FGMPDSs) which facilitates negotiation agents for Enhanced Market and Behavior-driven [2]. A new fuzzy system is used for accommodate heterogeneity in cloud requests workflows in clusters of heterogeneous Grid that at the same time is provided more reliability and QoS expected for scheduling distributed cloud-based requests by using a multi-criteria advanced reservation algorithm [4].

After evaluating linear and non-linear models to forecast future workload, Auto Regressive model from linear series models of Box and Jenkins and Neural network that it is non-linear model were selected. Because Auto Regressive model is one of the most widely used prediction models and Neural network is superior to traditional linear patterns in recent decades. These two selected models are doing prediction operations by using Fuzzy system output. We can estimate the number of processor required using these methods and switch rest servers to idle or turned off for lower power consumption. When some servers are idle lower energy is needed to turn them on and also rejection rate of requests is significantly reduced because they are converted faster from the off state to the on state in the times of need, in fact can be said that the resources are available but they have been in the state which use less energy it and their productivity is high.

This paper is written in six sections. After this introduction as section 1, section 2 includes investigation of methods and conducted researches in the field of workload prediction, resource consumption and consumption of energy in cloud computing environment. The concepts used in this study are explained in section 3. In section 4 the main theme of research will be defined that it is energy consumption management in cloud computing environment. Section 5 describes how to carry out the proposed approach. Section 6 explains simulation and evaluation indicators and at the end, the study is concluded and suggestions for future work are indicated in section 7.

2. Related Works

Unique business model of cloud computing provides obligation to a payment based on the use of virtual services which allows in the time required to increase capacity on the Internet and in fact, it is one approach for development required existing resources in Information Technology (IT) without investing in new infrastructure and related costs of persons. In this section, the main focus is on the models of forecasting resources and the amount of energy consumption to facilitate proactive scalability in the cloud.

In [9] the provision of resources according to the application providers has been analyzed and evaluated. So that hosted applications are tended to create scaling independent decision by evaluation of future resource utilization in the form of real-time (Such as CPU, memory, input/output of network …). So for increasing the efficiency and usability, they request in advance for added virtual samples through smart prediction, the majority of available jobs are based on the vision of the cloud service provider for investigate of providing resources. Scope of [9] is limited to the development of a model for efficiency prediction and just its statistical validity and the proposed prediction technique is more effective in the future and for several reasons continues work based on predicted resources after the error in the cloud:

1. Produces past data by implementing a client-server standard, like TCP-W in the Amazon EC2 [14] and ases that information to learn and experiment prediction models.
2. Using statistical learning algorithms the framework of prediction is developed and uses mechanism of moving window that still is effective case.
3. The accuracy of Prediction Framework is checked using the proposed evaluation criteria and its validity is determined [20].

In [10] linear models were evaluated to predict the average workload in the future and a detailed statistical study is done to tracking workload on actual different machines to examine Box-Jenkins models (Contains AR, MA, ARMA and ARIMA). Needs related to these models are measured to connect to an online prediction system in a wide range and predictive ability of models is evaluated with great accuracy. The main result is that predict ability is valuable and simple and practical AR model has competence for predicting workload.

An application will be able to choose a suiTable host to execute tasks if it has ability to predict the execution time of the tasks on any available host; In [11] guidelines for the Run Time Access (RTA) described and evaluated. RTA predicts the runtime of a task in a calculated confidence interval at the confidence level of the applications to determine the changes inherent in distributed systems and the anticipated processing of information. Confidence interval is an example abstract for applications but still there is not adequate information and statistical reasons in the process of scheduling; Therefore Dynda predicts execution time of the tasks with a method based on prediction of CPU loading and to predict the CPU loading is used AR(16) model. Much research has been done in the field of energy management and resource of clusters and data centers. Energy consumption of data centers has a unique growth and most of the energy is lost in unemployed systems. PowerNap is a method for energy consumption where server quickly moves between an active state and an idle state in response to loading with a force close to zero, called the "nap" [17]. In fact, the PowerNap method is used to eliminate energy consumption of idle servers.

In [11] a similar case without electricity is presented which allows remote access to memory of main servers, even when many of the other components are turned off. At first, a new case is presented that called the barely-alive and its memory can be available when a server is in this state although its CPU is completely off. Then a middleware will be considered in the cache memory for the barely-alive servers. Middleware is dynamically measured cache memory and uses distributed stack algorithm corresponding to the optimal access rate. The results show this middleware of cache memory and the barely-alive state have appropriate energy storage and create mechanisms to facilitate memory management in the range of clusters for applications in the cache memory level; the only problem of this method is the downside that is added due to the especial hardware requirements.

Green scheduling algorithm is used to energy storage in cloud computing using a neural network based on the prediction [25]. Green scheduling algorithm works such that first, according to the prediction, required dynamic workload on the servers is estimated, then unnecessary servers turned off to reduce the number of executive servers. Neural Networks used to prediction of the future workload request on the servers based on the last request. In fact, designed neural network is employed in view of the prediction of request and delay restart the server for predicts the user's request to turn off or on servers.

The main objective of this section was introduce the prediction methods that have been used to predict the required resources and reduce energy consumption in the past and inspired by the research conducted and the advantages and disadvantages of each. In the next section is introduced and described suggested method of this research.

3. Concepts

3.1. Fuzzy Logic

Human brain interprets inaccurate and incomplete information provided by sensory organs. The theory of fuzzy sets provides a systematic computing method to talk about linguistic information. This theory is based on the decisions on inputs and in the form of derived linguistic variables from membership functions that are formulated by using the specification of fuzzy set to derive the amount of belonging and the degree of membership in the set. The variables are equal by the terms of the Language linguistic of IF-THEN rules and reply of every law will be obtained from the Fuzzy concept. In implementing combination rules of inference, reply the law is based on the weights or the degree of membership for its inputs and center solutions calculated to produce the appropriate output [11].

3.2. Auto Regressive Model

A random process can mathematically be defined in the form of a random variable set $\{X(t), t \in T\}$ where T shows the set of time points and the process is defined in that points. If T is continuous (usually $-\infty < t < \infty$) random variable is shown at the t time with X(t) And if it is discrete (usually t = 0, +1, +2, ...) is shown with X_t. So a random process is a set of random variables which are arranged by the time. For only one outcome of the process, there is only one view for any random variable [3]. Suppose $\{Z_t\}$ is a random pure process with zero mean and variance σ_z^2 in which case $\{X_t\}$ called a p-order auto regressive process if:

$$X_t = \alpha_1 X_{t-1} + ... + \alpha_p X_{t-p} + Z_t \quad (1)$$

Actually, X_t is not regression on the independent variables, but is regression on the past values of X_t, Therefore called "Auto Regressive" [3].

3.3. Neural Network

Artificial Neural Networks are an attempt to model the neural information processing systems and they have a multi-layer structure hierarchy where a unit can communicate with other units. As the primary motivation, Artificial Neural Networks have the structure of biological systems and make up the computational paradigm alternative. Nervous systems have been integrated architecture from variable composition but all of them are formed of the blocks that are called the nerve cells or neurons. They can perform different functions. In fact, neurons receive signals and generate the corresponding response [22].

Artificial Neural Network is a non-linear pattern that is able to approximation various types of relationships in data. Neural Networks have flexible computing framework to model a wide range of nonlinear problems. The main advantage of these models compared to other linear models is that Neural Networks are comprehensive estimators and can approximate a wide range of functions with high degree of precision. The power of the Neural Networks are derived from parallel processing of data and feed forward Neural Network with one hidden layer is the most common form of neural network model for modeling and prediction.

4. Problem Definition

In fact, this research consists of two phases, an estimate of the amount of energy consumption of computing resources will be obtained by design of a proper Fuzzy system in the first phase and in the second phase, relatively accurate prediction of the amount of needed resources in the future based on the Fuzzy estimation is obtained by two famous method which include Auto regressive and Neural Network.

4.1. First phase: Energy Estimation Algorithm using Fuzzy Logic technique (EEFL)

4.1.1. Input and Output of Fuzzy System:

The variables used in this paper are:

- Job_ Number_Submited_Oneday: This input variable is the number of jobs submitted in one day.
- Sum_Job_RunTime_Oneday: It is one of the input variables which includes a total runtime of jobs in a day.
- Job_Successful_Percent_Oneday: This input variable includes percentage value of 1 (successful jobs) to 5 values (jobs canceled) and 0 (failed jobs) in a day.
- Sum_Energy_Amount_Estimated: It is the output variable that obtained by applying fuzzy system on the input variables and includes an estimate of the amount of energy consumed in a day.

So the designed fuzzy system has three inputs with an output that show in figure 1. Also it should be noted that the time of registered data in Grid5000 archive in second and includes data for about one year that it was used 5 months of registered data in this study and because of the need to predict the future, they have become to 24 hours that can predict the need to resource for next days based on data of previous day.

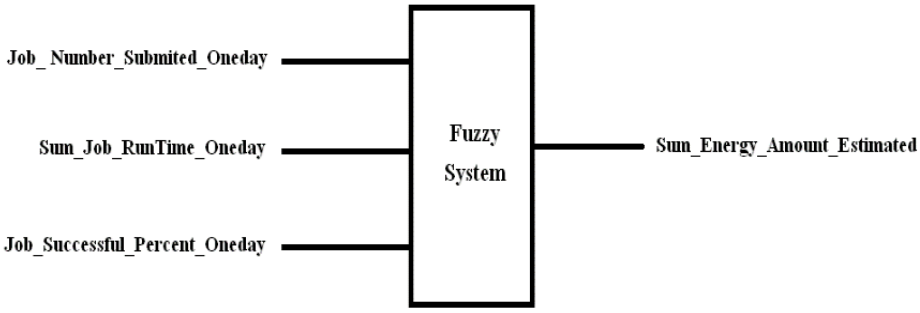

Figure 1. Designed Fuzzy System

Algorithm: Calculate Energy Amount Requirement for Every day with Fuzzy Logic System

Inputs: Job_Submit_Number, Job_RunTime, Job_Status
Outputs: Sum_Energy_Amount_estimated_Oneday

Method:
1. Submit job to the cloud
2. Register submit time, runtime, status for every job
3. Calculate number of submit, runtime, successful job for every day
4. Use Job_Number_Submited_Oneday, Sum_Job_RunTime_Oneday, Job_Successful_Percent_Oneday as inputs of Fuzzy Logic System
5. Apply Fuzzy Logic System on inputs
6. Get output from Fuzzy Logic System
7. Show output

Figure 2. The Algorithm to Estimate Daily Energy Consumption using Fuzzy Logic

4.2. Second Phase: The Proposed Algorithm for Predicting Amount of Required Energy of Processors

4.2.1. Energy Prediction Algorithm using Auto Regressive Model (EPAR): For predicting future required energy, initially value of estimated information by fuzzy system has been sent to Auto Regressive model for learning and then next values is predicted with this model and prediction accuracy is achieved by comparing the predicted values and the actual values. The algorithm related to do this job is shown in Figure 3.

Algorithm: Prediction with Auto Regressive model

Inputs: Energy amount estimated for n days ago
Outputs: Predict energy amount requirement for n days after

Method:
1. Select Least Square method for calculate α coefficient in Auto Regressive model
2. Calculate α
3. Multiple α in input data
4. Test Central Simple Moving Average method over result of step 3 with different values for m
5. Select and use best m value for use in Central Simple Moving Average method
6. Apply Auto Regressive model on result of step 5
7. Compare output data from Auto Regressive model with real data
8. Register error of Auto Regressive model

Figure 3. Energy Prediction Algorithm Using Auto Regressive Model

4.2.2. Energy Prediction algorithm using Neural Network (EPNN): Artificial neural network designed in this study is the multilayer Perceptron's network which can be seen in Figure 4.

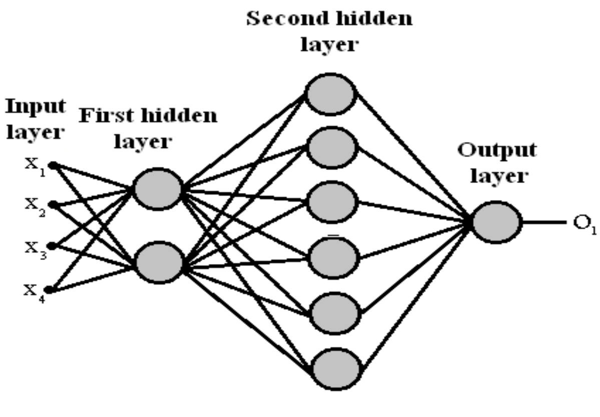

Figure 4. Designed Multilayer Neural Network

This neural network has four layers and has four inputs and one output. First layer is input layer and includes four input values. Second layer is the first hidden layer that includes tow nodes. Third layer is the second hidden layer that includes six nodes. Forth layer is output layer that includes one node.

In the neural network method like auto regressive model for predicting future required energy, value of estimated information by fuzzy system has been sent to Neural Network model for learning and then next values is predicted with this model and prediction accuracy is achieved by comparing the predicted values and the actual values in the frequent runs. The algorithm of this job is shown in Figure 5.

Algorithm: Prediction with Neural Network model

Inputs: Train data, Test data. Train data is included:

1. x_1: Energy amount estimated for n days ago
2. x_2: Energy amount estimated for n days ago with Central Simple Moving Average method by m=1
3. x_3: Energy amount estimated for n days ago with Central Simple Moving Average method by m=2
4. x_4: Energy amount estimated for n days ago with Central Simple Moving Average method by m=3

Outputs: Predict energy amount requirement for n days after

Method:

1. Design Neural Network model with 3 layer
2. Normalized train data
3. Use normalized train data as inputs of Neural Network model
4. Use FeedForward method for training Neural Network model and apply random weight with maximum 1000 iteration
5. Get output from NN model as output of train step
6. Compare output of train step with test data
7. Register error of NN model

Figure 5. Energy Prediction Algorithm using Neural Network

5. Proposed Method

This section introduces proposed method for energy consumption management using fuzzy logic technique in the estimation phase and using auto regressive and neural network in the prediction phase has been presented.

5.1. EEFL Configuration

With the variables of the problem each membership function should be determined. The membership functions for all variables are of 'trimf' type that represent the triangular membership function and are shown in figure 6.

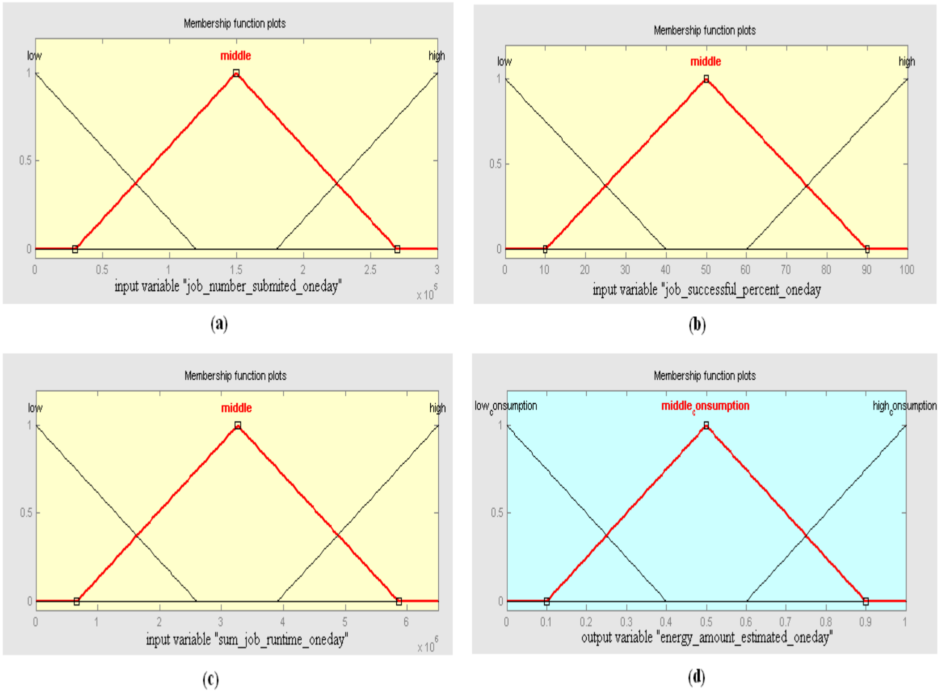

Figure 6. The Membership Function of the Fuzzy System Variables

These membership functions are formed of three parts ‘low’, ‘middle’ and ‘high’ that the first three diagrams show membership functions of input variables and the last diagram shows membership function of output variable.

After determining the fuzzy system components, some rules must be designed that apply on input variables and create an output variable. In fact, fuzzy logic can be used by these rules that are called fuzzy operators. Designed rules related to this study are shown in Table 1. After applying fuzzy rules on the input variables, energy consumption is obtained in the output. After obtaining fuzzy output, prediction based on it is started.

5.2. EPAR Configuration

In this study a strategy for making auto regressive model is used that is explained by Box and Jenkins and includes two major stages.

Table 1. Fuzzy Rules Related to designed Fuzzy System for estimate Energy Consumption

Rule number	Number of submitted job in one day	Percent of successful job in one day	Sum of runtime of jobs in one day	Sum of energy consumption in one day
1	low	low	low	low
2	low	low	high	middle
3	low	middle	high	high
4	middle	high	low	low
5	middle	high	middle	middle
6	high	low	high	high
7	high	middle	low	middle

8	high	high	low	low
9	high	middle	high	high
10	high	high	middle	middle

First Stage: Pattern Recognition

Obviously if exist certain dependencies between data during the time, a good opportunity happens so can predict future trends phenomenon with the help of those observations. The data used in study include the data of 151 days and variable under study is the output of a fuzzy system and prediction method is a univariate.

Second Stage: Pattern Fitting

In the fitting stage, by setting different parameters and compliance of obtained diagrams, a final diagram is obtained that have the least error. For this job, 75 percent of the available information is returned to auto regressive model until based on that information, it predict the remaining 25 percent and after that taken prediction is fitted with main diagram to determine the accuracy of prediction. In auto regressive model coefficient of the least square error has been used for calculation coefficient α. After obtaining the coefficient α, smoothing is needed to be done to analyze the results of the model and achieve a clear pattern. The easiest technique of this type is called Central Simple Moving Average (CSMA) of (1+2m) point. In this way, the actual observation X_t replaced with its average and m neighboring points so that shown in equation (2).

$$X_t = \frac{1}{2m+1}\sum_{j=-m}^{m} X_{t+j} \qquad (2)$$

According to equation (2) by using some investigations, a good value for (m) should be considered. Surely, fitting error is reduced with applying simple moving average method.

In fact, according to that all models of the Box and Jenkins Known as ARIMA are Linear-random and time series is assumed univariate in them and all the factors and effective communication in the form of a variable reflected in its values thus, for explain variability of variable can use its previous value of variable that is the most important source and predicted only with using previously information of the variable itself. According to this view if can achieve the process of generating a variable so predicting variable rather will be possible easily. But there is a problem when available variable has a non-linear behavior that in this case, the use of ARIMA models doesn't provide desirable results. As a result another method must be used to predict that the nonlinear nature of the available data does not interfere in prediction operations.

5.3. EPNN Configuration

In this section the steps that should be carried for operations of prediction in a neural network has been described.

First Stage: Data Partitioning

For using data to predict by neural network data should be divided into two parts trial data and test data. Trial data is the data that network is trained by them and test data is the data that used to test network performance. In fact, test data are the data that are given to the network to specify their network performance to predict the future. Input data of designed Neural Network in this study are output of fuzzy system. X_1 is exactly output of fuzzy system and has been entered to the neural network with no changes but X_2 , X_3 , X_4 are output of fuzzy system that CSMA applied with m=1, m=2 , m=3 on them. For training and test stage, 75% of the obtained data from Fuzzy Logic are given to neural network for training and 25% of the remaining data are given to neural network for test.

Second Stage: Training

Training stage can be done when training data are available with their answers. These data are called input data of training and their answers are called output data of training. In this study considering to have data problem, the most common learning algorithm is used that called Back propagation Algorithm. In Back Propagation Algorithm, stimulation function for every nerve considered to be the weighted sum of inputs of that nerve. Thus assuming w is corresponding weights between the input layer and the next layer, equation (3) is obtained.

$$A_j(\overline{x},\overline{w}) = \sum_{i=0}^{n} x_i w_{ji} \quad (3)$$

We can see clearly that the output of nerve stimulation function just depends on input and its corresponding weights therefore, for changing the output it must change weights. So that expressed previously, the purpose of the training process is to achieve the optimal output (or near-optimal). In this regard in the first step of implementation of the algorithm, the weight of all layers elected accidentally and the weights are corrected using Back Propagation Algorithm at every step. At the end of this stage, the system measures output according to what is learned.

Third Stage: Test

In test stage, to achieve the neural network error, the output of training stage compared with a portion of the data that are not logged in the system and called test input data; the obtained answers are called test output that with related diagrams will be shown in the next section.

6. Evaluation and Simulation

Implementation details and the results of the proposed method is expressed in this section and finally achieved results, their amount of the error, the reduction of energy consumption and ideally state for each of the servers is shown by obtained figures and Tables and the method having more accurate performance is selected as a suiTable method for prediction.

6.1. Grid5000 Platform

While now large distributed systems are supported by thousands of scientists but there are little data about the actual use of them. Due to strict licensing organization, quantitative tracking is done by researchers and their lawyers from archives of distributed system workloads. To solve this problem, Grid workload archive came into being that data of exchanged workload for the distributed systems stored in there at a certain time. In this archive a particular format is used for sharing workload information and data from 9 famous environments collected and analyzed using these tools which contains more than 2,000 users and sends more than 7 million jobs on the 130 sites containing over 10,000 source in a 13-year operating range [20].

Should be smoothed for analysis and to reach a 9 grid environment, Grid5000 platform is the only appropriate platform for this study according to studies conducted that is second archive between 9 famous Grid environments and is characterized by GWA-T-2. Grid5000 is an experimental Grid platform that includes 9 sites which are distributed in France and every site contains one or more cluster and there are a total of 15 clusters in the Grid5000. This means that 5000 processor is in it and every site has a 500 processor and only one site has 1000 processor and a total of 473 users will use it. Important attribute of this trial platform contains reconfigurable, ability to control and monitoring that provides test possibility on all layers of the software from network protocol to applications. As a result, record tracing of manipulating all batch scheduling of Grid5000 clusters is available since contains the beginning of Grid5000 project until November 2006 [12].

A prominent feature of the Grid5000 platform is that one job can request more than one processor and at the request of every job if it is possible to be assigned the same number of processors. In fact, one job can run parallel by multi-processor; also alongside this feature, every job run parallel and concurrent due to the multiplicity of processors and it is possible to increase the speed of response to user requests and following it, reduce the waiting time.

Grid5000 archive includes 1020195 records that have been saved in SQLite database and contains wide information. According to the factual and wide information obtained from this archive and because of the system under review did not crowd and the adequacy of the part of information from controlled archive, only 40,000 record is used for the purpose of this study which contains information about 151 days and its three parameters considered to be as input fuzzy system to be gained an estimate of the energy consumption amount.

6.2/ Estimation Phase

In this phase, a model based on fuzzy logic rules is designed and is used to estimate the amount of energy consumption in the cloud computing environment. This model has three inputs, triangular membership function for each fuzzy set, 27 main rule based on physics of problem and the characteristics of the environment under review and is doing operation estimate based on Mamdani fuzzy model.

Figure 7 shows a graph derived from the output of the fuzzy system. Given that data used in this study includes 151 days and variable under review is the output of a fuzzy system, the horizontal axis represents days and time and the vertical axis represents the output of the fuzzy system or the measured energy consumption with fuzzy system.

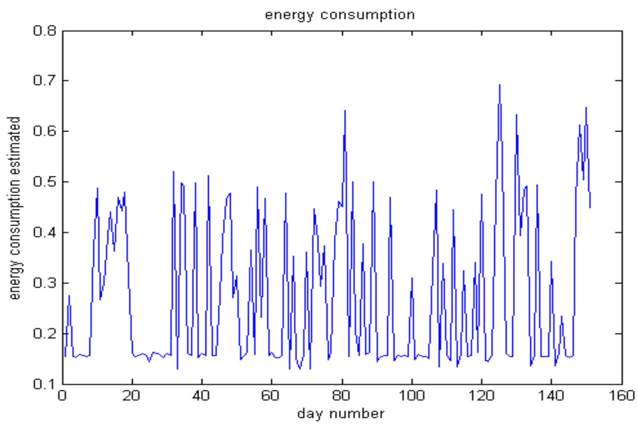

Figure 7. Estimated Energy Consumption using Fuzzy Logic

6.3. Prediction Phase

In prediction phase, at the first auto regressive method is used for predicting due to the linear nature of this method did not achieve appropriate result. Then a neural network is used for predict that due to has Predictive capability on the nonlinear data has favorable and accepTable result. It should be noted that training and testing operations in both methods, has the same data.

6.3.1. Predict using EPAR: The prediction result is not good after applying auto regressive method on the data from estimation phase, the result of this operation is shown in Figure 8.

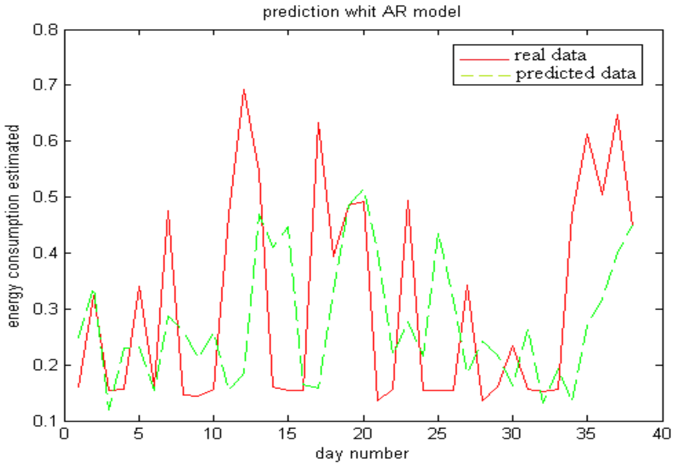

Figure 8. Fitting Prediction Result using auto Regressive Method on the Main Data

It is obvious that obtained error is caused by many changes in daily energy consumption on different days. As a result, require that prediction data should be smoothed for analysis and to reach a clear pattern in auto regressive method. In this study a Central Simple Moving Average method is used for smoothing data and by the trial and error method the m=3 is selected for smoothing that compared to other values which were given to m, it is better predicted output. Fitting result is shown in Figure 9.

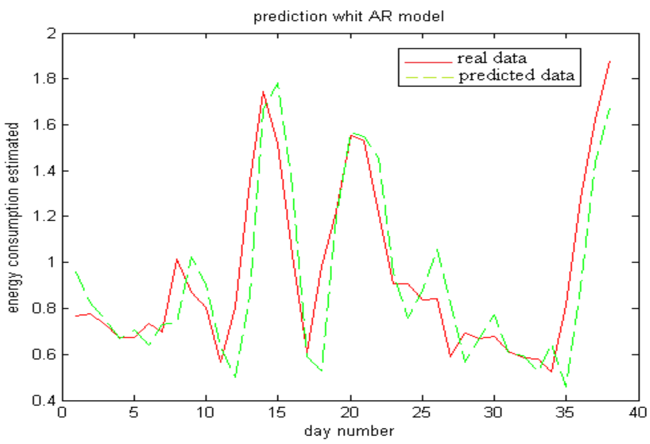

Figure 9. Fitting Prediction Result using Auto Regressive Method on Smoothing Main Data by CSMA Method

But because the result is not very favorable, an important tool among the topics Computational Intelligence used that called artificial neural networks.

6.3.2. Predict using EPNN: First, the input data should be normalized. This neural network has 4 inputs that all of them are normalized between 1 and -1 and after entry of learning data using Back propagation method, prediction operation is done based on them. Specifications of used neural network are shown in Table 2.

Table 2. Specifications of used Neural Network

Network Type	Transfer function of first hidden layer	Transfer function of second hidden layer	Transfer function of output layer	Learning function of BP method	Learning function of weights and bias	Maximum repetition	Range of normalized input value
Three layers neural network	tansig	purelin	tansig	trainlm	learngdm	1000 times	[-1,1]

Figure 10 shows predicted results in the first run.

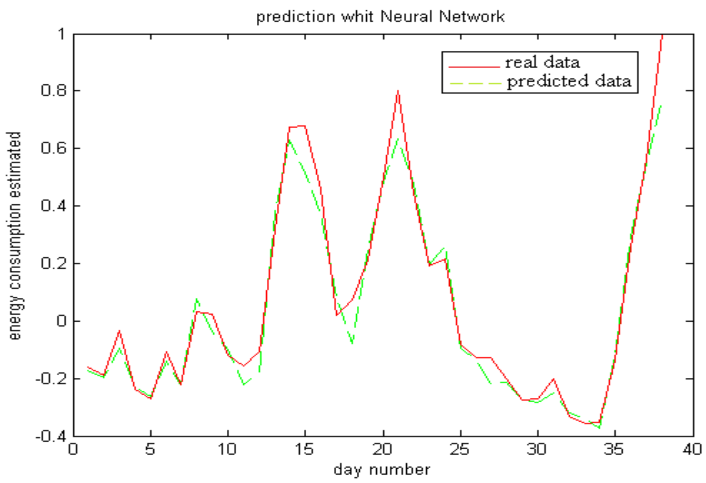

Figure 10. Fitting Prediction Result using Neural Network on the Main Data

6.4. Prediction Results of both Auto Regressive and Neural Network Methods

To evaluate the performance of these two methods, eight performance indicators were measured on their result that indicators include:

1. Maximum error: the most error of predicted value compared to the real value that obtained from subtracting the values.
2. Minimum Error: the least error of predicted value compared to the real value that calculated from subtracting values.
3. Mean Absolute Error: Mean Absolute Error is obtained from comparing a statistical estimator $\hat{\theta}$ according to the prediction parameter θ; its relation is shown in equation 4.

$$MAE = |E(\hat{\theta} - \theta)| = \frac{\sum_{i=1}^{n} (\hat{\theta}_i - \theta_i)}{n} \quad (4)$$

n shows the number of statistical estimators and in this study it means the number of days that prediction operation is done for them.

4. Mean Absolute Percentage Error: this error shows Mean Absolute Error in the form of percentage and is calculated by equation (5).

$$MAPE = MAE \times 100 \quad (5)$$

5. Normalized Mean Absolute Percentage Error: This error is calculated by equation (6).

$$NMAE = \frac{MAE}{\theta_{max} - \theta_{min}} \quad (6)$$

6. Root Mean Square Error: The calculation formula for this error is shown in equation (7).

$$RMSE = \sqrt{MSE(\hat{\theta})} = \sqrt{E((\hat{\theta} - \theta)^2)} = \sqrt{\frac{\sum_{i=1}^{n}(\hat{\theta}_i - \theta_i)^2}{n}} \quad (7)$$

7. Normalized Root Mean Square Error: This error is obtained by interval of the defined amounts from equation (8).

$$NRMSE = \frac{RMSE}{\theta_{max} - \theta_{min}} \quad (8)$$

8. Standard Deviation: This error is calculated by equation (9).

$$SD = \sqrt{\frac{\sum_{i=1}^{n}(\theta_i - \mu)^2}{n-1}} \quad (9)$$

These indicators show the difference between the predicted values by model or statistical estimator and the actual amount and they are good tools for comparing the prediction errors by a set of data and do not apply to compare several data sets. The results of measurement error indicators in the used model are shown in Table 3.

Table 3. The Result of Measurement 8 Error Indicators

Name of prediction method	Max Error	Min Error	ME	MAE	MAPE	NMAE	RMSE	NRMSE	SDR
Auto regressive	0.4978	-0.2978	0.0251	0.1445	14.4502	0.1090	0.0251	0.0190	0.3759
The total average error of Neural Networks method in eleven runs	0.2276	-0.0597	0.0259	0.0468	4.6756	0.0409	0.0259	0.0226	0.3288

As shown in Table 3 Multilayer Perceptron Neural Network with two hidden layers is superior than the auto regressive in terms of eight error index and this is because of the non-linear nature of neural network. In fact, the linear nature of Auto regressive method and the non-linear nature of neural network make relatively high error in auto regressive method compared to the neural network method. Characteristic of Perceptron neural networks is the ability to predict the types of non-linear data and according to the mean absolute error results, using this method is able to perform prediction operation considered in this study with 90% accuracy. So neural network method compared to auto regressive method acts more successful in predicting energy consumption and can be regarded as a basis for practical to predictions in the future.

6.5. Determine Server State using EPNN

Considering the designed system and since the main purpose of this study is energy saving, we can set the mode for servers that is wasted less energy. Therefore, in this study three state On, Off, Idle is considered for servers that using designed neural network can predict the amount of energy needed in the coming days and correspondent it, turn off unused servers and contract 10 percent of the servers can be off in idle state, because 10 percent of prediction error of neural network and the dynamic nature of cloud environments in the event of unforeseen needs, spent less energy and time to restart the servers.

To demonstrate how the operation prediction is and to determine the appropriate state for servers using design neural network, the amount of energy required in a later day predicted using data from previous days and the number of servers in each of the three target states is determined.

As it is expressed at the beginning of Section 6, Grid 5000 includes 5000 processor, for this reason in Table 4 the number of servers is determined in each of the states based on that each server has one CPU. To get the number of on servers used equation (10).

$$\textit{Number of On Servers} = \textit{Round} \ (\textit{Value of Prediction Requirement Energy} \times \textit{Number of Server in Cloud}) \quad (10)$$

The number of idle servers is obtained from equation (11).

$$\textit{Number of Idle Server} = \textit{Round} \ (\textit{Number of Server in Cloud} - \textit{Number of ON Servers}) \times 0. \quad (11)$$

The remaining servers in the site are the number of servers that can be turned off and is calculated by equation (12).

$$\textit{Number of Off Server} = \textit{Number of Server in Cloud} - (\textit{Number of On Servers} + \textit{Number of Idle Server}) \quad (12)$$

Considering the amount of required energy estimated by the fuzzy system (a value between 0 and 1), when all of the servers in one site are on, the amount of energy in that site is 1. Of course, with the shutdown the number of servers is saved significant amount of energy compared to when all servers in a site to be on. The percentage of this saving is obtained from equation (13).

$$\textit{Percentage of Energy Saving} = 100 - \textit{Percentage of Predicted Energy for Processors} \quad (13)$$

Table 4. Predict Future Energy and Determine Server State

Name of the prediction method	Number of all servers	Predicted energy to used processor	Number of On servers to processing	Number of Idle servers to processing	Number of Off servers to processing	Percent of energy saving
Neural Network	5000	0.2752	1376	363	3261	72.48

7. Conclusion

Using simulation method in this article, by providing information about all the days prior to simulator can predict the required energy for the next days. With careful reviewing of results of the simulation clearly we can see that energy consumption has decreased significantly and the goal of this study means removing the challenges of energy consumption and decrease it in cloud computing environment was achieved. Generally benefits of the proposed method in this study are as follows:

- Decrease energy consumption of processor in cloud environment
- Time saving and low cost of consumed resources
- Speed up of access to resources in the clouds in times of rising demand
- Ability to estimate the amount of use of cloud
- Reduce the rejection rate of requests to use cloud resources
- Better utilization of available resources

In this article, prediction operation is done for one day (*i.e.* 24 hours later) but the prediction interval can be smaller to be achieved more energy saving. But it should borne in mind that whatever predict interval will be small then computational overhead for prediction operation is greater and more times spent until the desired result is achieved.

Major work has been done in the field of prediction is in workload scope and CPU usage, in this study, the focus was on the same scope but in the further work we can review other factors such as reference to memory, energy consumption by memory, main board, fans and other attributes of servers. By doing this, the amount of energy consumed by them is predicted in addition to the energy consumed by the processor and thus to be earned a more comprehensive information about the future of system and besides reducing energy consumption by processors, reduction of energy consumption by other components can be evaluated.

References

[1] M. Mesbahi, A. M. Rahmani, A. T. Chronopoulos, "Cloud Light Weight: a New Solution for Load Balancing in Cloud Computing", International Conference on Data Science & Engineering, (2014), pp. 44-50.
[2] S. Adabi, A. Movaghar, A. M. Rahmani, H. beigy, H. Dastmalchy-Tabrizi, "A new fuzzy negotiation protocol for resource allocation", ELSEVIER, (2014), pp. 89-126.
[3] C. Chatfield, "The Analysis of Time Series: An Introduction", (1381).
[4] S. Adabi, A. Movaghar, A. M. Rahmani, "Bi-level fuzzy based advanced reservation of Cloud workflow applications on distributed Grid resources", Springer, (2013), pp. 175-218.
[5] D. S. Linthicum, "Cloud Computing and SOA Convergence in Your Enterprise", (2010).
[6] A. Li, X. Yang, S. Kandula, and M. Zhang, "CloudCmp: comparing public cloud providers", IMC, (2010), pp. 1-14.
[7] R. Buyya, J. Broberg, A. Goscinski, "Cloud Computing Principles and Paradigms", WILEY, (2011).
[8] V. Anagnostopoulou, S. Biswas, A. Savage, R. Bianchini, T.Yang, and F. T. Chong, "Energy Conservation in DatacentersThrough Cluster Memory Management and Barely-AliveMemory Servers", Proc.Of the Workshop on Energy-Efficient Design, (2009), pp. 7-12.
[9] A. Zade, "Fuzzy Logic Toolbox for Use with MATLAB", MathWorks, (1995).
[10] L. A.Zadeh, "Fuzzy sets", Information and Control 8, (1965), pp. 338-353.
[11] M. Sugeno, "Industrial Applications of Fuzzy Control", Elsevier, (1985).
[12] R. Bolze, F. Cappello, E. Caron, M. Dayd, F. Desprez, E. Jeannot, Y. Jgou, S. Lanteri, J. Leduc, N. Melab, G. Mornet, R. Namyst, P. Primet, B. Quetier, O. Richard, E.G. Talbi, T. Irena, "Grid'5000: a large scale and highly reconfigurable experimental grid testbed", International Journal of High Performance Computing Application 20 (4), (2006), pp. 481-494.
[13] J.a.N. Silva, L. Veiga, P. Ferreira, "Heuristic for resources allocation on utility computing infrastructures", in: MGC'08 Proceedings of the 6th International Workshop on Middleware for Grid Computing, ACM, New York, NY, USA, (2008), pp. 1–6.
[14] P.A. Dinda, D.R. O'Hallaron, "Host load prediction using linear models", Cluster Computing, (2001), pp. 265-280.
[15] R. Rojas, "Neural Networks: A Systematic Introduction", Springer-Verlag, (1996).
[16] P.A. Dinda, "Online prediction of the running time of tasks", Journal of Cluster Computing, (2002), pp. 225-236.
[17] T. V. Truong Duy, Y. Sato and Y. Inoguchi, "Performance Evaluation of a Green Scheduling Algorithm for Energy Savings in Cloud Computing", IEEE Computer, (2010), pp. 18-25.
[18] Y. Agarwal *et al*., "Somniloquy: Augmenting NetworkInterfaces to Reduce PC Energy Usage", Proc. Of NSDI, (2009), pp. 93-108.
[19] A. Huth, J. Cebula,"The Basics of Cloud Computing", US-CERT, (2011), pp. 1-4.
[20] A. Iosup, H. Li, M. Jan, S. Anoep, C. Dumitrescu, L. Wolters, D.H.J. Epema, "The Grid Workload Archive", Elsevier, (2007).
[21] A. Iosup, H. Li, C. Dumitrescu, L. Wolters, D.H.J. Epema, "The Grid Workload Format", (2006).
[22] R. Rojas, "Neural Networks: A Systematic Introduction", Springer, (1996).

