Advances in Intelligent Systems Research (AISR), volume 141
International Conference on Applied Mathematics, Modelling and Statistics Application (AMMSA 2017)

Research on Energy Consumption Modeling Method for Multi-Type Hybrid Applications

Fengyu Guo and Xiaoying Wang*
Department of Computer Technology and Applications, Qinghai University, Xining, Qinghai, China, 810016
*Corresponding author

***Abstract*—With the rapid development of cloud computing, a large number of datacenters were built. These datacenters provide excellent computing resource and service to people. However, they also bring huge energy consumption at the same time. Scientists all around the world try to solve this problem, for which the first step is to build an accurate power consumption model. Some related work aims to establish a model which only can be applied to a single-type application. Nevertheless, multi-type applications are often executed simultaneously in practice, and it is difficult to establish a multi-type application power consumption model. In this paper, we design a multi-type application power consumption model based on the hardware resource utilization ratio, recording the server power at different hardware resource utilization ratio. Using hardware resource utilization ratio as input and the power as output, we build the model by leveraging the BP neural network method. Finally, the accuracy of our method can reach 95%.**

Keywords-cloud computing; power modeling; hybrid applications

I. Introduction

In recent years, China's cloud computing industry developed rapidly. As an important IT infrastructure, datacenter also developed rapidly at the same time. In 2016, for example, China's datacenter industry grew by 40.4%, the market size is expected to reach 72 billion RMB [1]. The amount of datacenters in China has reached to above 400000, and the scale is also growing rapidly. The use of the TianHe NO.2 went into service in 2013, and by the end of 2016, the number of Tianhe No.2 's processors rise over 32000, and the number of processors of Shenwei Taihu Light which started in the same year reached 40960 [2]. However, the PUE (Power Utilization ratio Efficiency) of datacenters of China is greater than 2.2, which becomes a large gap with the international advanced level. According to the statistics data, the datacenter energy consumption in China occupies 1.5% of the total amount of the whole society [1]. Thus, how to reduce the energy consumption of the datacenters has become the academic subject which attracted widespread attention. Modeling energy consumption is the necessary basis for the research of energy consumption, since an accurate energy consumption model can provide guidance for the subsequent task scheduling. Scientists has made some achievements for the study of data center energy modeling, and the energy consumption of the datacenters has been a hot research topic. The current research on datacenter energy mainly focuses on single-type applications such as CPU-intensive or I/O-intensive ones, but in the actual work environment, hybrid-type applications are more common. Different types of applications lead to different utilization ratio of the server resources, and the use of different resources has a different impact on the overall power consumption of the system. In this paper, the energy consumption of the system is collected under the condition of different resource utilization ratio, and the energy consumption model is established to provide guidance for the subsequent task scheduling.

II. Related Work

Study on the server power consumption is a hot topic in recent years. Aiming to understand the server power consumption and how the load impact on the server power consumption, researchers use a series of methods to study the server. In order to learn the real-time power consumption of servers, they usually use two kinds of methods. 1) One is the direct measurement method, using power detection equipped special hardware to directly measure the power consumption and sum up the amount of multiple components to get the overall server power. The advantage of this method is the high accuracy, and the disadvantage is the high overhead and the limitation that the test must be carried out on specific hardware conditions, which is not universal. 2) The second method is non-invasive test, by which the usual practice is to test the hardware resource utilization ratio of server and server power, and then using mathematical modeling method to establish a model to forecast the power consumption of the server. This method has the advantages of low cost and high flexibility. The energy consumption models with different degrees of accuracy can be obtained by selecting different number of server resources as inputs. The disadvantage is that the model is complex and the model accuracy is low. Christoph Mobius et al. [3] provides a power consumption model based on CPU utilization ratio, and they provide three kinds of model which is the utilization ratio of CPU model, VM model and server model. The relative error of the model is between 0-15%, and the error is caused by the choice of Benchmark program in the modeling process. Basmadjian R et al. [4] divided a complex data center power modeling into four layers of independent model, i.e. called resource layer, server layer, storage layer and service layer respectively, and built an energy consumption model for the entire data center. The model is complex but credible. Rivoire S et al. [5] figured out a server model that can predict the average power, which does not consider the specific use of the server. The model has the advantages of simplicity and fast calculation, but for heterogeneous systems, this model is too rough and the prediction accuracy is too low. FanX et al. [6] proposed a linear power consumption model based on CPU utilization ratio. The main drawback of this model is that it is only suitable for CPU intensive tasks. The model proposed by Heath T et al. [7] is similar to that in the literature [6]. The difference is that the model considers the utilization ratio of CPU as well as the power consumption of the disk, which is more accurate than the model in [6].

Economou D et al [8] considered the CPU and disk utilization ratio and added a performance counter to the system, making the model very accurate. However, the performance counter technology is limited to several specifications of the CPU, and thus is not universal for heterogeneous systems. Chen Xi et al. [9] provides a power consumption model for multi-core scenarios, providing an online performance and power architecture that can quickly estimate the power consumption of the processor tasks.

III. APPLICATION SELECTION AND TEST METHODS

A. *Hybrid Type Applications*

The benchmark program is a standard test program, which is written in order to test the performance of a computer system. Benchmark programs can be divided into two categories according to the test object: one is the component test program, mainly used for the performance of the various components of the computer system testing, such as CPU, I/O and network, etc. The other is to test the performance of the entire computer system. In order to run the simulations at the state of mixed application and multiple types. Here, we use multi-type benchmark mixed execution mode for testing, ensuring the components of each part of system could be used, making the test state closer to the actual practice situation and the model has more practical significance [10][12]. The benchmark programs used in this experiment are shown in Table I.

TABLE I. STANDARD TEST PROGRAM

Application type	**Program name**
computation-intensive	ubench
memory-intensive	memtester
I/O-intensive	bonnie++,dd

B. *Resource Restriction Method*

In this paper, *CGroup* tools are used to control all kinds of resources. *CGroup* is the abbreviation of Control Groups, which is a mechanism which can restrict, record and isolate the process groups, such as CPU memory i/o, etc.. In order to control the amount of system resources used in the test program, it is necessary to configure the hardware resources of the system. Therefore, the *CGroup* mechanism is introduced to restrict the hardware resources used in the test program. For example, CPU resources can be limited by controlling the utilization ratio of CPU, I/O resources can be limited by limiting the data transmission speed of I/O, and the memory resources can be limited by limiting the available memory space.

IV. MODELING PROCEDURE

The power consumption model could be established by following steps, as shown in Fig.1 and Fig.2.

a) Executing various types of Benchmark standard test programs, recording the current power and CPU utilization ratio rate, memory utilization ratio, I/O transmission rate and network transmission rate respectively.

b) Using *CGroup* to limit the utilization ratio of various types of hardware resources, testing the server power at different hardware resource utilization ratio state.

c) According to the data from step a) and b) , using resource utilization ratio rate as input and the server power as output to build server power consumption model.

d) Verifying the correctness and accuracy of the model. Through the actual operation of some programs, monitoring server hardware resource utilization ratio and server power, comparing the result of power model with actual power analyzing the accuracy and correctness of the power consumption model.

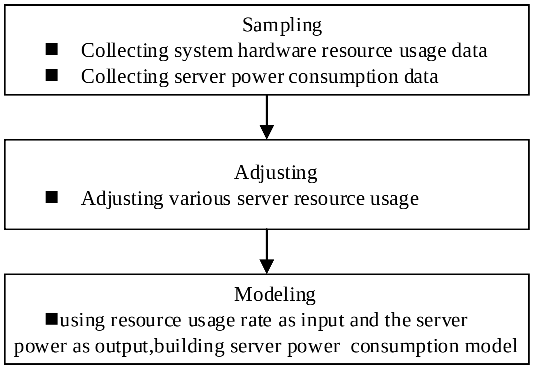

FIGURE I. BASIC PROCESS OF ENERGY MODELING

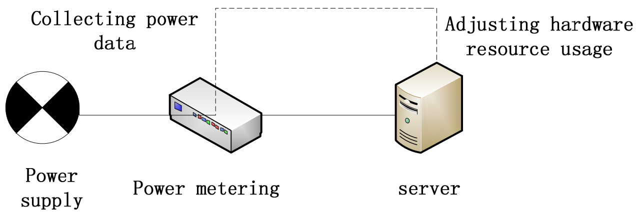

FIGURE II. SCHEMATIC DIAGRAM OF DATA SAMPLING

V. MODELING MECHANISM ANALYSIS

A. *Energy Consumption Feature Analysis*

In order to establish an accurate energy consumption model, we have to first analyze the energy consumption characteristics of the server. Server energy consumption can be divided into static energy consumption and dynamic energy consumption [11], where static energy consumption refers to the energy consumption when server is idle, which is usually a constant. Dynamic energy consumption refers to the energy consumption under the condition that the server is working, and the energy consumption is a dynamic value. Dynamic energy consumption contains CPU energy consumption, I/O energy consumption, memory energy consumption, network energy consumption, etc.. The total energy consumption of the server is the sum of static energy consumption and dynamic energy consumption.

Therefore, the total energy consumption of the server can be calculated as:

$$p_{server} = p_{base} + p_{cpu} + p_{mem} + p_{io} + p_{else} \quad (1)$$

where p_{server} refers to the total power consumption of the server, which can be measured by external power meter equipment; p_{base} refers to the static energy consumption of the server, i.e. the energy consumption of the server in standby mode which can also be measured by the power meter equipment; p_{cpu} refers to the server's CPU component power consumption, which generally cannot be measured with power meter; p_{mem} refers to server memory component energy consumption; p_{io} refers to the energy consumption of server disk components; p_{else} refers to the energy consumption of other components of the server. The energy consumption of the CPU, the energy consumption of the memory and the energy consumption of the disk I/O can account for more than 90% of the dynamic energy consumption in the dynamic energy consumption of the server, so the energy consumption modeling in this paper mainly considers the energy consumption of these three components, as follows:

$$p_{server} = p_{base} + p_{cpu} + p_{mem} + p_{io} \quad (2)$$

B. *Energy Consumption Modeling Based on Resource Utilization Ratio*

Traditional modeling methods are usually based on hardware performance counters or hardware utilization ratio [8], but modeling hardware performance counters need hardware support based on the relevant performance counters, and currently available hardware does not all support the performance counters. Therefore, in this paper we use the hardware utilization ratio for modeling, and when using *CGroup* to control the hardware resources, the utilization ratio of hardware resources can reflect the current use of hardware resources more directly.

In prior work, many of the energy consumption models are linear, which sums up the components energy consumption in accordance with their contribution ratio to the total power [11]. This kind of model is simple and easy to understand, but the accuracy is poor. The reason is that the server power consumption is not linearly related to the power consumption of each component. There is a certain coupling between the resource utilization ratio of different components. If the I/O rate rises to a certain extent, the memory utilization ratio will also increase. In this case, a simple linear model cannot fully describe the system power consumption. Therefore, the use of non-linear modeling method such as neural network is necessary [12]. In this paper, the energy consumption model based on multiple linear regression and the energy consumption model based on BP neural network are established, and the effects of these two models are compared.

VI. Power Modeling for Multi-type Hybrid Applications

A. *Linear Model Based on Hardware Utilization Ratio*

The principle of the linear model, as shown in Eq.(2), the input of the model is CPU power, memory power, the I/O power, the network power and the static power. The output is the power of entire server.

1) Power consumption model based on CPU utilization ratio: The power consumption model based on CPU utilization ratio is the most commonly-used power consumption model, as follows:

$$p_{cpu} = c_0 + c_1 * u_{cpu} \quad (3)$$

where c_0 is the static power constant of CPU, c_1 is the energy consumption contribution coefficient of CPU, u_{cpu} is the utilization ratio of CPU, and the value of c_0 and c_1 can be obtained by linear regression.

2) Power consumption model based on memory utilization ratio: When the server is running a CPU-intensive task, memory resources might also grow with the increase in CPU utilization ratio. Therefore, the energy consumption of the memory component contributes greatly to the overall energy consumption. The power consumption model mechanism based on memory utilization ratio is as follows:

$$p_{mem} = c_0 + c_1 * u_{mem} + c_2 * u_{cpu} \quad (4)$$

It can be seen from Eq.(4) that the memory power consumption is not only related to the utilization ratio ratio of memory, but also to the utilization ratio ratio of CPU, and there is a certain coupling relation between them.

3) Energy consumption model based on I/O utilization ratio: When the data intensive task is running on the server, the disk I/O rate has a greater impact on the overall power consumption. The I/O utilization ratio rate refers to the proportion of the current I/O transmission rate accounted for the largest I/O rate, and thus the modeling mechanism is as follows:

$$p_{io} = c_0 + c_1 * u_{io} + c_2 * u_{mem} \quad (5)$$

I/O related applications are usually used to deal with large amounts of data, which requires a lot of memory and storage, frequent I/O operations to read and write data. At the same time, requirements of CPU resources are less, most of the time CPU are waiting for the hard disk. Therefore, there is a coupling between the disk I/O speed and memory utilization ratio, which means that the utilization ratio of memory space might increase when the I/O rate increases, resulting in increased memory power consumption.

VII. ENERGY MODELING EXPERIMENT

A. Data Sampling Procedure

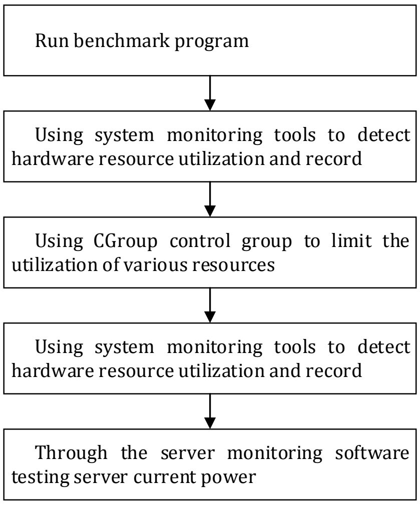

FIGURE III. DATA COLLECTION PROCEDURE

In order to test different server components, we need to run different types of benchmarks and the benchmark program running will usually take up a lot of related hardware resources. For example, running CPU-intensive benchmark will lead to CPU utilization ratio reaching 99% until the program ends. If you simply run this type of test program and record power consumption, the distribution of such data over the entire hardware resource utilization ratio interval is too concentrated, at around 99%, which is worthless for modeling and therefore it requires to control the utilization ratio of hardware resources, making the utilization ratio rate evenly distributed throughout the interval. In order to establish a power consumption model based on multiple hardware utilization ratio, it is necessary to run multiple types of benchmarks on the server at the same time and set multiple hardware resource utilization ratio ratios to record the power consumption of multiple resource utilization ratio combinations. The sample data set is as follows:

TABLE II. SAMPLE DATA SETS

Cpu utilization ratio ratio	**Memory utilization ratio ratio**	**I/O utilization ratio ratio**	**Server power (Watt)**
25%	15%	0	93
25%	40%	0	94
9%	98%	36%	90
75%	98%	81%	102
94%	98%	81%	104

We collected a total of 116 sets of mixed-type application of power consumption data, the data distribution is shown as follows:

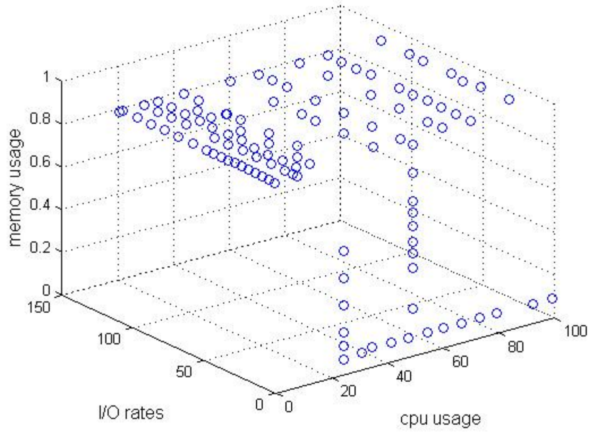

FIGURE IV. DISTRIBUTION OF CPU UTILIZATION RATIO,MEMORY UTILIZATION RATIO AND I/O RATES

The data in Fig.4 are evenly distributed in the CPU utilization ratio axis and the I/O ratio axis, but the memory utilization ratio axis is sparsely distributed, mainly due to the need to test the mixed type of applications. When I/O intensive applications are required a large number of disk read or write operations, reading and writing data need to take up a lot of memory space, resulting in higher memory utilization ratio, so most of the test data are distributed in the memory utilization ratio of the higher range.

B. Modeling Results Analysis

In order to verify the accuracy of multiple types of mixed application power models, we have established a single type of application power model and a mixed type model, and compare the two models with the actual power consumption of the server. Single-type application power model selects the typical power consumption model based on CPU utilization ratio. The research on this kind of model is more mature, the model is more accurate for the CPU-intensive scenario, but when multi-type applications are running on the server, including computing-intensive applications, data-intensive applications, and network-intensive applications, the model is less effective. In order to verify the accuracy of the model, the difference between the power model output and the actual power consumption is calculated and the difference is calculated as the percentage of the true value to measure the accuracy of the model. The accuracy formula is as follows:

$$Accuracy = 1 - \frac{p_{model} - p_{real}}{p_{real}} \quad (6)$$

1) Single power application modeling based on BP neural network: We use BP neural network method to model the compute-intensive application, using cpu utilization ratio as input and the server power as output, a total of 31 data is measured. the model output and the system measured power comparison as the picture shows:

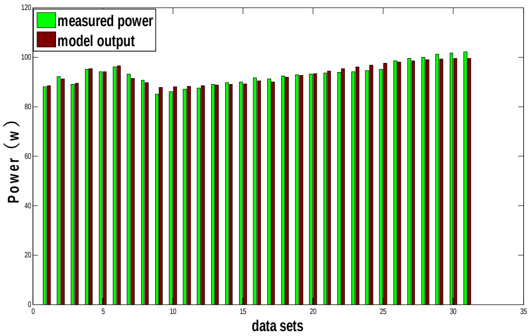

FIGURE V. FCOMPARISON OF SINGLE TYPE APPLICATION MODEL OUTPUT AND MEASURED DATATHE MODEL ERROR IS SHOWN IN THE FOLLOWING FIGURE:

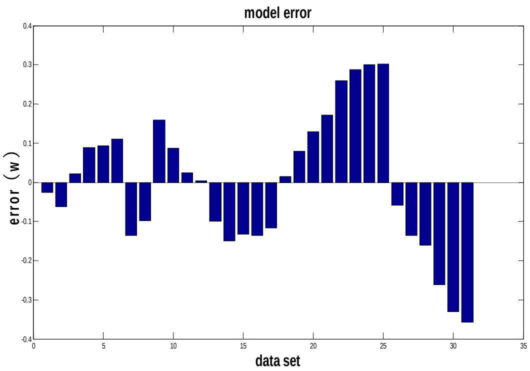

FIGURE VI. SINGLE-TYPE APPLICATION MODELING ERROR DIAGRAM

From the above figure we can see that the modeling effect of the model is good and the error is kept within ± 0.4W.

2) Power consumption modeling for multiple types of applications based on multiple linear regression: The modeling of a single-type application is more applicable when the server is running a single type of load, but in a real engineering environment, the server typically runs multiple types of applications, and in this case the single-type model will be limited and inaccurate. Hence, we need to build a hybrid-type power model, which adopts multiple linear regression and BP neural network method for modeling. The input parameters of the two models are CPU utilization ratio, memory utilization ratio and I/O utilization ratio. The model output is the system power consumption. A total of 116 sets of data are measured in the actual data sampling process, and the error analysis is as follows:

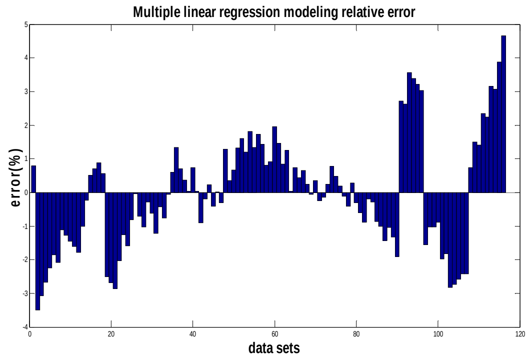

FIGURE VII. RELATIVE ERROR DIAGRAM OF MULTIPLE LINEAR REGRESSION MODELING

The energy consumption model based on multiple linear regression method can be described as follows:

$$p_{server} = (0.146,0.030,1.974\,)\begin{bmatrix} u_{cpu} \\ u_{mem} \\ u_{I/O} \end{bmatrix} + 87.690 \qquad (7)$$

3) Multi-type application power modeling based on BP neural network: Here we also select the above 116 sets of data training neural network model, set the number of training *f* to 50000 times, and then the modeling results are shown in Fig.8 and Fig.9.

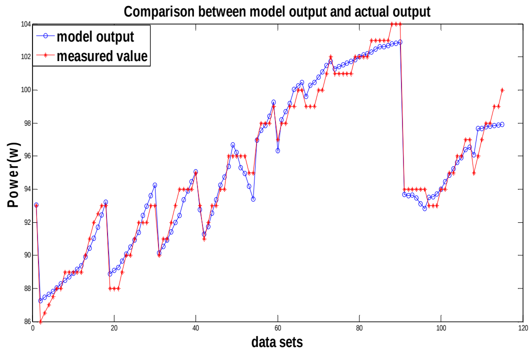

FIGURE VIII. COMPARISON OF ACTUAL OUTPUT AND OUTPUT OF MIXED TYPE APPLICATION MODEL BASED ON BP NEURAL NETWORK

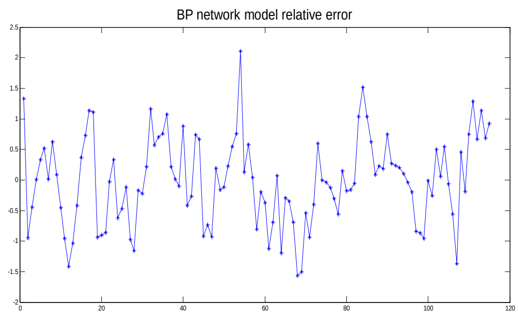

FIGURE IX. RELATIVE ERROR DIAGRAM OF POWER CONSUMPTION MODEL BASED ON BP NEURAL NETWORK

It can be seen from the above figure that the modeling error range based on BP neural network is kept within ± 0.2W and the relative error is kept within 2%.

By comparing the power consumption modeling based on multiple linear regressions and the power consumption modeling based on BP neural network, it can be seen that the neural network modeling has a good approximation to the nonlinear system, and the multiple linear regression model is too simple to describe the nonlinear system.

4) Comparison of multiple-type and single-type models: The multi-type application power model is more complex than the single-type application model, while the multi-type model can more accurately reflect the power consumption of the server. We tried to run a practical program to record the CPU utilization ratio, memory utilization ratio and I/O rate when the program in the running process and calculate these data through the above model, comparing the output of the three models with the measured server power, and the results are shown in Fig.10 and Fig.11.

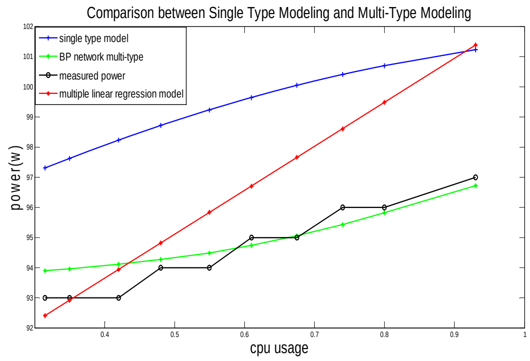

FIGURE X. COMPARISON BETWEEN DIFFERENT KINDS OF MODEL OUTPUT AND THE ACTUAL POWER WHEN THE CPU UTILIZATION RATIO INCREASED(I/O RATE: 0, MEMORY UTILIZATION RATIO: 8.8%)

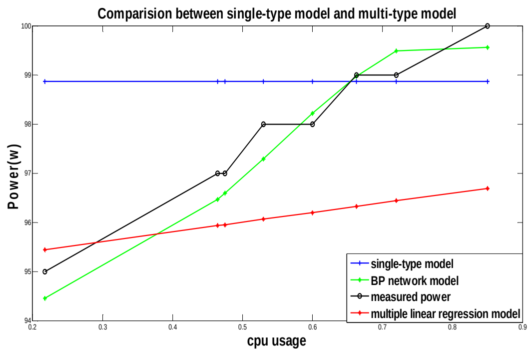

FIGURE XI. COMPARISON BETWEEN DIFFERENT KINDS OF MODEL OUTPUT AND THE ACTUAL POWER WHEN THE MEMORY UTILIZATION RATIO INCREASED (I/O RATE: 0 , CPU UTILIZATION RATIO :50%)

It can be seen from the above figure that the energy consumption model based on BP neural network can track the changes of server power consumption accurately with the change of server hardware resource utilization ratio, and the multivariate linear regression model is relatively poor to model the nonlinear system such as server power consumption. The power model of a single-type application works well under a single type of load, but when the server is running in a mixed-type scenario, multiple hardware resources of the system are used simultaneously, the single-type model take only one resource into account and thus the error is large. Power model based on multiple types of mixed applications can fully reflect the non-linear characteristics of power consumption system, and it is suitable for description for the server power consumption under hybrid type load condition.

VIII. CONCLUSION

In this paper, we run multiple kinds of CPU-intensive, memory-intensive, I/O-intensive tasks on an actual physical server, limiting the CPU utilization ratio, memory utilization ratio, and I/O rate of the server by the *CGroup* tool upon Linux O/S, recording the power consumption of the server under the combination of different resource utilization ratio rates, using the above three resource utilization ratio as the input and the server power as the output, and finally establish the multi-type server power model. By executing the program on the actual physical server to get actual hardware resource utilization ratio and power data, the hardware resource utilization ratio were brought into the multi-type application power model and a single-type application power model to get two sets of predictions value, comparing the two sets of predicted values with the measured power of the physical machine. It is found that the mixed-type application power model is more accurate and effective than the single-type application power model.

In this paper, we also found some drawbacks in the experimental process, e.g. in the process of data testing and recording the accuracy is affected by the monitoring software accuracy. In the future we will attempt to use high accurate sensors and monitoring software to collect data.

ACKNOWLEDGEMENTS

This paper is partially supported by The National Natural Science Foundation of China (No. 61363019, No.61563044 and No. 61640206), Open Research Fund Program of State key Laboratory of Hydroscience and Engineering (No.sklhse-2017-A-05), and National Natural Science Foundation of Qinghai Province (No. 2014-ZJ-718, No. 2015-ZJ-725).

REFERENCES

[1] Shen Hui, China's data center industry grew by nearly 40.4% in 2016 [N]. economic daily, 2017-01-11. (in Chinese)

[2] Jiang Fang, Cai Yugao, the national production of the world's fastest supercomputer “Shenwei-Taihu light " [N]. Xinhua news agency, 2016-06-20. (in Chinese)

[3] .Möbius, W. Dargie, A. Schill. Power Consumption Estimation Models for Processors, Virtual Machines, and Server s[J]. IEEE Transactions on Parallel & Distributed Systems, 2014, 25(6):1600-1614.

[4] Basmadjian R, Ali N, Niedermeier F, et al. A methodology to predict the power consumption of servers in data centres [C]// ACM SIGCOMM, International Conference on Energy-Efficient Computing and NETWORKING. ACM, 2011:1-10.

[5] Rivoire S, Ranganathan P, Kozyrakis C. A Comparison of High-Level Full-System Power Models. [C]// The Workshop on Power Aware

Computing and Systems, Hotpower 2008, December 7, 2008, San Diego, Ca, Usa, Proceedings. DBLP, 2008.

[6] Fan X, Weber W D, Barroso L A. Power provisioning for a warehouse-sized computer[C]// International Symposium on Computer Architecture. DBLP, 2007:13-23.

[7] Heath T, Diniz B, Carrera E V, et al. Energy conservation in heterogeneous server clusters [C]// ACM Sigplan Symposium on Principles and Practice of Parallel Programming, PPOPP 2005, June 15-17, 2005, Chicago, Il, Usa. DBLP, 2005:186-195.

[8] Economou D, Rivoire S, Kozyrakis C, et al. Full-system power analysis and modeling for server environments [J]. In Workshop on Modeling Benchmarking and Simulation (MOBS), 2006.

[9] Chen, Xi, Xu, et al. Performance and power modeling in a multi-programmed multi-core environment [J]. 2010:813-818.

[10] Wang Liang. Summa of Benchmark Performance Test [J]. Computer Engineering and Applications. 2006, 42(15):45-48. (in Chinese)

[11] Luo L, Wu WJ, Zhang F. Energy modeling based on cloud data center. Ruan Jian Xue Bao/Journal of Software, 2014,25(7):1371-1387 (in Chinese)

[12] Shi Yi, Dai Yuehua, Ren Jianbao,et al. Power Character Analyze for Applications on Multi-core Platform [J]. Computer& Digital Engineering, 2010, 38(8):38-41. (in Chinese)

[13] Li Yan Fei. power metering model and energyefficiency optimizing mechanism in cloudcomputing [D]. Beijing University of Posts and Telecommunications, 2012. (in Chinese)

[14] Xu Hao. research on neural networkbased virtual machine's powerprediction model [D]. Beijing University of Posts and Telecommunications,2015.(in Chinese)