Fine-grained energy consumption model of servers based on task characteristics in cloud data center

Citation:

DOI: http://www.dx.doi.org/10.1109/ACCESS.2017.2732458

© 2017, IEEE

Reproduced under the terms of the IEEE Open Access Publishing Agreement.

Downloaded from DRO:
http://hdl.handle.net/10536/DRO/DU:30104209
Fine-Grained Energy Consumption Model of Servers Based on Task Characteristics in Cloud Data Center

ZHOU ZHOU\textsuperscript{1,2}, JEMAL H. ABAWAJY\textsuperscript{3}, (Senior Member, IEEE), FANGMIN LI\textsuperscript{1}, ZHIGANG HU\textsuperscript{4}, MORSHED U. CHOWDHURY\textsuperscript{3}, ABDULHAMEED ALELAIWI\textsuperscript{3}, AND KEQIN LI\textsuperscript{6}, (Fellow, IEEE)

\textsuperscript{1}Department of Mathematics and Computer Science, Changsha University, Changsha 410083, China
\textsuperscript{2}Department of Computer Science, Hunan University, Changsha 410083, China
\textsuperscript{3}School of Information Technology, Deakin University, Geelong, VIC 3217, Australia
\textsuperscript{4}School of Software, Central South University, Changsha 410083, China
\textsuperscript{5}Software Engineering Department, College of Computer and Information Sciences, King Saud University, Riyadh 11451, Saudi Arabia
\textsuperscript{6}Department of Computer Science, State University of New York, New Paltz, NY 12561, USA

Corresponding author: Jemal H. Abawajy (jemal.abawajy@deakin.edu.au)

This work was supported in part by the National Natural Science Foundation of China under Grant 61572525, Grant 61373042, Grant 61602525, and Grant 61404213, in part by the China Scholarship Council, Deakin University, and in part by the Deanship of Scientific Research at King Saud University, Riyadh, Saudi Arabia, through the Research Group Project under Grant RGP-VPP-318.

\textbf{ABSTRACT} In this paper, we address the problem of accurately modeling the cloud data center energy consumption. As minimizing energy consumption has become a crucial issue for the efficient operation and management of cloud data centers, an energy consumption model plays an important role in cloud datacenter energy management and control. Moreover, such model is essential for guiding energy-aware algorithms, such as resource provisioning policies and virtual machine migration policies. To this end, we propose a holistic cloud data center energy consumption model that is based on the principal component analysis and regression methods. Unlike the exiting approaches that focus on single system component in the datacenter, the proposed approach takes into account the energy consumption of the processing unit, memory, disk, and network interface card as well as the application characteristics. The proposed approach is validated through extensive experiments with the SPECpower benchmark. The experimental results show that the proposed energy consumption model achieves more than 95% prediction accuracy.

\textbf{INDEX TERMS} Energy consumption model, energy consumption contribution, task characteristic, principal component analysis.

\textbf{I. INTRODUCTION}

CLOUD computing provides access to a large pool of shared computational resources as a service over Internet in an on-demand, self-service, automatically scalable and pay-per-use model [1], [2]. Although Cloud computing provides many benefits, the high energy consumption of cloud datacenter is a serious concern [3]. The global data center electricity consumption in 2013 is estimated to be more than 4.35 gigawatt with estimated annual growth rate of up to 15% [4], [5]. Moreover, it is reported that only 10-15% of the supplied electricity to the data center is consumed by servers in datacenters [6], [7]. The high Cloud datacenters energy consumption have received significant attention recently due to its (i) high operating costs, (ii) adverse effect on the environment, and (iii) significant impact on performance.

As Cloud datacentres energy consumption has been steadily increasing over the last few years, the minimization of cloud datacenter power and energy consumption has become a challenging problem. A variety of energy-aware algorithms and mechanisms have been proposed to manage and control energy consumption in Cloud datacentres. Energy consumption model plays an important role in Cloud datacenter energy management and control [8]. Thus, any practical approach for minimizing Cloud datacenter energy consumption requires an accurate modelling of the Cloud datacenter energy consumption. An energy consumption model is essential for guiding energy-aware algorithms such as resource provisioning policies and mechanisms such as virtual machine migration policies. Moreover, it affects the pricing mechanism which cloud service providers charge.
their customers. Therefore, it is necessary to propose an accurate energy consumption model to perform effective management and control.

Existing approaches [12], [16], [19] on energy management models in datacenters primarily focus on CPU energy consumption [32], while ignoring the energy consumption by other subsystems such as memory, disk and NIC subsystems. As CPU is only one of the critical resources in cloud datacenters, datacenter energy consumption minimization techniques should consider all resources contributing to the overall energy consumption at the same time. With Cloud datacenters using huge storage subsystems to store and process data and the increasing communication traffic seen by the datacenters make the disk and NIC subsystems significant contributors to the energy consumption of the datacenters. Therefore, in addition to the energy consumed by CPU and memory subsystems, the energy consumed by the disk and NIC subsystems should be considered in building the energy consumption model for Cloud datacenters. Furthermore, existing approaches do not consider application characteristics when modeling energy consumption model for Cloud datacenters. The fact that different applications impose different resource requirements, considering application characteristics in the development of the model also becomes a primary concern.

In this paper, we propose a holistic Cloud datacenter energy consumption model that is based on the Principal Component Analysis (PCA) and regression methods. Unlike the existing approaches that focus on a single system components in the datacentre, the proposed approach takes into account the energy consumption of the processing unit, memory, disk and NIC (Network Interface Card) as well as the application characteristics. The experimental results of the proposed approach show that the proposed energy consumption model achieves more than 95% prediction accuracy. The main contributions of the paper are summarized as follows.

- A novel holistic Cloud data center energy consumption model that considers CPU, memory, disk and NIC subsystems, as well as the application characteristics (CPU intensive task, transactional web task, and I/O intensive task);
- Principal Component Analysis (PCA) and regression methods are used to analysis each subsystem parameter’s contribution accounting for total energy consumption
- Extensive experimental analysis to validate the proposed model using widely adopted benchmarks to evaluate the power and performance characteristics of servers [19].
- Comparison of the proposed model with three baseline energy consumption models, the Ramon Model [12], Linear Model [16] and Cubic Model [19].

The rest of the paper is organized as follows. In Section 2, the related works are discussed. Section 3 describes the methodology we used to develop the model. Section 4 presents the feature extraction and selection while the energy consumption modeling is described in Section 5. The performance analysis, results and discussion are discussed in Section 6 and Section 7, respectively. Section 8 presents the conclusion remarks.

II. RELATED WORK

The need for managing energy consumption level have become an important is in various domains [23], [24], [33]–[36]. Generally speaking, computational resources (e.g., CPU, memory, disk, and networking) and cooling system such as air conditioning equipment are the main energy consumption sources in datacenters. There are many energy-aware algorithms such as resource provisioning policies and virtual machine migration policies that aim to minimize energy consumption of Cloud datacenter. For example, a three-threshold energy-saving algorithm based on the empirical power model is proposed in [23] and [24]. Beloglazov and Buyya [25] explored the virtual machine migration based on an empirical power model. This power model can be obtained through recording energy consumption and CPU utilization at different load level. An approach that tracks per-VM power consumption is proposed in [26]. These energy-aware algorithms generally depend on the underlying power models used to develop them. Therefore, an accurate power model is the prerequisite to achieve the fine-grained power control and management in Cloud datacenter.

Prior works on power modeling focus on three main aspects: (i) performance-monitor-counter (PMC) based models [9]–[15]; (ii) resource utilization based model [16]–[22]; and (iii) their usage to guide energy-aware algorithms [23]–[26]. The PMC-based approach have three main steps. In the first step, events related to hardware units such as CPU, memory, disk, and NIC are monitored. In the second step, the events are analysed and those events that are related to the PMC set are screened out. In the final step, the energy consumption model is built based on the relationship between the PMC events and energy consumption by the system components.

Min et al. [9] proposed a surrogate model that is based on the PMC method. The model can sustain the absolute estimation error of 5.32 percent when running the SPEC benchmark. In addition, the authors validated the nonlinear relationship between the server energy consumption and CPU utilization. In [10] and [11], the authors used the microprocessor performance counters to account for the entire system power consumption. Although the approach is promising, the power model did not solve the high relative error problem. The PMC-based model proposed by Bertran et al. [12] utilized the CPU and memory models for the virtualized environments. The analysis confirmed that: (1) virtual machines (VMs) assigned the same amount of CPU cycles do not consume equal amount of energy; (2) PMC-based method can be used in virtualized environments; (3) DVFS (Dynamic Voltage and Frequency Scaling) method does not affect the accuracy of the power model. The methodology proposed in this paper is promising, albeit it did not consider the consumption of disk and NIC.
In [13] and [14], the authors leveraged PMC method to build real-time power model. In order to accurately measure the energy consumption of a virtual machine, authors in [15] proposed a new virtual machine power model based on a concept called a ‘relative PMC’. Based on the power model, the authors proposed a virtual machine scheduling algorithm to reduce the energy consumption and minimize SLA (Service Level Agreement) violations. However, the approach is complicated as it collects too many events leading to high overheads. Furthermore, the approach is not suitable to extend to other servers or VMs in a data center.

The resource utilization based approaches leverage the resource utilization of a server (such as CPU utilization, memory utilization, and so on) to construct energy consumption model. Garg et al. [16] proposed an approach based on the CPU utilization. This model does not reflect the true energy consumption in data centers [18]. Beloglazov et al. [17] leveraged the maximum power consumption, idle power consumption and CPU utilization of a server to build power model. Hsu et al. [18] proposed an exponential model based on CPU intensive tasks. Zhang et al. [19] argue that the relationship between the energy consumption and the CPU utilization is not linear and instead it is a cubic. Thus, the authors proposed a modified power model named Cubic Model in order to improve the accuracy of the power model. The model proposed in [20] and [21] estimates the energy consumption of the system component (such as CPU) and then build a linear power model based on statistics. The E-mc² framework [22] models the requirements of energy consumption in Cloud computing systems. This framework is easy to perform although the accuracy of power model should be further improved.

There are many energy-aware algorithms such as resource provisioning policies and virtual machine migration policies that aim to minimize energy consumption of Cloud datacenter. For example, a three-threshold energy-saving algorithm based on the empirical power model is proposed in [23] and [24]. Beloglazov and Buyya [25] explored the virtual machine migration based on an empirical power model. This power model can be obtained through recording energy consumption and CPU utilization at different load level. An approach that tracks per-VM power consumption is proposed in [26]. These energy-aware algorithms generally depend on the underlying power models used to develop them. Therefore, an accurate power model is the prerequisite to achieve the fine-grained power control and management in Cloud datacenter.

In summary, although both the disk and NIC subsystems consume considerable energy as compared to CPU and memory subsystems, they are generally ignored in the development of the model. Moreover, different application characteristics lead to different energy consumption. Thus, an accurate Cloud datacenter energy consumption model must consider not only the CPU, memory, disk and NIC subsystems but also the application characteristics. By accounting for CPU, memory, disk and NIC subsystems contribution to the total energy consumption as well as the application characteristics, our approach tackles the shortcomings of the exiting models.

III. METHODOLOGY

Fig. 1 shows the general steps used to develop the proposed energy consumption model. The methodology consists of the feature extraction, feature selection, modelling and
evaluation steps. The feature extraction step is responsible for collecting features of the resources and applications relevant to the energy consumption modeling. This step can be performed by using either resource utilization based method or PMC-based method. Some features extracted in this step may be related to power model while others may not be related to power model.

The feature selection step is responsible for finding good feature representation. This step can be accomplished by deploying approaches such as Correlation Matrix (CM) or Principal Component Analysis (PCA). The power model is then built using the subset features returned from the feature selection step. In this paper, we will use regression methods to build the model. An effective power model is characterized by its accuracy, representativeness, and extensibility. Finally, in the ‘evaluate’ step, the accuracy of the model is assessed in order to ensure its effectiveness.

IV. FEATURE EXTRACTION AND SELECTION

In this section, we discussed the steps performed to produce the collection of features used to build the energy consumption model.

A. FEATURE EXTRACTION

In building a proper power model, it is necessary to include appropriate subsystem parameters related to energy consumption. Let $P_{\text{system}}$ be the total power that can be consumed by a server in a datacenter. Therefore, $P_{\text{system}}$ can be modeled with the following equation:

$$P_{\text{system}} = P_{\text{CPU}} + P_{\text{memory}} + P_{\text{disk}} + P_{\text{network}} + \sigma \quad (1)$$

where $P_{\text{CPU}}$, $P_{\text{memory}}$, $P_{\text{disk}}$, and $P_{\text{network}}$ are the power of CPU, memory, disk, and NIC, respectively. The parameter $\sigma$ can be considered as a constant and represents the power of other subcomponents of a system excluding CPU, memory, disk, and NIC. The parameter $P_{\text{CPU}}$ can be modeled with the following equation [22]:

$$P_{\text{CPU}} = (P_{\text{max}} - P_{\text{idle}}) \times U + P_{\text{idle}} \quad (2)$$

where $P_{\text{max}}$ represents the maximum power of the server, $P_{\text{idle}}$ is the power consumed when the server is idle while $U$ denotes the CPU utilization of the server. As the value of $P_{\text{CPU}}$ is related to $U$, we choose parameter “Processor Time” as energy-consumption representative of the CPU. Note that the “Processor Time” refers to the percentage of an elapsed time that the processor spends executing a non-idle thread. We can monitor the value of “Processor Time” to get the CPU utilization.

The parameter $P_{\text{memory}}$, it can be modeled as follows [22]:

$$P_{\text{memory}} = P_{\text{PRE}} + P_{\text{ACT}} + P_{\text{RD}} + P_{\text{WR}} + P_{\text{REF}} \quad (3)$$

where $P_{\text{PRE}}$, $P_{\text{ACT}}$, $P_{\text{RD}}$, $P_{\text{WR}}$, and $P_{\text{REF}}$ are the power of pre-charge ($P_{\text{PRE}}$), activate ($P_{\text{ACT}}$), read ($P_{\text{RD}}$), write ($P_{\text{WR}}$) and refresh ($P_{\text{REF}}$), respectively. As the value of $P_{\text{memory}}$ is associated with writing and reading, we choose “Memory Used” and “Page Fault/Sec” parameters as energy consumption representative of memory subsystem. “Memory Used” and “Page Fault/Sec” represent the memory utilization and an average number of error pages per second, respectively.

For the parameter $P_{\text{disk}}$, it could be modeled with the following equation [22]:

$$P_{\text{disk}} = P_{\text{READ}} + P_{\text{WRITE}} + P_{\text{IDLE}} \quad (4)$$

where $P_{\text{READ}}$, $P_{\text{WRITE}}$, and $P_{\text{IDLE}}$ represent the power needed for reading, writing and remain idle, respectively. We select parameters “Disk time” and “Disk Bytes/Sec” as the energy-consumption representative of the disk. “Disk time” is the percentage of elapsed time that the selected disk drive was busy servicing the read or write requests. The “Disk Bytes/Sec” refers to the total number of bytes sent to the disk (write) and retrieved from the disk (read) over a period of one second.

As for the parameter of $P_{\text{network}}$, it can be modeled as [22]:

$$P_{\text{network}} = C_0 + C_1 \times \frac{S}{B} \quad (5)$$

where parameters $C_0$ and $C_1$ can be considered as constants, parameter $S$ is the file size in MB; parameter $B$ is the bandwidth in MB/s. We choose parameters “Bytes Total/Sec” and “Current Bandwidth” as energy-consumption representative of the NIC. “Bytes Total/Sec” is the rate at which the network adapter is processing data bytes; “Current Bandwidth” is the bandwidth.

<table>
<thead>
<tr>
<th>Component name</th>
<th>Parameter</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPU architecture</td>
<td>2 × (intel)Xeon E5-2620 Six-Core</td>
</tr>
<tr>
<td>CPU frequency</td>
<td>12×2.0 GHZ</td>
</tr>
<tr>
<td>Level 1 cache</td>
<td>6 × 32 KB</td>
</tr>
<tr>
<td>Level 2 cache</td>
<td>6 × 256 KB</td>
</tr>
<tr>
<td>Level 3 cache</td>
<td>15 MB</td>
</tr>
<tr>
<td>Memory size</td>
<td>20 GB DDR3</td>
</tr>
<tr>
<td>Disk size</td>
<td>2×1TB</td>
</tr>
<tr>
<td>NIC</td>
<td>Intel quad-port Gigabit network adapter</td>
</tr>
</tbody>
</table>

B. FEATURE SELECTION

In this section, we discuss the process used to select a subset of relevant features from those extracted in the preceding section for use in building power model. In order to determine the subset features, we use a Principal Component Analysis (PCA) [28]. To select a subset features, we deployed three application domains: CPU-intensive applications, Transactional Web and I/O intensive applications on DELL PowerEdge R720 with the configuration shown in Table 1. The SPEC CPU2006 benchmark [27] is a typical example of CPU-intensive task, and it includes “401.bzip2”, “403.gcc”, “429.mcf”, “453.povray” and “450.soplex” subsets. For the transactional web application, we used the HP LoadRunner [29], which is a typical transactional web application. For I/O-intensive application, we used Iozone dataset [30], which is a typical I/O intensive task.
Table 2 shows the feature values (Processor Time, Memory Used, Page Fault/Sec, Disk Time, Disk Bytes Total/Sec, Bytes Total/Sec, and Current Bandwidth) and corresponding energy consumption under three application domains (CPU-intensive applications, Transactional Web and I/O intensive applications).

For example, regarding CPU Intensive application, when “Processor Time” = 4.23%, “Memory Used” = 4.47%, “Page Fault/Sec” = 512.78, “Disk Time” = 0.66, “Disk Bytes/Sec” = 4102.28, “Bytes Total/Sec” = 562.00, and “Current Bandwidth” = 9.22 × 10^18, the total energy consumption is 122.49 W. Similarly, for Transactional Web application, when “Processor Time” = 6.90%, “Memory Used” = 4.29%, “Page Fault/Sec” = 28192.04, “Disk Time” = 2.86, “Disk Bytes/Sec” = 689229.22, “Bytes Total/Sec” = 64.13, and “Current Bandwidth” = 9.22 × 10^18, the total energy consumption is 107.00 W.

How these features influence the energy consumption? Which feature is related to energy consumption? Which feature is not related to energy consumption? To solve these problems, we make a Principal Component Analysis (PCA) [28] for factors’ contribution, and the each factors’ contribution is listed in Table 3.

As shown in Table 3, the top three features (i.e., Processor Time, Disk Bytes/Sec and Disk Time) contribute significantly while Page Fault/Sec, “Memory Used” and “Bytes Total/Sec” contribute very little while “Current Bandwidth” does not contribute at all. This is because a CPU-intensive application also called compute-intensive task requires a lot of processing power as compared to other resources.

The transactional web application is similar to CPU-intensive application regarding the contribution of the features. As shown in Table 3, the top three features (i.e., Processor Time, Disk Bytes/Sec and Disk Time) contribute

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Application Contribution (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CPU Intensive</td>
</tr>
<tr>
<td>Processor Time</td>
<td>62</td>
</tr>
<tr>
<td>Disk Bytes/Sec</td>
<td>19</td>
</tr>
<tr>
<td>Disk Time</td>
<td>14</td>
</tr>
<tr>
<td>Page Fault/Sec</td>
<td>4</td>
</tr>
<tr>
<td>Memory Used</td>
<td>1</td>
</tr>
<tr>
<td>Bytes Total/Sec</td>
<td>0</td>
</tr>
<tr>
<td>Current Bandwidth</td>
<td>0</td>
</tr>
</tbody>
</table>
significantly while Page Fault/Sec and "Memory Used" contribute very little with and "Bytes Total/Sec" and "Current Bandwidth" contributing nothing. The results for the I/O-intensive application is similar to the other two applications with the top three features (i.e., Processor Time, Disk Bytes/Sec and Disk Time) contribute significantly while Page Fault/Sec and "Memory Used" contribute very little with and "Bytes Total/Sec" and "Current Bandwidth" contributing nothing.

Therefore, we choose the non-zero features (i.e., Processor Time, Disk Bytes/Sec, Disk Time, Page Fault/Sec, and Memory Used) to build the energy consumption model.

V. ENERGY CONSUMPTION MODELLING

We used the subset features from the previous section and a regression method to construct the energy consumption model. In this section, we use four modeling methods namely a linear regression, a power regression, an exponential regression, and a polynomial regression in combination with the representative parameter (see Section 3.2) and EViews 8.0 [31] software to build the energy consumption models.

In the following subsection, we use y, x1, x2, x3, x4, x5, and x6 to represent the features used for use in building power model as shown in Table 4.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>y</td>
<td>Energy consumption</td>
</tr>
<tr>
<td>x1</td>
<td>Processor Time</td>
</tr>
<tr>
<td>x2</td>
<td>Disk Bytes/Sec</td>
</tr>
<tr>
<td>x3</td>
<td>Disk Time</td>
</tr>
<tr>
<td>x4</td>
<td>Page Fault/Sec</td>
</tr>
<tr>
<td>x5</td>
<td>Memory Used</td>
</tr>
<tr>
<td>x6</td>
<td>Bytes Total/Sec</td>
</tr>
</tbody>
</table>

A. MULTIVARIATE LINEAR REGRESSION MODEL

The multivariate linear regression model is defined as follows:

\[ y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_m x_m + \epsilon \] (6)

where y is the real energy consumption, \(\beta_0, \beta_1, \beta_2, \ldots, \beta_m\) are the regression coefficients and \(\epsilon\) represents a stochastic error. For CPU2006 benchmark [27], which is CPU intensive task, the energy consumption model of linear regression is given in Eq. (7).

\[
y = 102.9169 + 1.967511 x_1 - 1.37 \times 10^{-5} x_2 \\
- 0.001408 x_3 + 1.29 \times 10^{-5} x_4 + 2.528892 x_5 
\] (7)

For the transactional web task LoadRunner [29], the energy consumption model based on the linear regression is given in Eq. (8):

\[
y = -869.7 - 14.18 x_1 - 8.68 \times 10^{-5} x_2 + 22.92 x_3 \\
+ 0.002449 x_4 + 234.2339 x_5 - 0.067755 x_6 
\] (8)

For I/O intensive task Iozone dataset [30], the energy consumption model based on the linear regression is given in Eq. (9).

\[
y = 111.5943 + 9.173805 x_1 - 1.51 \times 10^{-6} x_2 \\
+ 2.037900 x_3 - 0.000781 x_4 - 19.46270 x_5 
\] (9)

B. POWER REGRESSION MODEL

The power regression model is represented as follows:

\[
y = \beta_0 x_1^{b_1} x_2^{b_2} \cdots x_m^{b_m} + \epsilon \] (10)

where y is the real energy consumption, \(\beta_0, \beta_1, \beta_2, \ldots, \beta_m\) are the regression coefficients and \(\epsilon\) represents a stochastic error. The energy consumption model based on power regression approach for CPU2006 benchmark [27] is shown in Eq. (11).

\[
y = e^{4.840775 \times (x_1^{0.219818} \times (x_2^{-0.056527} \times (x_3^{0.067893} \\
\times (x_4)^{0.000708} \times (x_5)^{0.096609}))} 
\] (11)

For the transactional web task LoadRunner [29], the energy consumption model based on the power regression is given in Eq. (12):

\[
y = e^{-8.920533 \times (x_1^{0.198811} \times (x_2^{-0.008926} \times (x_3^{-0.028378} \\
\times (x_4)^{-0.016527} \times (x_5)^{-2.920025} \times (x_6)^{-0.014455})} 
\] (12)

For I/O intensive task Iozone dataset [30], the energy consumption model based on the power regression is given in Eq. (13).

\[
y = e^{5.626638 \times (x_1^{0.038072} \times (x_2^{-0.000339} \times (x_3^{-0.054210} \\
\times (x_4)^{0.010081} \times (x_5)^{-0.751834})} 
\] (13)

C. EXPONENTIAL REGRESSION MODEL

The exponential regression model is defined as follows:

\[
y = \beta_0 e^{\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_m x_m} + \epsilon \] (14)

where y is the real energy consumption, \(\beta_0, \beta_1, \beta_2, \ldots, \beta_m\) are the regression coefficients and \(\epsilon\) represents a stochastic error. The energy consumption model based on exponential regression approach for CPU2006 benchmark [27] is shown in Eq. (15).

\[
y = e^{4.614954 \times (0.016 x_1 - 1.35 \times 10^{-7} x_2 + 0.00085 \\
\times x_3 + 1.56 \times 10^{-7} x_4 + 0.022 x_5)} 
\] (15)

For the transactional web task LoadRunner [29], the energy consumption model based on the exponential regression is given in Eq. (16):

\[
y = e^{-4.67} \times e^{(-0.13 x_1 - 8.18 \times 10^{-7} x_2 + 0.217 x_3 \\
\times + 2.30 \times 10^{-7} x_4 + 2.23 x_5 - 0.0006 x_6)} 
\] (16)
Z. Zhou et al.: Fine-Grained Energy Consumption Model of Servers Based on Task Characteristics

For I/O intensive task Iozone dataset [30], the energy consumption model based on the exponential regression is given in Eq. (17).

\[
y = e^{4.741696 \times e^{-6.53 \times 10^{-6} \times x_4 - 0.1736 \times x_5}} (17)
\]

**D. POLYNOMIAL REGRESSION MODEL**

As for polynomial regression, the regression model is defined as follows:

\[
y = \beta_0 + \beta_1 (x_1)^2 + \beta_2 x_2 + \cdots + \beta_m x_m + \epsilon \quad (18)
\]

where \(y\) is the real energy consumption, \(\beta_0, \beta_1, \beta_2, \ldots, \beta_m\) are the regression coefficients and \(\epsilon\) represents a stochastic error. The energy consumption model for CPU2006 benchmark [27] based on the polynomial regression model is shown in Eq. (19).

\[
y = 111.4598 + 0.151606 \times (x_1)^2 - 1.83 \times 10^{-5} \times x_2 \\
+ 0.420755 \times x_3 + 1.08 \times 10^{-7} \times x_4 + 1.816320 \times x_5 \\
- 0.0797 \times x_6 \quad (19)
\]

For the transactional web task LoadRunner [29], the energy consumption model based on the polynomial regression is given in Eq. (20):

\[
y = -334.1569 - 0.115852 \times (x_1)^2 - 6.70 \times 10^{-5} \times x_2 \\
+ 16.867 \times x_3 + 0.000406 \times x_4 + 102.1 \times x_5 \\
- 0.0797 \times x_6 \quad (20)
\]
For I/O intensive task Iozone dataset [30], the energy consumption model based on the polynomial regression is given in Eq. (21).

\[
y = 78.99736 + 1.459156 \times (x_1)^2 - 1.51 \times 10^{-6} \times x_2 \\
+ 2.667544 \times x_3 - 0.000969 \times x_4 - 12.17560 \times x_5
\]  

(21)

VI. PERFORMANCE ANALYSIS
In this section, we discuss the experimental analysis of the energy consumption model proposed in this paper. All experiments were run on DELL PowerEdge R720 with 2.0 GHZ (2× Six-core), 20 GB RAM and 2 TB disk storage. The parameters and values for the server configuration is given in Table 1. The benchmark for CPU intensive task is SPEC CPU2006 [27] (it includes “401.bzip2”, “403.gcc”, “429.mcf”, “453.povray” and “450.soplex” subsets), and the benchmark for transactional web task is HP LoadRunner [29], and for I/O intensive task is Iozone [30].

To evaluate the accuracy of energy consumption model, we define the following metric:

\[
\text{Power error} = \frac{\text{Power predict} - \text{Power true}}{\text{Power true}}
\]  

(22)

where \(\text{Power predict}\) is the predicted value of the energy consumption by the model, \(\text{Power true}\) is the true value of the energy consumption, and \(\text{Power error}\) is the relative error of the energy consumption. The true value of the energy consumption is measured using the Power Bay-SSM tool.

We compared the proposed approach with three baseline approaches: the Ramon Model [12], the Linear Model [16] and the Cubic Model [19]. The Ramon Model focuses on CPU and memory, while the Linear Model and the Cubic Model focus on CPU alone.

VII. RESULTS AND DISCUSSION
In this section, we discuss the experimental results for the seven models (i.e., linear regression, power regression, exponential regression, polynomial regression, Ramon Model [12], Linear Model [16] and Cubic Model [19]) under the execution of various applications.

A. ANALYSIS FOR CPU INTENSIVE TASK
Fig.2 show the energy consumption of the seven models while Fig.3 shows the relative error generated by the seven models for the CPU intensive task CPU2006 [27].

Compared to Ramon Model [12], Linear Model [16] and Cubic Model [19], the four modeling methods (linear regression, power regression, exponential regression, and polynomial regression) perform slightly better. The reason is two folds. Firstly, the four modeling methods consider all components related to energy consumption such CPU, memory, disk, and NIC during the construction of power model, while Ramon Mode only takes into account the consumption of the CPU and memory, and Linear Model and Cubic Model only consider consumption of CPU. Secondly, the four modeling methods leverage PCA method to improve the accuracy of power model based on the application characteristics.

B. ANALYSIS FOR TRANSACTIONAL WEB TASK
For the transactional web task LoadRunner [29], Fig.4 and Fig.5 show the energy consumption and relative error under the transactional web task, respectively.

The four modeling methods (linear regression, power regression, exponential regression, and polynomial regression) perform better than Ramon Model [12], Linear Model [16] and Cubic Model [19]. The reason includes two aspects. On the one hand, the characteristic of the transactional web task determines that this task visits memory and network frequently. Therefore, only considering CPU or memory factor is not enough to build power model. Conversely, the four modeling methods not only consider CPU and memory factors, but also disk and NIC factors. On the other hand, the four modeling methods utilize PCA method to improve the accuracy of power model based on task characteristics. Fig.4 and Fig.5 also
illustrate that Ramon Model is better than Linear Model and Cubic Model, this can be explained by the fact that Ramon Model considers both CPU and memory factors, while Linear Model and Cubic Model only consider CPU factor.

**C. ANALYSIS FOR I/O INTENSIVE TASK**

Fig.6 and Fig.7 show the energy consumption and the relative error for the seven models under the I/O intensive task, respectively.

In comparison with the Ramon Model [12], the Linear Model [16] and the Cubic Model [19], the proposed four models (i.e. linear regression, power regression, exponential regression, and polynomial regression) improve more than 2% accuracy of the energy consumption model. The reason can be explained by the fact that the four modeling methods consider task characteristics and all factors related to energy consumption (such as CPU, memory, disk, and NIC) during the construction of power modeling. Fig.6 and Fig.7 also show that Ramon Model has better performance than Linear Model and Cubic Model, this reason is that Ramon Model considers both CPU and memory factors, while Linear Model and Cubic Model only consider CPU factor.

**D. THE COMPARISON OF THE FOUR MODELING METHODS**

Fig.8 and Fig.9 respectively show the comparison for the four modeling methods (linear regression, power regression, exponential regression, and polynomial regression) in terms of energy consumption and relative error. Fig.8 and Fig.9 reveal that, no matter the task belongs to CPU intensive task or transactional web task or I/O intensive task, power regression leads to the highest accuracy of power model in most cases. Therefore, we recommend using power regression to build power model in further research.
VIII. CONCLUSION
This paper proposed an energy consumption model for data-centres based on application characteristics (such as CPU intensive task, transactional web task, and I/O intensive task) and various subsystems (i.e., CPU, memory, disk and NIC). Experimental results show that: (1) during the process of building energy consumption model, considering all components related to energy consumption such as CPU, memory, disk, and NIC is more effective than only consider CPU and memory; (2) taking into account the task characteristics (CPU intensive task, transactional web task, and I/O intensive task) provides a better performance than only focusing on CPU intensive task during the construction of the power model. Moreover, the energy consumption model proposed in this paper is more accurate than the existing ones. The proposed energy consumption model can be extended to other servers in data centers, so as to guide the energy-saving algorithm to save energy-consumption.

ACKNOWLEDGEMENT
This work was done while the first author had a visiting position at the School of Information Technology, Deakin University, Australia. The help of Maliha Omar is also sincerely appreciated.

REFERENCES
The Standard Performance Evaluation Corporation. SPEC CPU2006


ZHOU ZHOU received the Ph.D. degree from Central South University, Changsha, China, in 2017, majoring in computer science. He is currently a Lecturer with Changsha University. He also holds a post-doctoral position at Hunan University for two years (2017–2019). His research interests include cloud computing, energy consumption model, energy-efficient resource management, and virtual machine deployment.

JEMAL H. ABAWAJY (SM’11) is currently a Full Professor with the Faculty of Science, Engineering and Built Environment, Deakin University, Australia. He has authored/co-authored over 250 refereed articles and supervised numerous Ph.D. students to completion. He has also served on the editorial board of numerous international journals including the IEEE TRANSACTIONS ON CLOUD COMPUTING. He has delivered over 50 keynote and seminars worldwide and has been involved in the organization of over 30 international conferences in various capacity including chair and general co-chair.

FANGMIN LI received the Ph.D. degree majoring in computer science from Zhejiang University, Hangzhou, China, in 2001. He is currently a Professor with Changsha University. He also holds ten Chinese patents. He has authored several books on embedded systems and over 30 academic papers in cloud computing and wireless networks. His current research interests include cloud computing, energy-efficient resource management, and wireless communications and networks.

ZHIGANG HU received the M.S. and Ph.D. degrees from Central South University in 1988 and 2002, respectively. He is currently a Professor with the School of Software, Central South University. His research interests include high performance computing, cloud computing, energy-efficient resource management, and virtual machine deployment.

MORSHED U. CHOWDHURY received the Ph.D. degree in computing from Monash University, Australia, in 1999. He is currently a Senior Lecturer with the School of Information Technology, Deakin University, Burwood, Australia. He has over 12 years of industry experiences and has authored over 100 research papers. He has organized a number of international conferences and served as a member of program committees of several international conferences. He has also acted as reviewer of many IEEE and Elsevier journal papers.

ABDULHAMEED ALELAIWI received the Ph.D. degree in software engineering from the College of Engineering, Florida Institute of Technology-Melbourne, USA. He is currently a Professor with School of Software Engineering Department, College of Computer and Information Sciences, King Saud University, Riyadh, Saudi Arabia. He has authored/co-authored many publications, including refereed IEEE/ACM/Springer journals, conference papers, books, and book chapters. His research interest includes software testing analysis and design, cloud computing, and multimedia.

KEQIN LI (F’14) is currently a SUNY Distinguished Professor of computer science. He is also an Intellectual Ventures Endowed Visiting Chair Professor with Tsinghua University, China. He has over 520 refereed research publications. His research interests mainly include the design and analysis of algorithms, parallel and distributed computing, and computer networking. He is currently or has served on the Editorial Board of the IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, the IEEE TRANSACTIONS ON COMPUTERS, THE IEEE TRANSACTIONS ON SERVICES COMPUTING, THE IEEE TRANSACTIONS ON SUSTAINABLE COMPUTING.