An Energy Consumption Model and Analysis Tool for Cloud Computing Environments

FeiFei Chen, Jean-Guy Schneider, Yun Yang, John Grundy and Qiang He
Faculty of Information and Communication Technologies
Swinburne University of Technology
Melbourne, Australia 3122
{feifeichen, jschneider, yyang, jgrundy, qhe}@swin.edu.au

Abstract—Cloud computing delivers computing as a utility to users worldwide. A consequence of this model is that cloud data centres have high deployment and operational costs, as well as significant carbon footprints for the environment. We need to develop Green Cloud Computing (GCC) solutions that reduce these deployment and operational costs and thus save energy and reduce adverse environmental impacts. In order to achieve this objective, a thorough understanding of the energy consumption patterns in complex Cloud environments is needed. We present a new energy consumption model and associated analysis tool for Cloud computing environments. We measure energy consumption in Cloud environments based on different runtime tasks. Empirical analysis of the correlation of energy consumption and Cloud data and computational tasks, as well as system performance, will be investigated based on our energy consumption model and analysis tool. Our research results can be integrated into Cloud systems to monitor energy consumption and support static or dynamic system-level optimisation.

Keywords—green computing; Cloud computing; energy consumption; performance analysis

I. INTRODUCTION

Cloud computing is a new and promising paradigm which delivers computing as a utility [1]. It provides computation, software, data access, and storage services through the Internet. Key advantages include that users can scale on demand their computing and data storage services without the traditional large upfront investment in computing infrastructure. This has led to huge investments over the last few years in building large-scale data centres, due to the massive growth in demand for high performance, Cloud data and computational services. As Cloud computing becomes more widespread, these increasing data storage and computation needs raise the energy consumption of their large infrastructures. Thus energy consumption has become a critical concern in designing modern Cloud systems. Firstly, the high energy consumption of data centres often results in consumption of electricity produced by “brown” generation facilities, resulting in high emission of carbon dioxide, with negative impacts on the environment. Secondly, a common economic objective of Cloud providers is to minimise their total deployment and operational costs. High energy consumption directly contributes to both deployment and operational costs. As analysed in [2], the electricity consumption for powering the data centres in the USA alone is projected to reach 100 billion kWh at the cost of $7.4 billion by 2011. This power consumption contributes up to 42% of a data centre’s monthly budget [3]. Therefore, energy consumption, as well as its impact on system performance, operating cost and the environment, have become critical issues in Cloud environments [4].

Many efforts have been made to improve energy efficiency in Cloud environments. Some simple techniques provide basic energy management for servers in Cloud environments, i.e., turning on and off servers, putting them to sleep or using Dynamic Voltage/Frequency Scaling (DVFS) [5] to adjust servers’ power states. DVFS adjusts CPU power (consequently the performance level) according to the workload. However the scope of DVFS optimisation is limited to CPUs. Another approach for improving energy efficiency is to adopt virtualisation techniques to get better resource isolation and reduce infrastructure energy consumption through resource consolidation and live migration [6]. Using virtualisation techniques, several energy-aware resource allocation policies and scheduling algorithms have been proposed to optimise total energy consumption in Cloud environments [7]. However, energy consumption and system performance of data centres vary greatly with different system resource configuration and allocation strategies, as well as the workload and types of running tasks in the Cloud [8]. By its nature, these Cloud workloads is a highly variable and application-specific. In addition, system performance should not be impacted while energy consumption is being minimised. Thus in order to achieve this, a thorough understanding of energy consumption patterns of each individual type of task in Cloud environments e.g., data retrieval and data processing, and how energy consumption of such tasks is affected by different workloads or different system configurations, is required. In addition, the correct energy consumption models and energy consumption profiles are needed.
However, there are still some major challenges that need to be addressed:

- What determines the energy consumption of specific tasks?
- How do we characterise and profile the energy consumption of different tasks?
- What is the relationship between energy consumption and workload of tasks?
- What is the relationship between energy consumption and system performance?

In order to identify these challenges, we propose a new energy consumption model and an analysis tool for Cloud environments. Our energy consumption model gives a detailed description of each parameter used to calculate the energy consumption in Cloud environments. The analysis tool takes the energy consumption model as input and characterises energy consumed by each task. It helps identify the relationship between energy consumption and running tasks in Cloud environments, as well as system configuration and performance. The idea is to use our energy consumption analysis tool and empirical energy and task analysis results to statically plan task organisation and scheduling on available cloud platforms, or to dynamically monitor energy consumption and support system-level optimisation (or both).

We briefly summarise the state-of-the-art of energy consumption models and analysis approaches in Section II. In Section III, we introduce our energy consumption model for Cloud environments. The energy consumption analysis tool and validation framework are described in Section IV. In Section V, we describe our empirical validation approaches. Finally, we conclude the paper and discuss directions for future work in Section VI.

II. RELATED WORK

Energy consumption in Cloud computing environments has quickly become a popular research topic. Several efforts have been made to build energy consumption models and develop energy-aware cost models for optimising the total cost, i.e., deployment cost plus operational cost, in Cloud environments. Li et al [9] propose a cost model for calculating the total cost of ownership and utilisation cost in Cloud environments. They also developed suites of metrics for this calculation. However their calculation granularity is a single hardware component. Similarly, Jung et al [10] focus on power consumed by physical hosts. Their energy consumption models do not take into account the impact of specific workloads running on specific hardware. In addition, a consumer-provider Cloud cost model has been proposed by Mach and Schikuta [11]. Their energy consumption calculation is based on the number of Java Virtual Machine (JVM) instances on each server. However, it is hard to measure the actual numbers of JVM because of the dynamic nature of JVM life cycle. Moreover, Lee and Zomaya [12] propose an energy model of Cloud tasks for developing energy-conscious task consolidation algorithms to reduce energy consumption in Cloud environments. However, the energy model simply assumes the relation between CPU utilisation and energy consumption is with linear increasing.

Energy saving policies in Cloud environments has also been investigated in the past few years. Liu et al [13] describe a new cloud infrastructure to dynamically consolidate VMs based on CPU utilisation of servers, in order to identify idle machines. These idle machines can be turned off to save energy. Verma et al [14] use the characteristics of VMs, such as cache footprint and the set of applications running on the VMs, to drive power-aware placement of VMs. VirtualPower [15] is proposed to exploit power management decisions of guest VMs on virtual power states. The virtual power states of guest VMs are considered as preconditions to run local and global energy management policies across the computation.

Research efforts have also been made in profiling and analysing the energy consumption in Cloud environments. In existing research outcomes, profiling and analysis are conducted by actively using energy benchmarks or closely monitoring the energy profile of individual system components, such as CPU, cache, disk and memory, at runtime. A framework for energy optimisation and development of an energy-aware operation system has been developed based on the availability of energy models for each hardware component [16]. Chen et al [17] propose a linear power model that presents the behaviour and power consumption from individual components to a single work node. Joule meter is a power meter for VMs [18]. This makes use of software components to monitor the resource usage of VMs and then converts it to energy consumed based on the power model of each individual hardware resource.

Some of the abovementioned works have made some initial efforts in benchmarking the energy and system performance. However, none of them has identified the relationship between energy consumption and runtime tasks with different configurations in Cloud environments as well as system performance. In this paper, we propose a new energy consumption model and an analysis tool for Cloud environments to address these issues.

III. OVERVIEW OF ENERGY CONSUMPTION MODEL

In this section, we introduce our energy consumption model for Cloud environments under different system configurations.

A. Model Structure

Most of the current state-of-the-art research on energy efficiency has predominantly focused on the optimisation of processing resources, i.e. computation servers, since processing resources contribute to the major part of total energy consumption. However, earlier research results indicate that at least 30% of total energy is consumed by communication links, switching and aggregation resources [19]. In addition, data retrieval on storage resources also contributes to a significant part of energy consumption, with data size growing significantly [20]. Although energy
consumed by the cooling system is significant in Cloud data centre environments [21], cooling overheads can be approximately modelled as a fixed figure of the total energy consumed [22]. Therefore, we focus on storage, computation and communication resources as the energy consumed by these are both significant and highly dynamic in Cloud environments.

Energy consumed during idle time is also a fixed part of the total energy consumption. The dynamic part of energy consumption is the additional energy consumed by running tasks in the Cloud. We further divide the energy consumption into two parts: fixed energy consumption (energy consumed during idle time) and variable energy consumption (additional energy consumed by Cloud tasks).

### B. Energy Calculation Units

Instead of measuring energy consumption of individual hardware components we treat a single task running in a Cloud environment as the fundamental unit for energy profiling. This is because workload and type of task have significant impact on both energy consumption and system performance [8]. A task can be defined as a three-tuple denoted as \((\text{input, process, output})\). For instance, a task to zip a file can be denoted as \((\text{source file, zip, target file})\). If the input can be partitioned and processed separately, we treat each partition as an individual task. For example, a 10GB file may be processed as a single file or divided into two 5GB files to be processed by separate tasks.

Although all tasks utilise the abovementioned three resources, the percentage of each resource used by different tasks is different. For example, some tasks focus on computation while others focus on data retrieval or update. In order to demonstrate the impact of energy consumption produced by different types of tasks, we divide tasks into three types, as defined:

1. For a task \(i\), if it is data-intensive, it has an ID \(ts_i\). The energy consumption of the task is defined as \(EC_{ts_i}\). The process number of task \(ts_i\) is defined as \(PT_{ts_i}\). The size of data processed by task \(ts_i\) is defined as \(DS_{ts_i}\). Size of data transmitted is defined as \(DT_{ts_i}\).

2. For a task \(i\), if it is computation-intensive, it has an ID \(tc_i\). The energy consumption of the task is defined as \(EC_{tc_i}\). The process number of task \(tc_i\) is defined as \(PT_{tc_i}\). The size of data processed by task \(tc_i\) is defined as \(DS_{tc_i}\). Size of data transmitted is defined as \(DT_{tc_i}\).

3. For a task \(i\), if it is communication-intensive, it has an ID \(tt_i\). The energy consumption of the task is defined as \(EC_{tt_i}\). The process number of task \(tt_i\) is defined as \(PT_{tt_i}\). The size of data processed by task \(tt_i\) is defined as \(DS_{tt_i}\). Size of data transmitted is defined as \(DT_{tt_i}\).

### C. Calculation Formula

As discussed earlier, total energy consumption is composed of fixed energy consumption, defined as \(E_{Fix}\), and variable energy consumption, defined as \(E_{Var}\). The total energy consumption defined as \(E_{Total}\) is formulated as follows:

\[
E_{Total} = E_{Fix} + E_{Var}
\]  \hspace{1cm} (1)

In this energy consumption model, we focus on the additional energy consumption on storage, computation and communication resources. The three kinds of additional energy consumption are defined as follows:

1. Energy consumption of storage resources is denoted by \(E_{Storage}\).
2. Energy consumption of computation resources is denoted by \(E_{Comp}\).
3. Energy consumption of communication resources is denoted by \(E_{Comm}\).

Given the above, formula (1) can be transformed into:

\[
E_{Total} = E_{Fix} + E_{Storage} + E_{Comp} + E_{Comm}
\]  \hspace{1cm} (2)

Total energy consumed in cloud environments is the summary of energy consumed by all tasks. However, different types of task simultaneously exist in Cloud environments. The total energy consumption of two tasks is not equivalent to the sum of individually consumed energy due to scheduling overhead and interference. Additional energy consumption will be generated by scheduling overhead, denoted as \(E_{Sche}\). Therefore, the total energy consumption can be defined as:

\[
E_{Total} = E_{Fix} + \sum_{i=1}^{n} EC_{ts_i} + \sum_{i=1}^{n} EC_{tc_i} + \sum_{i=1}^{n} EC_{tt_i} + E_{Sche}
\]  \hspace{1cm} (3)

According to formula (3), in order to calculate the total energy consumption, we need to identify the energy consumed by each task and the scheduling operation.

For each task, energy consumption is tightly coupled with task workload. For instance, for a computation-intensive task, the energy consumption increases with the number of processes used by the task [8]. Although there are other specific application related factors that can influence the energy consumed by a task, e.g. the encoding and decoding algorithms of a data-intensive task, we focus on the factors related to workload because those factors, shared by most tasks, directly and largely influence energy consumption. In addition, system configurations have significant impact on energy consumption. Energy consumption increases dramatically when the number of VMs configured on a physical machine increases [23]. Hence, the energy consumed by each task is determined by the number of processes, the size of data to be processed, the size of data transmitted and system configuration denoted as \(C_i\). Thus, the energy consumption of each type of task can be defined as:

\[
EC_{ts_i} = f_i\left(PT_{ts_i}, DS_{ts_i}, DT_{ts_i}, C_i\right)
\]  \hspace{1cm} (4)
\[ EC_{t_0} = f_0\left(PT_{t_0}, DS_{t_0}, DT_{t_0}, C_t\right) \]  \hspace{1cm} (5)

\[ EC_{t_0} = f_1\left(PT_{t_0}, DS_{t_0}, DT_{t_0}, C_t\right) \]  \hspace{1cm} (6)

One of our objectives is to find regularities between input and output of the formulas defined above by doing empirical analysis. The validating approaches will be introduced in Section V.

IV. ENERGY CONSUMPTION ANALYSIS TOOL AND PROFILING FRAMEWORK

In this section, we first introduce our prototype energy consumption analysis tool. Then, we present our energy profiling framework.

A. Architecture of Our Energy Consumption Analysis Tool

We have been developing a tool to calculate and analyse total energy consumption. Our energy consumption model defined in Section III is integrated in this tool. Figure 1 shows the architecture of our energy consumption analysis tool.

The core component of the tool is the Analysis Engine. It takes our energy consumption model and application task parameters as input, as well as the performance data collected from the Cloud by Data Collection Engine. The Data Collection Engine collects two kinds of data: 1) the energy consumed by each task; and 2) the value of system performance parameters, e.g. response time, CPU utilisation, memory utilisation, disk I/O and network throughput.

B. Energy Profiling Framework

ERL (Energy Research Lab) is a leading-edge research lab at Swinburne University of Technology. With the energy monitoring facilities in this lab, we can precisely monitor the actual energy consumption of our private cloud servers and network devices. In this lab, the energy consumption of each individual hardware device can be measured. The measurement of energy is managed by PowerNode, power usage profiling equipment invented by GreenWave Reality. The result of energy measurement is reported to the GreenWave Gateway, used for creating a mesh-based Home Area Network (HAN). The Gateway then sends the data to the GreenWave Reality data centre. The information in the data centre can be viewed via a website or desktop display.

The framework of our cloud energy profiling system is presented in Fig. 2. Based on the task descriptions, our workload generator component generates workloads to stress load our private Cloud infrastructure. A PowerNode monitor is connected to each hardware device, including servers, data stores and switches, to measure the energy consumed by each device while Cloud tasks are running. When the tasks are finished, the energy consumption analysis tool collects energy data via the PowerNode Reality data centre and task performance data from Cloud Controller. The system configuration file encapsulates the settings of various hardware and software components, e.g. number and type of hosts, virtual machines, switches, data stores, etc at multiple system levels. According to these data and system configuration descriptions, the analysis tool can demonstrate the correlation of the energy consumption, the tasks, the Cloud platform configuration, and system performance.

![Figure 2. Framework of the energy profiling system.](http://www.greenwavereality.com)

V. EMPIRICAL VALIDATION APPROACHES

In this section, we describe the approaches that we will perform on our cloud test bed in our Energy Research Lab. Our goal is to characterise energy usage and performance in Cloud environments by measuring and analysing the impact of various tasks and system configurations. Our energy consumption model will be validated based on this empirical data.

Different tasks can impose substantially different resource requirements, and even the resources requirements of a particular task can vary over time. The energy consumption of the Cloud depends not only on the hardware configurations of hosting servers, but also on the runtime tasks in the Cloud. In addition, since the energy consumed by the same task executing on different servers...
can vary significantly, and the total energy consumption of two co-located workloads is not equivalent to the sum of individually consumed energy, we also need to study the interference caused by resource contention when multiple tasks are running simultaneously.

We have been focusing on the energy consumption of three types of tasks: data-intensive tasks, computation-intensive tasks and communication-intensive tasks. In order to measure the additional energy consumption of the Cloud, the energy consumed in idle state is measured as the benchmark. Then, we profile and analyse the energy consumption of the single task and multiple tasks of the same type, as well as the corresponding system performance. In order to simplify the problem, the energy consumption of mixed types of tasks will be profiled and analysed in our future work. As analysed in Section III, we aim to identify the relationship between the input and output in the energy consumption model. The inputs of the model are the task parameters and the system configurations. The output of the model is the additional energy consumed by the tasks. In addition, we will also analyse the system performance with various workloads of different tasks. The energy consumption profiling analysis metrics of all three task types are presented in Table 1.

### Table 1. Profiling Metrics

<table>
<thead>
<tr>
<th>Task Type</th>
<th>Task Parameter</th>
<th>System Configuration</th>
<th>Energy Consumed</th>
<th>Performance Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data-Intensive</td>
<td>(PT_{\text{DS}}) (DS_{\text{DT}}) (DT_{\text{ti}})</td>
<td>Description of hardware and software resources allocated</td>
<td>(EC_{\text{DS}}) (E_{\text{DS}})</td>
<td>CPU Utilization, Memory Utilization, Bandwidth of Disk I/O, Task Execution Time</td>
</tr>
<tr>
<td>Computation-Intensive</td>
<td>(PT_{\text{c}}) (DS_{\text{c}}) (DT_{\text{ci}})</td>
<td>Description of hardware and software resources allocated</td>
<td>(EC_{\text{c}}) (E_{\text{c}})</td>
<td>CPU Utilization, Memory Utilization, Task Execution Time</td>
</tr>
<tr>
<td>Communication-Intensive</td>
<td>(PT_{\text{n}}) (DS_{\text{n}}) (DT_{\text{ni}})</td>
<td>Description of hardware and software resources allocated</td>
<td>(EC_{\text{n}}) (E_{\text{n}})</td>
<td>CPU Utilization, Memory Utilization, Network Bandwidth Task Execution Time</td>
</tr>
</tbody>
</table>

**A. Energy Consumption of Data-Intensive Tasks**

A data-intensive task usually needs to process a large amount of data in different data storage servers within the same data centre. It requires high local disk I/O bandwidth in order to meet customers’ performance requirements. Although in reality, the storage servers could be deployed in different data centres located in different geographic locations, we only consider the energy consumption in one data centre for the purpose of simplicity.

As investigated in [24], the energy consumption of the same hard disk is not linear with the data transferred to or from the disk because of the data processing overhead. Thus, we focus on the correlation of energy consumption and the data transferred in or out the storage server. We profile and analyse the energy consumption of tasks with different data sizes, as well as system performance.

**B. Energy Consumption of Computation-Intensive Tasks**

A computation-intensive task usually requires a number of isolated processes to perform the computation. In a Cloud environment, different VMs are allocated to deal with different processes. These VMs are hosted by different servers and VM migration is executed if the server’s capacity reaches the limit or the server cannot meet performance requirement. The migration of VMs can increase energy consumption significantly [25]. However, the energy consumption might increase with the number of processes within the same server since the overhead of scheduling will increase accordingly. We focus on the energy consumed by different computation workloads.

**C. Energy Consumption of Communication-Intensive Tasks**

A communication-intensive task requires many network resources to transmit large amounts of data. Switches form the basis of the interconnection fabric of a Cloud network. Therefore, switches are the main energy consumers among network resources. Traditionally, the energy consumption of a switch depends on the hardware parameters, such as type of switch, number of ports and port transmission rates. However, the energy consumption may increase with the amount of data stream because of the processing overhead. In addition, the total energy consumption might be impacted by the network congestion because of the imbalance between the computation speed and the communication speed. We investigate this issue by applying different network workloads.

**VI. CONCLUSION AND FUTURE WORK**

Analysing the dynamics of energy consumption in Cloud environments is necessary and valuable for developing efficient energy-saving resource management and techniques for Green Cloud Computing. In this paper, we have presented an energy consumption model for calculating the total energy consumption in Cloud environments. We have also described an energy consumption analysis tool and empirical energy analysis approaches being used in our investigation. We treat a single task as a unit and measure the energy produced by the task under various configurations. The correlation of system performance and energy consumed are extracted based on the analytical results. These experimental research results are crucial for developing energy management mechanisms to reduce the energy consumption while achieving the expected system performance for Cloud environments. We practice the empirical validation approach in our Energy Research Lab.

As future work, we will investigate the energy consumption of mixed types of tasks in the Cloud. In addition, we will integrate an energy cost rate and an “energy dirtiness rate” into our energy consumption model to factor in differing environment impact of different cloud...
energy sources. A new energy cost model is to be investigated in order to minimise the total energy cost. Moreover, with the proliferation of Cloud computing, much effort has been initiated on executing scientific workflows in Cloud environments and evaluating the trade-off between system performance and resources costs. We also plan to extend our research to the energy cost of scientific workflows that is composed of multiple types of tasks on larger scales. It is necessary to research the energy cost of scientific workflows, as it helps investigate scientific workflow task consolidating strategies.

ACKNOWLEDGMENT

The research work reported in this paper is partly supported by Australian Research Council under Discovery Project DP110101340.

REFERENCES


