Hindawi
Security and Communication Networks
Volume 2017, Article ID 2376484, 15 pages
https://doi.org/10.1155/2017/2376484

Research Article

Data Placement for Privacy-Aware Applications over Big Data in Hybrid Clouds

Xiaolong Xu,[1,2,3,4] Xuan Zhao,[3] Feng Ruan,[5] Jie Zhang,[3] Wei Tian,[1,2] Wanchun Dou,[3] and Alex X. Liu[4]

[1] *School of Computer and Software, Nanjing University of Information Science and Technology, Nanjing, China*
[2] *Jiangsu Engineering Centre of Network Monitoring, Nanjing University of Information Science and Technology, Nanjing, China*
[3] *State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, China*
[4] *Department of Computer Science and Engineering, Michigan State University, East Lansing, MI, USA*
[5] *School of Information and Control, Nanjing University of Information Science and Technology, Nanjing, China*

Correspondence should be addressed to Wanchun Dou; douwc@nju.edu.cn

Received 17 August 2017; Accepted 24 September 2017; Published 8 November 2017

Academic Editor: Md Z. A. Bhuiyan

Nowadays, a large number of groups choose to deploy their applications to cloud platforms, especially for the big data era. Currently, the hybrid cloud is one of the most popular computing paradigms for holding the privacy-aware applications driven by the requirements of privacy protection and cost saving. However, it is still a challenge to realize data placement considering both the energy consumption in private cloud and the cost for renting the public cloud services. In view of this challenge, a cost and energy aware data placement method, named CEDP, for privacy-aware applications over big data in hybrid cloud is proposed. Technically, formalized analysis of cost, access time, and energy consumption is conducted in the hybrid cloud environment. Then a corresponding data placement method is designed to accomplish the cost saving for renting the public cloud services and energy savings for task execution within the private cloud platforms. Experimental evaluations validate the efficiency and effectiveness of our proposed method.

1. Introduction

The rapid development of science and technology makes the network information increase exponentially, and the continuous accumulation of network data brings opportunities and challenges for big data. Big data gives plenty of benefits to humanity in many fields including network, health care, transportation, finance, military, and politics. Recommendation service, prediction service, and computing service can be realized through big data storage and analysis [1–3]. As huge amount of data could lead to system crash with the traditional data storage techniques, it is essential to realize the big data storage [4, 5]. Distributed file systems and databases are beneficial to big data storage. The emergence and development of cloud computing contributes to the big data storage, access, and processing, as cloud computing provides ubiquitous and various resources, to respond the explosive growth of data accumulation.

Cloud computing is a powerful technology that can provide humorous cloud services for the customers everywhere through the Internet, which aggregates geodistributed resources, to accomplish higher throughput and computing ability [6]. The customers could benefit from the public cloud services, as they are not necessary to build the infrastructure and manage the data center [7]. Currently, data privacy issues have received a lot of attention due to the increasing concern of the privacy and the data value protections, since the individuals often suffer heavy blows from privacy leaks [8, 9]. How to protect data and improve the security level of private data has become a hot topic of cloud computing [10]. Generally, it is an effective way to place these datasets in the private cloud; thus hybrid cloud for big data storage should be taken into consideration for privacy-aware applications [11, 12].

Table 1: Key terms and descriptions for cost, access time, and energy analysis in hybrid cloud.

Notation	Description
V	All the available VM instances $V = \{v_1, v_2, \dots, v_M\}$
v_m	mth $(1 \leq w \leq W)$ VM in V
D	The dataset need to be placed $D = \{d_1, d_2, \dots, d_W\}$
d_w	wth $(1 \leq w \leq W)$ dataset in D
l_w^m	The binary variable to judge whether d_w is placed on v_m
$r_{m',w}$	The data access time for $v_{m'}$ $(1 \leq m' \leq M)$ extracting the d_w
T	The tasks need to be performed, $T = \{t_1, t_2, \dots, t_N\}$
t_n	nth $(1 \leq n \leq N)$ task in T
I_n^m	The binary variable to judge whether t_n is place on v_m
$c_{n,w}$	The time cost for the task t_n to extract the dataset d_w
F_m	The binary variable to judge whether v_m is on public cloud
O	The total access time of public datasets
O'	The access time for obtaining the datasets in private cloud
PE	The baseline power consumed by the active PMs
VE	The power consumed by the active VMs is calculated by
IE	The power consumed by the VMs in the idle mode
TE	The power consumed by the switches due to data access
E	The total power consumption to perform the tasks

Nowadays, an increasing number of applications, especially for the scientific workflows, for example, weather forecasting flows, are deployed in the hybrid cloud.

Data placement has a direct impact on the data access efficiency and the cost for data storage, as the locations of big data could affect the overhead for the service renting and the access time for data extraction. Therefore, reasonable and efficient data placement methods are essential to the performance of big data processing [13–15]. For the data placement in hybrid clouds, the payment of public cloud services and the energy consumption generated in the private cloud are key factors to determine the locations of the datasets for the execution of the privacy-aware applications.

With the above observations, it is still a challenge to realize data placement for privacy-aware applications over big data in the hybrid cloud, considering the cost saving in the public cloud and the energy saving in the private cloud. In view of this challenge, we design an efficient data placement method to deal with the above challenge. Our main contributions are threefold. Firstly, we undergo cost, access time, and energy analysis over big data in hybrid cloud. Secondly, a corresponding cost and energy aware data placement method, named CEDP, is designed to address the resource provisioning problem for the privacy-aware applications over big data in the hybrid cloud. Finally, a sequence of experimental analysis is conducted to validate the efficiency and the effectiveness of our proposed method.

The rest of this paper is organized as follows. In Section 2, formalized concepts are presented for cost, access time, and energy analysis over big data in hybrid cloud. Section 3 specifies our proposed method. The comparison analysis and performance evaluation are described in Section 4. Section 5 presents the related work, and Section 6 concludes the paper and gives outlook for the future work.

2. Cost, Access Time, and Energy Analysis over Big Data in Hybrid Cloud

In this section, cost and access time for data placement in the public cloud are analyzed. Besides, the access time and the energy consumption analysis for data placement in the private cloud are also presented. Table 1 specifies the key terms and description for cost, access time, and energy analysis over big data in the hybrid cloud.

2.1. Cost and Access Time Analysis in Public Cloud. In the cloud environment, the datasets and the tasks both need to be hosted in the form of VMs. Suppose there are M VM instances that are available for hosting tasks and datasets across the public clouds and the private cloud data centers, denoted as $V = \{v_1, v_2, \dots, v_M\}$. Suppose there are W datasets that need to be stored in the hybrid cloud platforms, denoted as $D = \{d_1, d_2, \dots, d_W\}$.

Let l_w^m be a binary variable to judge whether d_w $(1 \leq w \leq W)$ is placed on v_m $(1 \leq m \leq M)$, which is measured by

$$l_w^m = \begin{cases} 1, & d_w \text{ is placed on } v_m \\ 0, & \text{Otherwise.} \end{cases} \tag{1}$$

The reserved datasets need to be extracted from one VM to another. When the VM $v_{m'}$ $(1 \leq m' \leq M)$ needs to extract d_w, the data access time is denoted as $r_{m',w}$, which is calculated by

$$r_{m',w} = \begin{cases} l_w^m \cdot \dfrac{|d_w|}{\sum_{i=1}^{\kappa_{m,m'}} bw_i}, & i > 0 \\ 0, & i = 0, \end{cases} \tag{2}$$

where $\kappa_{m,m'}$ is the number of links between v_m and $v_{m'}$, $|d_w|$ represents the data size of d_w, and bw_i $(0 \leq i \leq \kappa_{m,m'})$ is the bandwidth of the ith link.

Suppose there are N tasks that need to be performed in the hybrid cloud environment, denoted as $T = \{t_1, t_2, \ldots, t_N\}$. The tasks are executed on the VMs whether in the public cloud or in the private cloud. Thus, these tasks have placement relationships with the VM instances. Let I_n^m be the binary variable to judge whether t_n $(1 \leq n \leq N)$ is placed on v_m, which is measured by

$$I_n^m = \begin{cases} 1, & t_n \text{ is placed on } v_m \\ 0, & \text{Otherwise.} \end{cases} \quad (3)$$

Thus, the time cost for the nth $(1 \leq n \leq N)$ task t_n to extract the dataset d_w, denoted as $c_{n,w}$, is calculated by

$$c_{n,w} = \sum_{m'=1}^{M} I_n^{m'} \cdot r_{m',w}. \quad (4)$$

As big data is now expanding explosively in both academia and industry, the execution of one task may need several datasets for supporting. For the data extraction in the public cloud, we mainly focus on the bandwidth cost for the data transferring.

Let F_m be the binary variable to judge whether v_m is placed on the public cloud, which is measured by

$$F_m = \begin{cases} 1, & v_m \text{ is placed on public cloud} \\ 0, & \text{Otherwise.} \end{cases} \quad (5)$$

The datasets could be extracted by the tasks in both public cloud and private cloud. The access time by the tasks in the public cloud is calculated by

$$O_{\text{pub}} = \sum_{n=1}^{N} \sum_{w=1}^{W} \sum_{m=1}^{M} \sum_{m'=1}^{M} F_{m'} \cdot F_m \cdot I_n^m \cdot r_{m',w} \cdot G_n^w, \quad (6)$$

where G_n^w is a binary variable to judge whether d_w is necessary for the execution of t_n.

$$G_n^w = \begin{cases} 1, & t_n \text{ requires } d_w \text{ for execution} \\ 0, & \text{Otherwise.} \end{cases} \quad (7)$$

The access time for the datasets in the public cloud by the tasks in the private cloud is calculated by

$$O_{\text{pri}} = \sum_{n=1}^{N} \sum_{w=1}^{W} \sum_{m=1}^{M} \sum_{m'=1}^{M} F_{m'} \cdot (1 - F_m) \cdot I_n^m \cdot r_{m',w} \cdot G_n^w. \quad (8)$$

Then the total access time for extracting the datasets from the public cloud is calculated by

$$\begin{aligned} O &= O_{\text{pub}} + O_{\text{pri}} \\ &= \sum_{n=1}^{N} \sum_{w=1}^{W} \sum_{m=1}^{M} \sum_{m'=1}^{M} F_{m'} \cdot F_m \cdot I_n^m \cdot r_{m',w} \cdot G_n^w \\ &\quad + \sum_{n=1}^{N} \sum_{w=1}^{W} \sum_{m=1}^{M} \sum_{m'=1}^{M} F_{m'} \cdot (1 - F_m) \cdot I_n^m \cdot r_{m',w} \cdot G_n^w \\ &= \sum_{n=1}^{N} \sum_{w=1}^{W} \sum_{m=1}^{M} \sum_{m'=1}^{M} F_{m'} \cdot I_n^m \cdot r_{m',w} \cdot G_n^w. \end{aligned} \quad (9)$$

The bandwidth cost for the data transferring in the public cloud is calculated by

$$C = \sum_{n=1}^{N} \sum_{w=1}^{W} \sum_{m=1}^{M} \sum_{m'=1}^{M} F_{m'} \cdot I_n^m \cdot r_{m',w} \cdot G_n^w \cdot \theta_m, \quad (10)$$

where θ_m is the expenditure of v_m for data transferring per unit time.

2.2. Access Time and Energy Consumption Analysis in Private Cloud. For a private cloud data center, the cloud providers need to take into account the time cost and the power consumption while allocating the datasets. The access time for obtaining the datasets in private cloud is calculated by

$$\begin{aligned} O' = \sum_{n=1}^{N} \sum_{w=1}^{W} \sum_{m=1}^{M} \sum_{m'=1}^{M} (1 - F_{m'}) \cdot (1 - F_m) \cdot I_n^m \cdot r_{m',w} \\ \cdot G_n^w. \end{aligned} \quad (11)$$

Suppose there are Q PMs denoted as $P = \{p_1, p_2, \ldots, p_Q\}$ that are available to host the private datasets. And the tasks in the private clouds are also deployed on the PMs in P.

The energy consumption in the private cloud for the execution of the privacy-aware applications mainly refers to the energy consumed by the PM base power, active VMs, and the unused VMs, and the energy consumption due to data transferring. The PMs in the sleep mode also consume a certain amount of power, but it is far less than the energy consumed by the active PMs in the order of magnitude, that could be neglected [16, 17].

The baseline power consumed by the active PMs is calculated by

$$\text{PE} = \sum_{q=1}^{Q} \alpha_q \cdot \eta_q, \quad (12)$$

where α_q $(1 \leq q \leq Q)$ and η_q are the baseline power consumption rate and the total running time for p_q.

The power consumed by the active VMs is calculated by

$$\text{VE} = \sum_{m=1}^{M} (1 - F_m) \cdot \beta_m \cdot \delta_m, \quad (13)$$

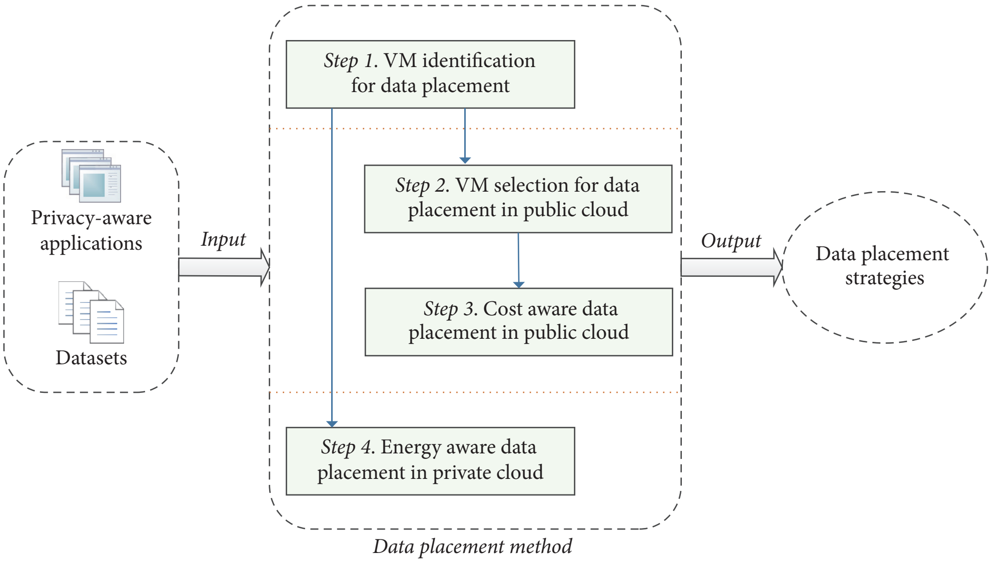

Privacy-aware applications
Datasets
Input
Step 1. VM identification for data placement
Step 2. VM selection for data placement in public cloud
Step 3. Cost aware data placement in public cloud
Step 4. Energy aware data placement in private cloud
Output
Data placement strategies
Data placement method

Figure 1: Specification of our proposed method.

where β_m, δ_m are the power consumption rate and the running time for v_m, respectively.

The power consumed by the VMs in the idle mode is calculated by

$$\mathrm{IE} = \sum_{m=1}^{M} \left(1 - F_m\right) \cdot \beta'_m \cdot \delta'_m, \tag{14}$$

where β'_m and δ'_m are the power rate and the idle time of v_m, respectively.

The power consumed by the switches due to data access is calculated by

$$\mathrm{TE} = \sum_{n=1}^{N} \sum_{w=1}^{W} \sum_{m=1}^{M} \left(1 - F_m\right) \cdot c_{n,w} \cdot G_n^w \cdot \theta_{n,w} \cdot \gamma, \tag{15}$$

where θ_m is the number of switches between t_n and d_w, and γ is the active power rate for each switch.

Then the total power consumption to perform the tasks with data extraction processes is calculated by

$$E = \mathrm{PE} + \mathrm{VE} + \mathrm{IE} + \mathrm{TE}. \tag{16}$$

Then the objectives for data placement over big data in the hybrid cloud are min E and min C.

3. Cost and Energy Aware Data Placement Method for Privacy-Aware Applications

In this section, a cost and energy aware data placement method is proposed for privacy-aware applications over big data in the hybrid environment. In this method, we aim to reduce the cost for renting cloud services and achieve energy savings in the private cloud.

3.1. Method Overview. In this paper, a cost and energy aware data placement method is proposed to address the challenges of data placement problem for the privacy-aware applications in the hybrid cloud environment.

Figure 1 shows the specification of our proposed method. The input of our method is the privacy-aware applications with task distribution in the hybrid cloud, and the datasets that need to be placed in the hybrid cloud. Our method consists of four main steps, i.e., VM identification for data placement, VM selection for data placement in public cloud, cost aware data placement in public cloud, and energy aware data placement in private cloud.

For each dataset, the VMs that need to access it are identified in Step 1. Then in the public cloud, we choose available VMs to host the datasets, which need to be placed in the public cloud, through Step 2. For the VMs obtained by Step 2, we conduct cost aware data placement through Step 3, so that the optimal data placement strategies with minimum cost are designed for the datasets that need to be placed in the public cloud. For the datasets with privacy preservation requirements, they are necessary to place in the private cloud. Energy aware data placement are designed in Step 4 to achieve energy savings while allocating VMs to store these datasets. The ultimate output of our method is the data placement strategies.

3.2. VM Identification for Data Placement. In the hybrid cloud, both the datasets and the tasks combined in the privacy-aware applications physical resources from cloud platforms for hosting, which could be responded by the VMs. Generally, the resource capacity of PMs and the resource requirements from tasks and datasets are specified by the amount of the resource units, that is, the VM instances [16]. For many public cloud vendors, such as Amazon, they provide many types of VM instances, including CPU-intensive instances and I/O optimized instances.

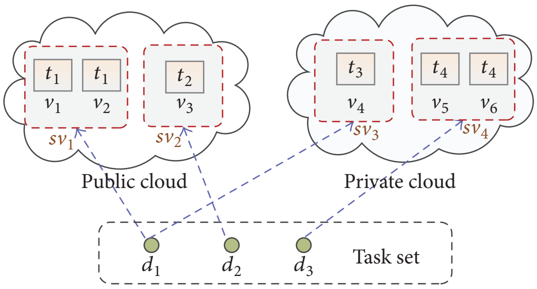


FIGURE 2: An example of special VM identification with tasks (t_1~t_4) and datasets (d_1~d_3) deployed on VMs (v_1~v_6) in the hybrid cloud.

```
Input: The dataset $D$
Output: The special VM set $VS$
(1)  for $i = 1$ to $|D|$ do
(2)    for $j = 1$ to $N$ do
(3)      if $t_j$ requires $v_i$ for execution then
(4)        Add $t_j$ to $r_i$
(5)        flag = 1, $k = 1$
(6)        while flag == 1&& $k \le M$ do
(7)          if $I_n^k == 1$ then
(8)            Add $v_k$ to $vs_i$
(9)            flag = 0
(10)         else $k = k + 1$
(11)         end if
(12)       end while
(13)     end if
(14)   end for
(15)   Classify the VMs in $vs_i$ as several special VMs
(16) end for
(17) Return $SV$
```

ALGORITHM 1: Special VM identification for data access.

Definition 1 (resource requirement of t_n). The resource requirement of t_n mainly refers to the VM instance type and the number of VM instances, which is denoted as $r_n = \{vt_n, cou_n\}$, where vt_n and cou_n are the VM instance type and the required total amount of VM instances of t_n, respectively.

To satisfy the requirements of a dataset that needs to be stored, one or more VM instances with the same specification are requested, and these instances could be regarded as a special VM.

Definition 2 (special VM). For the VM instances that deployed to perform the same task or store the same dataset, it could be treated as a special VM.

Currently, in the big data era, large-scale datasets could be shared for multiple tasks, and one task may need several different datasets for execution. To place the dataset efficiently, the special VMs that rented for hosting the tasks, which require the datasets for execution, should be identified. For the dataset d_w in D, the special VM set is denoted as sv_w; then D has a corresponding special VM set $SV = \{sv_1, sv_2, \ldots, sv_W\}$.

Figure 2 shows an example of special VM identification. In this example, there are three datasets (i.e., d_1, d_2, and d_3) that need to be stored in the hybrid cloud. d_1 needs to be accessed by tasks t_1 and t_3, d_2 needs to be accessed by t_2, and d_3 needs to be accessed by t_4. t_1 requires v_1 and v_2 for execution, t_2 requires v_3 for execution, t_4 requires v_5 and v_6 for execution. In this example, the two VM instances v_1 and v_2 occupied by t_1 could treated as a special VM sv_1, v_3 is treated as sv_2, v_4 is treated as special VM sv_3, and v_5 and v_6 are treated as special VM sv_4.

The VMs identified according to the task distribution and the dataset access requirements should be specified as the special VMs. For the dataset d_w, the corresponding special VMs are put in the VM set sv_w. Then all special VM sets for all datasets are recorded as $SV = \{sv_1, sv_2, \ldots, sv_W\}$.

Algorithm 1 specifies the key idea of special VM identification for data access. The input is the dataset D. This algorithm should traverse all the datasets (Line (1)) and all the tasks (Line (2)). For each dataset, we find the VMs of the tasks that need to access the dataset (Lines (3) to (12)). Finally, the output is the special VM set SV.

3.3. VM Selection for Data Placement in Public Cloud. The datasets should be placed on the VMs, thus the available

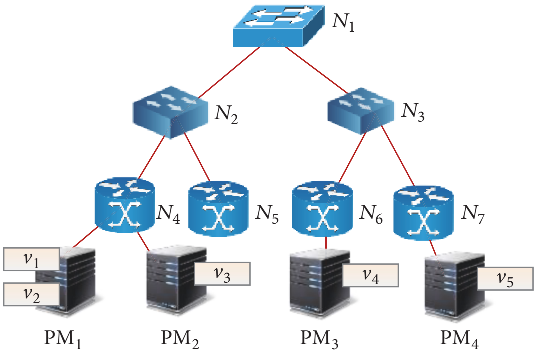


Figure 3: An example of VM distribution in a FatTree-based data center network with core switch N_1, aggregation switches N_2 and N_3, and edge switches N_4~N_7.

VMs on the cloud should be identified to store the datasets. For the PMs and VMs in the private cloud, the resource scheduler could be aware of the map relationship between PMs and VMs. However, in the public cloud, the resource scheduler can only select the available VMs that cloud vendors provided.

In the public cloud platforms, when renting VMs for storing the datasets. We would like to choose the VM instances with lowest cost. Generally, the more the renting time of bandwidth is, the more cost the users need to pay. Thus, the access time for tasks extracting the datasets should be taken into consideration. FatTree is a typical network topology for cloud datacenters. For most VMs connected to different switches, the data access time is almost the same. In this section, we conduct VM selection process, to select the VMs that could store the datasets in the public cloud. And these VMs should be sorted by the distances between the selected VMs and the VM for holding tasks, identified in Algorithm 1.

Figure 3 shows an example of VM distribution in a FatTree-based data center network. There are seven switches (i.e., N_1~N_7), distributed as a tree network. In these switches, N_1 is the core switch, N_2 and N_3 are the aggregation switches, and the switches N_4~N_7 are edge switches. There are three PMs (i.e., PM_1~PM_3) connected to the edge switches. In this example, there are five running VMs distributed on these PMs, where v_1 and v_2 are placed on PM_1, v_3 is placed on PM_2, v_4 is placed on PM_3, and v_5 is placed on PM_4.

The data access time is a key objective that users take into consideration, which is closely relevant to the distances between the VMs where the task hosts and the datasets locates. The distance calculation relies on the locations on the FatTree network. The distance between two VMs on the same PM is 0. For example, as shown in Figure 3, v_1 and v_2 are placed on the same PM; the distance between v_1 and v_2 is 0. The distance between two VMs on different PMs depends on the number of links between these two VMs. For example, the distance between v_1 and v_3 is 2. Furthermore, if the PMs are connected to two different edge switches, but they have the same aggregation switch, the distance is double than the distance between two VMs connected to the same edge switch. For example, the distance between v_4 and v_5 is 4. Besides, if the VMs are connected to the different aggression switch, the distance of them is triple than the distance of VMs connected to the same edge switch. For example, the distance between v_1 and v_4 is 6, and the distance between v_3 and v_4 is also 6.

Based on the process of distance calculation, the identified VMs, which are available for hosting the datasets, could be sorted by the increasing order of the distance values. The datasets placed in the public cloud also be accessed by the tasks deployed in the private cloud. In the private cloud datacenter, the network is also built based on FatTree; thus the data access between VMs in these two kinds of cloud platforms needs to access the core switch, the aggregation switch, and the core switch in both public and private cloud, which is an edge-to-edge communication across clouds and platforms. So, in this section, we mainly focus on the access time within the public cloud platform.

For the dataset d_w, which is arranged to store in the public cloud, there are several tasks in the public cloud should access d_w; the special VMs in sv_w should be updated by removing the special VMs in private clouds. The corresponding VMs are selected to hold d_w, which are put in the VM set cv_w. For all the datasets, the VM set list is denoted as $CV = \{cv_1, cv_2, \ldots, cv_W\}$.

Algorithm 2 specifies the key process of VM selection for data placement. The input is the VM node set SV. This algorithm traverses all the VM set in SV (Line (1)), and, for each VM set, the VMs in the private cloud removed (Line (2)). For each VM in the VM set, we select the VMs in the public cloud and calculate the distances between the selected VM and the VM in the VM set (Lines (3) to (12)). Then the selected VMs are put in the VM set CV, and it is sorted in the increasing order of distance (Line (13)). The final output is the identified VM set CV.

3.4. Cost Aware Data Placement in Public Cloud. After the processing of VM selection in Algorithm 2, the VMs that could be allocated to store the datasets in public cloud are obtained. As, in the public cloud, the cost and the access time are closely relevant, especially, in the FatTree network, in this section, we mainly focus on the cost for the public cloud services.

```
Input: The VM set $VS$
Output: The identified VM set $CV$
for $i = 1$ to $|VS|$ do
    Remove the VMs in private cloud from $vs_i$
    for $j = 1$ to $vs_i$ do
        for $k = 1$ to $M$ do
            if $v_k$ is in the public cloud then
                if $v_k$ is not in $cv_i$ then
                    Add $v_k$ to $cv_i$
                    Calculate the distance between $v_k$ to $vs_{i,j}$
                end if
            end if
        end for
    end for
    Sort the VMs in $cv_i$ in the increasing order of distance
end for
Return $CV$
```

Algorithm 2: VM selection for data placement.

The cost mainly depends on the service time and the unit payment fee for VM renting. As we know there are different VM instances provided by the cloud vendors, and the cost for these VM instances are various; thus, to achieve cost efficiency, we should select the optimal data placement strategy with minimum cost for the datasets that need to be placed in the public cloud.

Definition 3 (data placement strategy of d_w). The dataset placement strategy of d_w consists of the VM instances that need to rent for storing d_w, denoted as s_w.

For all the datasets in D, the relevant data placement strategy set is denoted as $S = \{s_1, s_2, \ldots, s_W\}$. After the processing by Algorithm 1, we get the special VMs for each dataset access, which are used to hold the tasks that need to access the dataset. Although the datasets in the public cloud could be accessed by the tasks both running in the public cloud and the private cloud, the datasets only can use the public cloud services for storing, due to the resource limit in the private cloud. The VMs that could be employed to respond the resource requirements could be achieved by Algorithm 2.

Then for each dataset in the public cloud, we try to select the suitable data placement policy, to save the cost expenditure for cloud service renting. As there are multiple data placement policies for each dataset, the placement policy with the minimum cost, calculated by formula (10), is selected as the final data placement strategy.

Algorithm 3 specifies the key idea of cost aware data placement. The input for this algorithm is the dataset D that need to be placed in the hybrid cloud. The special VMs for each dataset are identified by Algorithm 1 (Line (1)). Then we traverse all the datasets (Line (2)) and select the datasets that need to be placed in the public cloud (Line (3)). The VM instances are selected to respond to the resource requirements of each dataset in public cloud (Line (5)). Then multiple iterations are undergoing to find the data placement policy with lowest cost (Lines (7) to (16)). The output of this algorithm is the data placement strategy S.

3.5. Energy Aware Data Placement in Private Cloud.

After the data placement in Section 3.4, the data placement strategies for the datasets that need to be placed in the public cloud are all designed. For the privacy-aware applications of users, some tasks contained in them are deployed in their own datacenter, that is, the private cloud constructed by themselves. In this scenario, the resource scheduler could know the specification of the task distribution and the map relationship between VMs and PMs. Similar to the network topology of the public cloud, the private cloud data center also constructed based on the FatTree network. Thus, when allocating VMs to store the datasets in the private cloud, the access time is not a key issue to take care of. In the private cloud, we mainly focus on reducing the energy consumption due to data access and task execution.

In Section 2, the energy consumption is specified as the energy consumed by the running PMs, the active VMs, the idle VMs, and the switches due to data transferring. The energy consumed due to data access could be specified as the following three scenarios:

(1) The datasets could be placed on the PMs that the tasks located which need to access the datasets. In this case, the energy consumption of switches due to data transferrin could be neglected. For example, the VMs v_1 and v_2 in Figure 4 share the data storage of PM_1; thus there are no data transferring through any switch.

(2) The datasets also could be placed on the VMs which are connected to the same edge switches with the tasks which need to access the datasets. Then the data access only across one switch, and the energy for data transferring only occurs in this switch. For example, the VMs v_1 and v_3 are placed on PM_1 and

```
Input: The dataset $D$
Output: The dataset placement strategy $S = \{s_1, s_2, \ldots, s_W\}$
Algorithm 1 Special VM identification for data access
for $i = 1$ to $|D|$ do
    if $d_i$ needs to be placed in public cloud then
        Update $vs_i$ and $r_i$
        Get the VM instances by Algorithm 2
        Classify $CV$ as special VMs, denoted as $sv_i$
        $C = \text{MAX}CV$, $j = 1$
        while $j \leq |sv_i|$ do
            if $sv_{i,j}$ can hold $d_i$ then
                Calculate the total cost TC by (10)
                if TC $< C$ then
                    $C =$ TC
                else $j = j + 1$
                end if
            end if
        end while
        Update $s_i$ with cost $C$
    end if
end for
Return $S$
```

Algorithm 3: Cost aware data placement.

PM_2 separately, and the data transferring between these two VMs only employs the switch N_4.

(3) The datasets also cloud be placed on the PMs with the different switches to the PMs that hosted the tasks need to access the datasets. In this situation, whether the datasets placed on which PM, the energy consumed due to data transferring is same, as the data access use five switches, that is, two edge switches, two aggregation switches, and one core switch. For example, in Figure 4, the energy consumption due to data access between v_1 and v_4, will use the edge switch N_4, the aggregation switch N_2, the core switch N_1, the aggregation switch N_3, and the edge switch N_6.

From the above analysis, the occupation of the VMs which are near to the VMs the task hosts, which need to access the dataset, will cause fewer energy consumption. Besides, for the energy consumption for PM running, the main idea to save the energy consumption is to make full use of the running PMs and try best to reduce the number of running PMs. If the VMs are placed on the PMs with the tasks, it can achieve energy saving from the perspective of both data access and PM running. Thus, the PMs are sorted in the decreasing order of the distances between the VM to host the tasks and the VM identified for hosting the dataset. Then, we select the PM through multiple iterations; at last we select the PM to host the dataset with the minimum energy consumption, calculated by formula (16).

The data placement strategy for the datasets in the private cloud could be improved as $s_w = \{nm_w, pm_w\}$, where nm_w and pm_w are the amount of VM instances and the VM location of d_w, respectively.

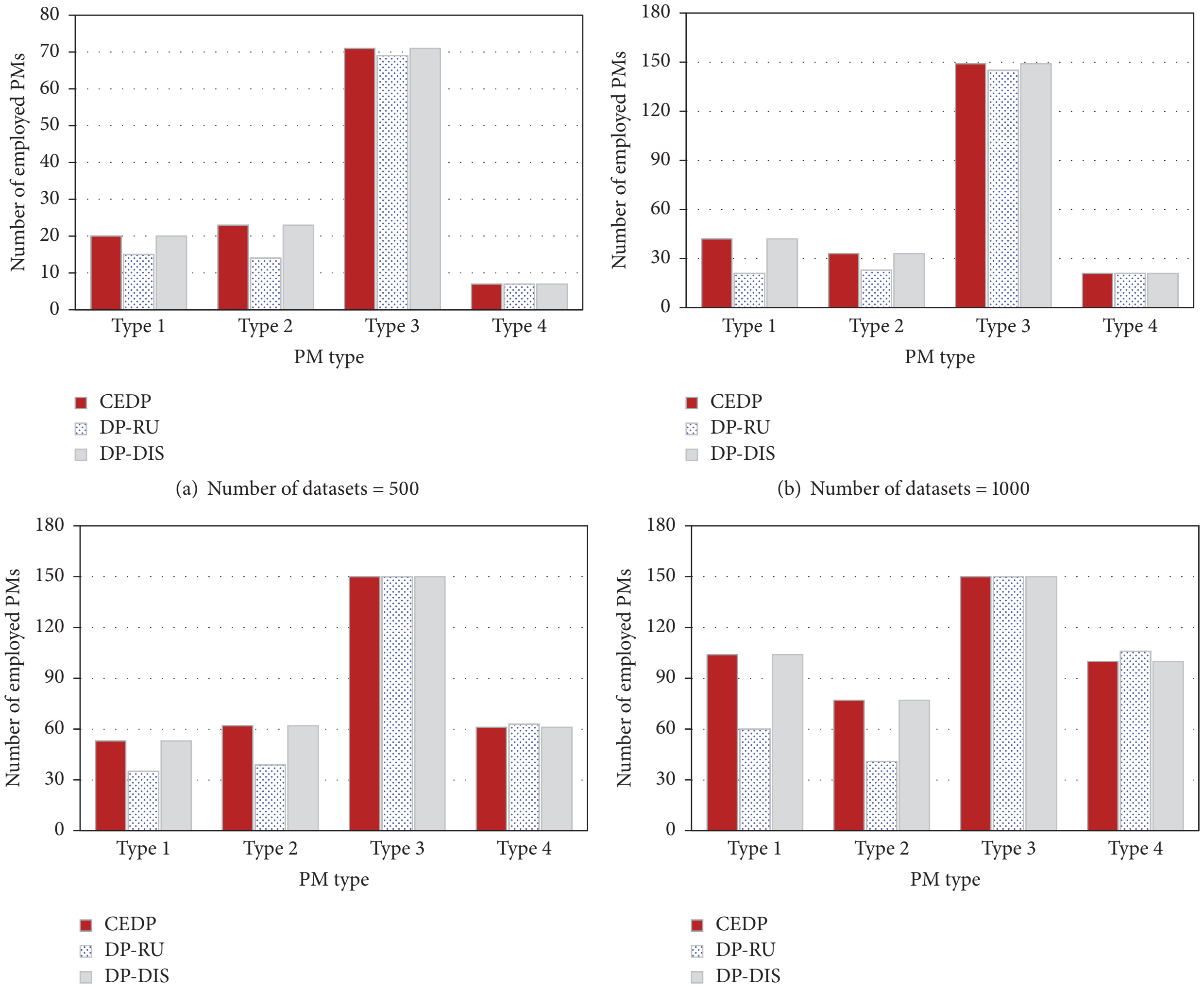
Algorithm 4 shows the key idea of energy aware data placement in private cloud. In this algorithm; the input is the dataset D. The special VMs for hosting the tasks in the private cloud are identified by Algorithm 1 (Line (1)). Then all the datasets are traversed to check whether the dataset needs to place in the private cloud (Line (2) and Line (3)). For each dataset, we traverse the PM list, to select the PMs that can hold it (Lines (4) to (8)). The PM list is sorted in the decreasing order of VM distance between the VMs selected by Algorithm 1 and the PMs (9). We find the PM with energy consumption for data placement through multiple iterations (Lines (10) to (20)). The final output of this algorithm is the updated data placement strategy set S.

4. Experimental Evaluation

In this section, we use the cloud simulator CloudSim to simulate the hybrid cloud environment and the data placement method CEDP.

4.1. Experimental Context. In this paper, 4 different scales datasets with VM distributions and task distributions are generated to validate our proposed method. Besides, the bid datasets are also provided with 4 different scales. The above datasets are stored in the Google disk (https://drive.google.com/open?id=0B0T819XffFKrQVFoOHM2TU1zZHM). Our method is validated on the physical node, equipped with the processor (Intel Core i5-5300U CPU @2.30 GHz) and 8.00 GB memory.

The parameters used in our simulation are specified in Table 2. We use 4 types of PMs (150 PMs for each type) to construct our private cloud platform. And the energy consumption rate settings are similar to our previous work in [16–18].



(a) Number of datasets = 500

(b) Number of datasets = 1000

(c) Number of datasets = 1500

(d) Number of datasets = 2000

FIGURE 4: Comparison of the number of employed PMs by CEDP, DP-RU, and DP-DIS with 4 different scales of datasets placed in private cloud.

TABLE 2: Parameter Settings.

Parameter	Domain
Number of hosts in private cloud	600
PM type	4
PM baseline energy consumption (W)	{86, 93.7, 192, 342}
Number of VMs for each type of PM	{6, 8, 18, 34}
VM type in public cloud	4
Cost for each type of VM (cents/h)	{2, 2.5, 3, 3.5}
Number of datasets	{500, 1000, 1500, 2000}
Unit cost for data access (cents/GB·time)	1.5
Bandwidth (MB)	80
Switch power (W)	300
Data size for each dataset (GB)	[1, 50]

```
Input: The dataset $D$
Output: The dataset placement strategy $S = \{s_1, s_2, \ldots, s_W\}$
Algorithm 1 Special VM identification for data access
for $i = 1$ to $|D|$ do
    if $d_i$ needs to be placed in private cloud then
        for $j = 1$ to Q do
            if $p_j$ can hold $d_i$ then
                Add $p_j$ to $cp_i$
            end if
        end for
        Sort $cp_i$ in the decreasing order of VM distance
        between the VM in $svi$ and the VM in $cp_i$
        Calculate the energy consumption $\mathrm{ec}_1$ after
        allocating $d_i$ to $cp_{i,1}$ by (16)
        $MC = \mathrm{ec}_1$, num = 2
        while num $\leq |cp_i|$ do
            Calculate the energy consumption $\mathrm{ec}_{\mathrm{num}}$ after
            allocating $d_i$ to $cp_{i,\mathrm{num}}$ by Eq. (16)
            if $\mathrm{ec}_{\mathrm{num}} < MC$ then
                $MC = \mathrm{ec}_{\mathrm{num}}$
            end if
            num = num + 1
        end while
        Update $s_i$ according to $MC$ and the relevant PM
    end if
end for
Return $S$
```

Algorithm 4: Energy aware data placement in private cloud.

We use 4 datasets with different scale of datasets that need to be placed in the hybrid cloud. And 20% of them are privacy-aware data, which should be placed in the private cloud. For the public cloud, there are 4 types of VMs that are presented for data placement.

4.2. Performance Evaluation. The performance evaluation is conducted from two aspects, that is, the public cloud and the private cloud. For the private cloud, we mainly focus on the energy consumption and the access time. However, for the public cloud, we mainly validate the method performance through the comparison analysis on cost for VMs renting. As our work is the first to privatize a data placement policy for privacy-aware applications over big data in hybrid cloud, two benchmark methods are employed for comparison analysis. On is a resource utilization aware data placement method, named DP_RU, which aims to optimize the resource utilization for the cloud datacenters. The other is a distance-aware data placement method, named DP_DIS, which aims to place the datasets near the tasks which needs to access them.

(1) Evaluation on Energy Consumption in Private Cloud. The energy consumption is closely relevant to the number of the employed PMs. Figure 4 shows the comparison of the employed PMs by CEDP, DP_RU, and DP_DIS with 4 different scales of datasets for data placement in private cloud. In Figure 4, it is intuitive that our proposed method employs the same number of PMs with DP_DIS. It is because that our method considers the data access time, which also depends on the distance between the tasks and the datasets. From Figure 4, we can find that, in most cases, DP_RU applies fewer PMs than CEDP and DP_DIS, because DP_RU is a greedy algorithm to achieve high resource usage, regardless of the data access time.

Although we employ more PMs than DP_RU, it does not mean DP_DIS is more energy efficient than CEDP, as the data access processes also consume a certain amount of energy.

Figure 5 shows the comparison of the total energy consumption with different scale of datasets by using CEDP, DP_RU, and DP_DIS. As shown in Figure 5, CEDP and DP_DIS achieve the same energy consumption after data placement in the private cloud. And these two methods achieve better energy efficiency than DP_RU, although DP_RU employs fewer PMs than CEDP. We can detect that there is more energy consumed by the switches due to data access.

(2) Evaluation on Access Time in Private Cloud. As the applications need big data for processing, the tasks need to access the placed data frequently. The access time is a key attribute to measure the quality of cloud service. Figure 6 (including 4 subfigures) shows the comparison of the access time by CEDP, DP_RU, and DP_DIS with different scale of datasets placed in private cloud. For these 4 datasets, there are 100, 200, 300, and 400 privacy-aware datasets, separately. To

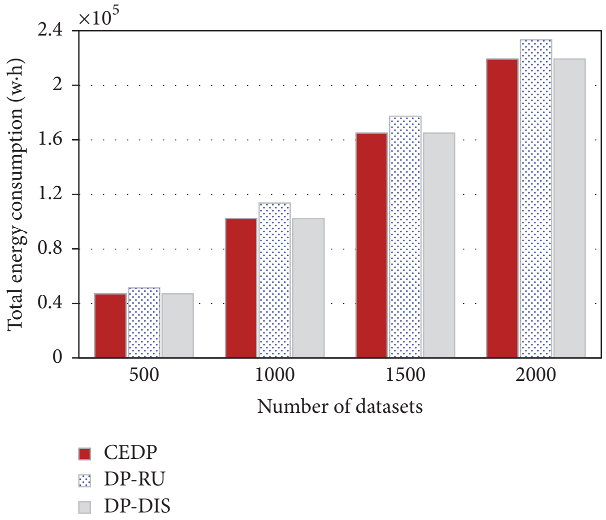
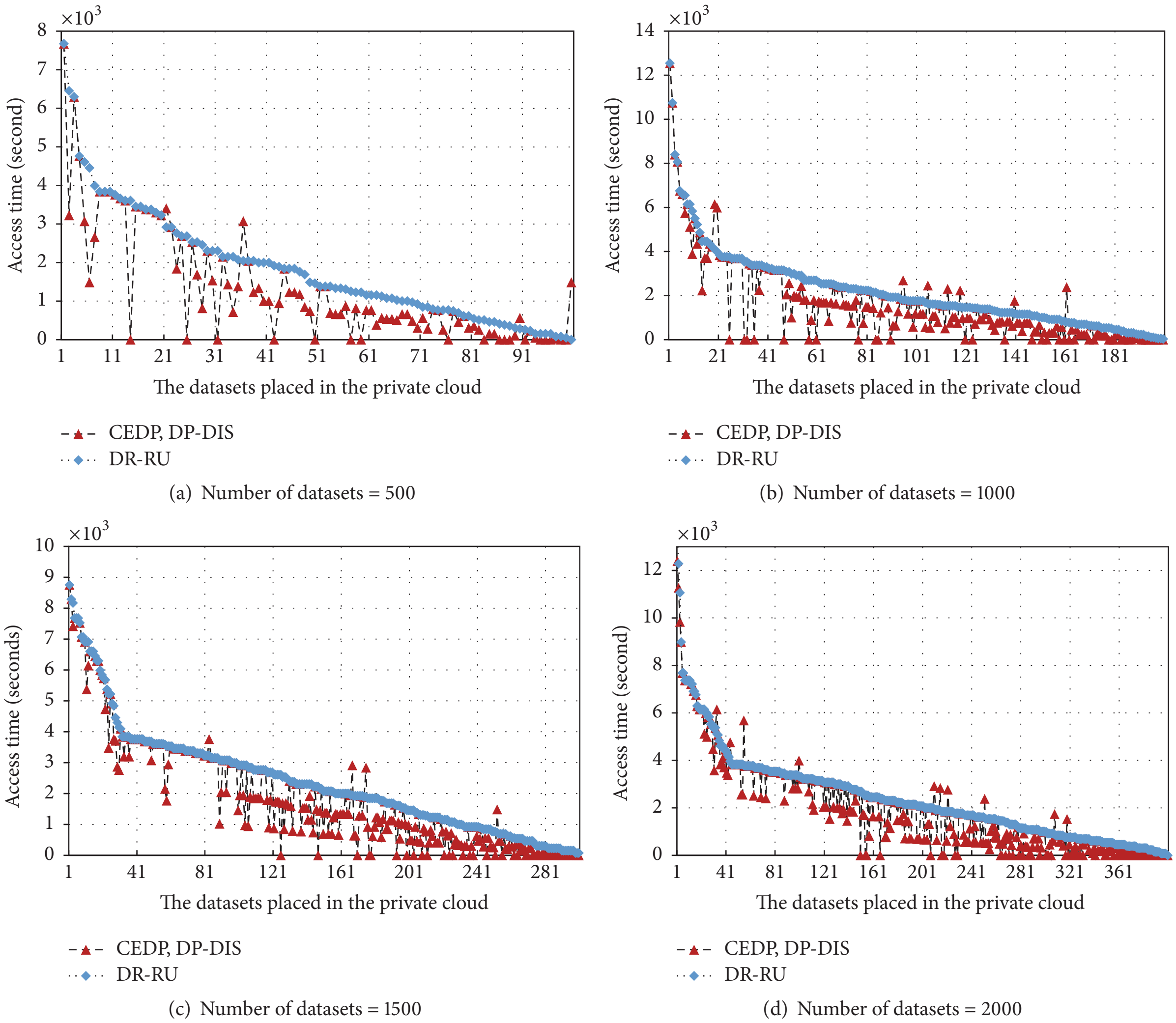


Figure 5: Comparison of total energy consumption with different scale of datasets by using CEDP, DP-RU, and DP-DIS.



(a) Number of datasets = 500

(b) Number of datasets = 1000

(c) Number of datasets = 1500

(d) Number of datasets = 2000

Figure 6: Comparison of access time by CEDP, DP-RU, and DP-DIS with different scale of datasets placed in private cloud.

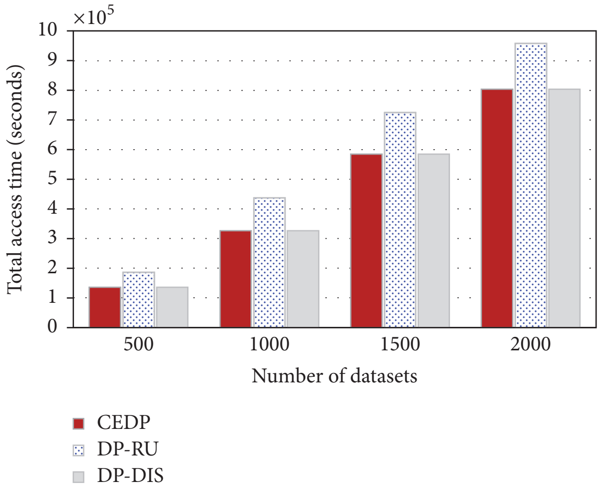


Figure 7: Comparison of total access time with different scale of datasets by using CEDP, DP-RU, and DP-DIS.

better show the comparison analysis, we sort the experimental results in the decreasing order of access time, achieved by DR_RU. From Figure 6, we can find that our method CEDP obtains optimal access time than DP_RU, in most cases. For example, in Figure 6(c), when the total number of datasets is 1500, there are 300 privacy-aware datasets that should be placed in the private cloud, where there are 296 datasets that could obtain better access time by CEDP among 300 datasets than DP_RU. Obsoletely, there are still some accidental cases that DP_RU achieves better access time than CEDP. As there are multiple datasets that could be provided for the same task, a dataset has been placed in advance, and there are no spare PMs that connected to the same edge switch or the aggression switch. Our proposed method CEDP is a global optimization method that can achieve better time efficiency than DP_RU from a global perspective. Therefore, for some of the datasets, it is reasonable that there are some accidental cases. Overall, CEDP could obtain time efficiency than DP_RU.

Figure 7 shows the comparison of total overall access time with different scale of datasets by using CEDP, DP_RU, and DP_DIS. It is intuitive from Figure 7 that CEDP could get the same access time as DP_DIS, and both of them are superior compared to DP_RU. For example, when the number of datasets is 2000 and CEDP and DP_DIS get the overall time near 8×10^5 seconds, whereas DP_RU achieves near 9.5×10^5 seconds overall access time. CEDP and DP_DIS are both distance-aware data methods; thus they are time-sensitive.

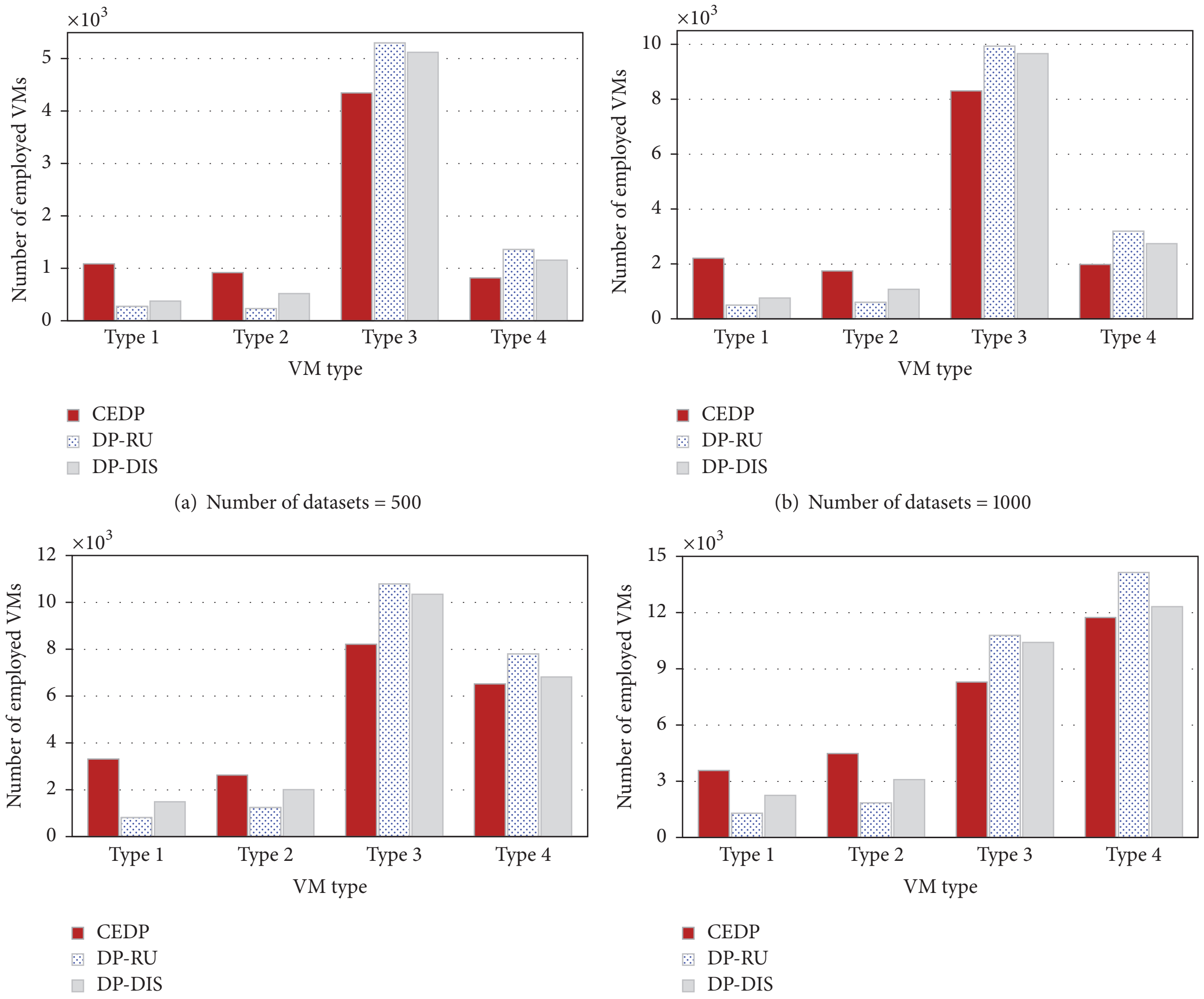
(3) Evaluation on Cost in Public Cloud. For the performance evaluation in the public cloud, the cost for VMs renting is one of the most key metrics. The renting fee is closely relevant to the VM instance type. Thus, we analyze the number of employed VMs for data placement in public cloud. Four figures in Figure 8 show the comparison analysis of the number of employed VMs by CEDP, DP-RU, and DP-DIS with different scale of datasets placed in public cloud. From Figure 8, we can find that CEDP employs cheaper VM instances (i.e., type 1 and type 2) than DP_RU and DP_DIS. Besides, CEDP employs fewer expensive VM instances (i.e., type 3 and type 4) than DP_RU and DP_DIS. For example, in Figure 8(a), CEDP employs over 100 VM instances with type 1 and type 2, whereas DP_RU and DP_DIS both employ less than 50. But CEDP employs fewer VMs with respect to type 3 and type 4 VMs.

Then we conduct the statistics of total cost for these 3 methods. Figure 9 shows the comparison of total cost with different scale of datasets by using CEDP, DP-RU, and DP-DIS. In Figure 9, we can find that our method could achieve cost savings compared to DP_RU and DP_DIS, as we present a cost-sensitive method for data placement in public cloud.

5. Related Work

Big data needs a huge mass of computing resources and storage resources, to promote the development of cloud technology [19–21]. Data placement in cloud environment has been widely concerned to improve the quality of cloud services.

Data Placement over Cloud. Due to the necessity and importance of data placement, there exist multiple methods to place users' data over multiple clouds [13–15, 22–24]. Fan et al. [13] constructed a tripartite graph in GBDP (genetic based data placement) scheme and demonstrated validation of the scheme. Jiao et al. [14] proposed an optimization approach leveraging graph cuts to optimize multiobjective data placement in multicloud for socially aware services. Yu and Pan [15] showed a location-aware associated data placement scheme to improve the associated data location and the localized data serving and at the same time ensure the balance between nodes. Agarwal et al. [22] presented a system named Volley which can analyze the logs of data center requests and output migration recommendations to address data placement problem. Yu and Pan [23] proposed the sketch-based data placement (SDP) to lower the overhead and keep the benefits of the data placement. Su et al. [24] proposes that better features can be provided by multicloud



(a) Number of datasets = 500

(b) Number of datasets = 1000

(c) Number of datasets = 1500

(d) Number of datasets = 2000

Figure 8: Comparison of the number of employed VMs by CEDP, DP-RU, and DP-DIS with different scale of datasets placed in public cloud.

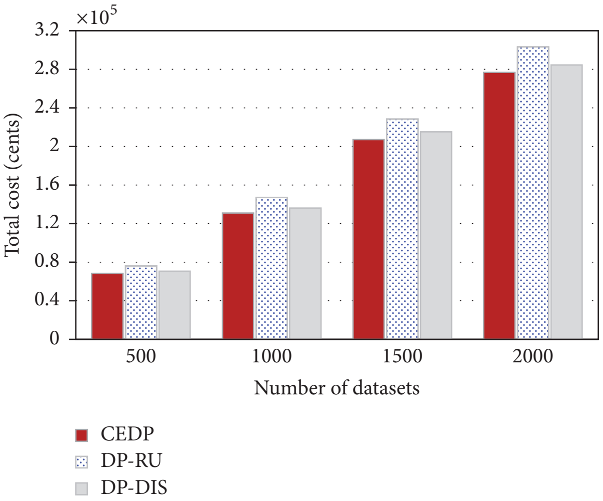


Figure 9: Comparison of total cost with different scale of datasets by using CEDP, DP-RU, and DP-DIS.

storage and presented a systematic model Triones to formulate data placement in multiple clouds storage by using erasure coding.

Although cloud computing can provide rich resources of computing and storage and data placement can also maximize the efficiency of cloud usage and expenditure reduction, the data privacy problems become the greatest concern for more and more people [25]. Hybrid cloud that combine public cloud and private cloud can protect user privacy by placing private data on private cloud [26].

Hybrid Cloud. The study of hybrid cloud is also increasing. Mixed cloud on the task scheduling, virtual machine scheduling, privacy, and other related work have made some progress. Some were discussed about separating private data from public data and placing them in trusted private cloud and untrusted public cloud, respectively [27–30]. Zhou et al. [27] presented a set of techniques for privacy-aware data retrieval by splitting data and storing on hybrid cloud. Huang and Du [28] proposed a scheme to achieve image data privacy over hybrid cloud efficiently and proposed a one-to-one mapping function for image encryption. Wang and Jia [29] described several methods about protecting data security in hybrid cloud and discussed an authentication intercloud model. Abrishami et al. [30] presented a scheduling algorithm to protect data privacy while minimizing the cost and satisfying the users' limitation. Tasks, datasets, and virtual machines scheduling in hybrid cloud to maximize the benefits and minimize the cost were studied in [31–34]. Zhou et al. [31] produced a three-stage framework to explore the benefits of uploading applications to hybrid cloud. Qiu et al. [32] described a model for heterogeneous workloads scheduling and an online algorithm for tasks preemptive scheduling. Zinnen and Engel [33] used HGP to estimate task execution times and proved that the former result is the same as optimization with unknown generating distributions. Bakshi [34] introduced a secure hybrid cloud approach and the virtual switching technologies. Although these papers take improving the overall efficiency of the hybrid cloud by scheduling into account, it does not consider the effect of the storage location of the data on overall efficiency when the task has been properly allocated in public and private cloud. Our work considers the impact of data placement in a hybrid cloud environment, paying attention to the energy loss on the private cloud and the rental price on the public cloud.

To the best of our knowledge, there are few works which focus on the data placement problem in the hybrid cloud for privacy-aware applications over big data, considering both the cost in public cloud and the energy consumption in private cloud.

6. Conclusion and Future Work

In the big data era, data placement becomes increasingly important for data accessing and analysis, as the datasets are often too large to host with the computing task. Cloud platforms are proved to be powerful to host the data-intensive tasks. Besides, the data privacy is also a key concern for both academia and industry; thus it is necessary to undergo data placement in the hybrid cloud. In this paper, we propose an energy and cost aware data placement method driven by the requirements of the privacy-aware applications in the hybrid cloud. Our method aims to reduce the energy consumption in the private cloud and save cost for renting the VMs in the public cloud.

For future work, we will try to realize our method for the real-world workflow applications, such as weather forecasting, where the raw data should be stored in the private cloud, and the intermediate data could be stored in the public cloud.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research is supported by the National Science Foundation of China under Grants no. 61702277, no. 61672276, no. 61772283, no. 61402167, and no. 61672290, the Key Research and Development Project of Jiangsu Province under Grants nos. BE2015154 and BE2016120. Besides, this work is also supported by the Startup Foundation for Introducing Talent of NUIST, the Open Project from State Key Laboratory for Novel Software Technology, Nanjing University, under Grant no. KFKT2017B04, the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD) fund, Jiangsu Collaborative Innovation Center on Atmospheric Environment and Equipment Technology (CICAEET), and Project "Six Talent Peaks Project in Jiangsu Province" under Grant no. XYDXXJS-040.

References

[1] J. Heidrich, A. Trendowicz, and C. Ebert, "Exploiting big data's benefits," *IEEE Software*, vol. 33, no. 4, pp. 111–116, 2016.

[2] X. Wu, X. Zhu, G.-Q. Wu, and W. Ding, "Data mining with big data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 1, pp. 97–107, 2014.

[3] L. Mertz, "What can big data tell us about health? Finding gold through data mining," *IEEE Pulse*, vol. 7, no. 5, pp. 40–44, 2016.

[4] J. Li, Z. Xu, Y. Jiang, and R. Zhang, "The overview of big data storage and management," in *Proceedings of the 13th IEEE International Conference on Cognitive Informatics and Cognitive Computing, ICCI-CC 2014*, pp. 510–513, London, UK, August 2014.

[5] Q. Liu, W. Cai, J. Shen, Z. Fu, X. Liu, and N. Linge, "A speculative approach to spatial-temporal efficiency with multi-objective optimization in a heterogeneous cloud environment," *Security and Communication Networks*, vol. 9, no. 17, pp. 4002–4012, 2016.

[6] E. Chovancová, L. Vokorokos, and M. Chovanec, "Cloud computing system for small and medium corporations," in *Proceedings of the 13th IEEE International Symposium on Applied Machine Intelligence and Informatics, SAMI 2015*, pp. 171–174, svk, January 2015.

[7] K. Peng, R. Lin, B. Huang, H. Zou, and F. Yang, "Link importance evaluation of data center network based on maximum flow," *Journal of Internet Technology*, vol. 18, no. 1, pp. 23–31, 2017.

[8] B. Nelson and T. Olovsson, "Security and privacy for big data: A systematic literature review," in *Proceedings of the 4th IEEE International Conference on Big Data, Big Data 2016*, pp. 3693–3702, usa, December 2016.

[9] S. Yu, "Big Privacy: Challenges and Opportunities of Privacy Study in the Age of Big Data," *IEEE Access*, vol. 4, pp. 2751–2763, 2016.

[10] C. Perera, R. Ranjan, and L. Wang, "End-to-end privacy for open big data markets," *IEEE Cloud Computing*, vol. 2, no. 4, pp. 44–53, 2015.

[11] H. M. Musse and L. A. Alamro, "Cloud Computing: Architecture and Operating System," in *Proceedings of the 2016 Global Summit on Computer &amp; Information Technology (GSCIT)*, pp. 3–8, Sousse, Tunisia, 2016.

[12] A. Sill, "Standards for Hybrid Clouds," *IEEE Cloud Computing*, vol. 3, no. 1, pp. 92–95, 2016.

[13] W. Fan, J. Peng, X. Zhang, and Z. Huang, "Genetic Based Data Placement for Geo-Distributed Data-Intensive Applications in Cloud Computing," in *Advances in Services Computing*, Springer International Publishing, 2016.

[14] L. Jiao, J. Lit, W. Du, and X. Fu, "Multi-objective data placement for multi-cloud socially aware services," in *Proceedings of the 33rd IEEE Conference on Computer Communications, IEEE INFOCOM 2014*, pp. 28–36, Toronto, Canada, May 2014.

[15] B. Yu and J. Pan, "Location-aware associated data placement for geo-distributed data-intensive applications," in *Proceedings of the 34th IEEE Annual Conference on Computer Communications and Networks, IEEE INFOCOM 2015*, pp. 603–611, hkg, May 2015.

[16] X. Xu, W. Dou, X. Zhang, and J. Chen, "EnReal: An Energy-Aware Resource Allocation Method for Scientific Workflow Executions in Cloud Environment," *IEEE Transactions on Cloud Computing*, vol. 4, no. 2, pp. 166–179, 2016.

[17] W. Dou, X. Xu, S. Meng et al., "An energy-aware virtual machine scheduling method for service QoS enhancement in clouds over big data," *Concurrency Computation*, vol. 29, no. 14, Article ID e3909, 2016.

[18] X. Xu, W. Wang, T. Wu, W. Dou, and S. Yu, "A Virtual Machine Scheduling Method for Trade-Offs Between Energy and Performance in Cloud Environment," in *Proceedings of the 4th International Conference on Advanced Cloud and Big Data, CBD 2016*, pp. 246–251, chn, August 2016.

[19] Y. Zhao, Y. Li, S. Lu, I. Raicu, and C. Lin, "Devising a Cloud Scientific Workflow Platform for Big Data," in *Proceedings of the 2014 IEEE World Congress on Services (SERVICES)*, pp. 393–401, Anchorage, AK, USA, June 2014.

[20] L. Qi, X. Xu, X. Zhang et al., "Structural balance theory-based E-commerce recommendation over big rating data," *IEEE Transactions on Big Data*, 2016.

[21] X. Xu, W. Dou, X. Zhang, C. Hu, and J. Chen, "A traffic hotline discovery method over cloud of things using big taxi GPS data," *Software: Practice and Experience*, vol. 47, no. 3, pp. 361–377, 2017.

[22] S. Agarwal, J. Dunagan, N. Jain, S. Saroiu, A. Wolman, and H. Bhogan, "Automated Data Placement for Geo-Distributed Cloud Services. Usenix Symposium on Networked Systems Design and Implementation," in *Proceedings of the NSDI*, pp. 17–32, San Jose , Calif, Usa, 2010.

[23] B. Yu and J. Pan, "Sketch-based data placement among geo-distributed datacenters for cloud storages," in *Proceedings of the 35th Annual IEEE International Conference on Computer Communications, IEEE INFOCOM 2016*, San Francisco, Calif, USA, April 2016.

[24] M. Su, L. Zhang, Y. Wu, K. Chen, and K. Li, "Systematic data placement optimization in multi-cloud storage for complex requirements," *Institute of Electrical and Electronics Engineers. Transactions on Computers*, vol. 65, no. 6, pp. 1964–1977, 2016.

[25] K. Bakshi, "Secure hybrid cloud computing: Approaches and use cases," in *Proceedings of the 2014 IEEE Aerospace Conference*, Big Sky, Mon, USA, March 2014.

[26] L. F. Bittencourt, E. R. M. Madeira, and N. L. S. Da Fonseca, "Scheduling in hybrid clouds," *IEEE Communications Magazine*, vol. 50, no. 9, pp. 42–47, 2012.

[27] Z. Zhou, H. Zhang, X. Du, P. Li, and X. Yu, "Prometheus: privacy-aware data retrieval on hybrid cloud," in *Proceedings of the 32nd IEEE Conference on Computer Communications (IEEE INFOCOM '13)*, pp. 2643–2651, April 2013.

[28] X. Huang and X. Du, "Achieving big data privacy via hybrid cloud," in *Proceedings of the 2014 IEEE Conference on Computer Communications Workshops, INFOCOM WKSHPS 2014*, pp. 512–517, can, May 2014.

[29] J. K. Wang and X. Jia, "Data security and authentication in hybrid cloud computing model," in *Proceedings of the 2012 IEEE Global High Tech Congress on Electronics, GHTCE 2012*, pp. 117–120, chn, November 2012.

[30] H. Abrishami, A. Rezaeian, G. K. Tousi, and M. Naghibzadeh, "Scheduling in hybrid cloud to maintain data privacy," in *Proceedings of the 5th International Conference on Innovative Computing Technology, INTECH 2015*, pp. 83–88, Galicia, Spain, May 2015.

[31] B. Zhou, F. Zhang, J. Wu, and Z. Liu, "Cost Reduction in Hybrid Clouds for Enterprise Computing," in *Proceedings of the 2017 IEEE 37th International Conference on Distributed Computing Systems Workshops (ICDCSW)*, pp. 270–274, Atlanta, GA, USA, June 2017.

[32] X. Qiu, W. L. Yeow, C. Wu, and F. C. M. Lau, "Cost-minimizing preemptive scheduling of mapreduce workloads on hybrid clouds," in *Proceedings of the 2013 IEEE/ACM 21st International Symposium on Quality of Service, IWQoS 2013*, pp. 213–218, Montreal, Canada, June 2013.

[33] A. Zinnen and T. Engel, "Deadline constrained scheduling in hybrid clouds with Gaussian processes," in *Proceedings of the 2011 International Conference on High Performance Computing and Simulation, HPCS 2011*, pp. 294–300, tur, July 2011.

[34] K. Bakshi, "Secure hybrid cloud computing: Approaches and use cases," in *Proceedings of the 2014 IEEE Aerospace Conference*, pp. 1–8, Big Sky, Mon, USA, March 2014.

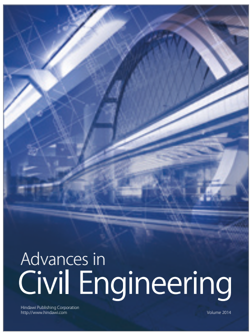
Journal of
Engineering

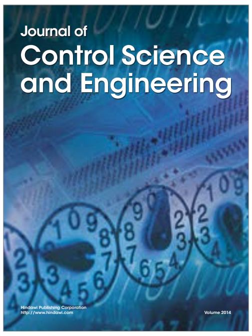
The Scientific World Journal

International Journal of
Rotating Machinery

Journal of
Sensors

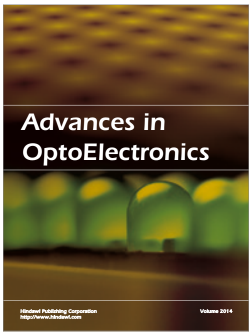
International Journal of
Distributed Sensor Networks


Advances in
Civil Engineering


Journal of
Control Science and Engineering

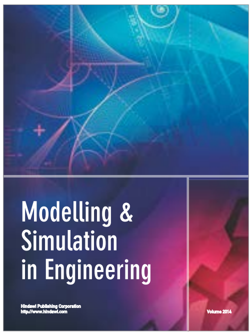
Journal of
Robotics

Journal of
Electrical and Computer Engineering

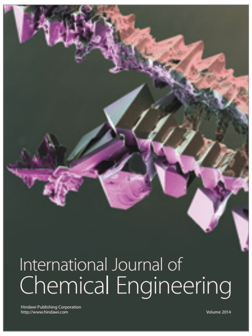

Advances in
OptoElectronics

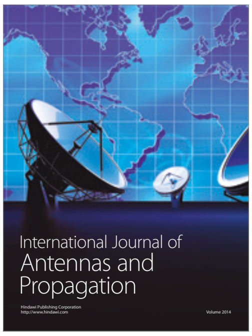
VLSI Design

International Journal of
Navigation and Observation

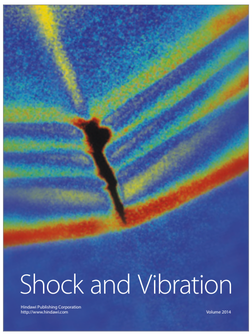

Modelling &
Simulation
in Engineering

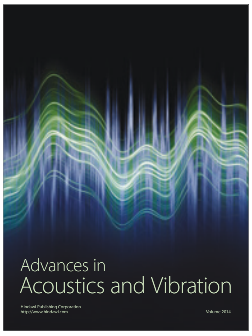
International Journal of
Aerospace Engineering


International Journal of
Chemical Engineering


International Journal of
Antennas and Propagation

Active and Passive
Electronic Components


Shock and Vibration


Advances in
Acoustics and Vibration