

Research Article

A Low Energy Consumption Storage Method for Cloud Video Surveillance Data Based on SLA Classification

Yonghua Xiong,^{1,2} Chengda Lu,¹ Min Wu,¹ Keyuan Jiang,² and Dianhong Wang¹

¹*School of Automation, China University of Geosciences, Wuhan 430074, China*

²*Department of Computer Information Technology & Graphics, Purdue University Calumet, Hammond, IN 46323, USA*

Correspondence should be addressed to Min Wu; wumin@cug.edu.cn

Received 29 October 2015; Revised 4 February 2016; Accepted 29 February 2016

Academic Editor: Ansgar Gerlicher

Copyright © 2016 Yonghua Xiong et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the continuous expansion of the amount of data with time in mobile video applications such as cloud video surveillance (CVS), the increasing energy consumption in video data centers has drawn widespread attention for the past several years. Addressing the issue of reducing energy consumption, we propose a low energy consumption storage method specially designed for CVS systems based on the service level agreement (SLA) classification. A novel SLA with an extra parameter of access time period is proposed and then utilized as a criterion for dividing virtual machines (VMs) and data storage nodes into different classifications. Tasks can be scheduled in real time for running on the homologous VMs and data storage nodes according to their access time periods. Any nodes whose access time periods do not encompass the current time will be placed into the energy saving state while others are in normal state with the capability of undertaking tasks. As a result, overall electric energy consumption in data centers is reduced while the SLA is fulfilled. To evaluate the performance, we compare the method with two related approaches using the Hadoop Distributed File System (HDFS). The results show the superiority and effectiveness of our method.

1. Introduction

Cloud computing is defined by the US National Institute of Standards and Technology (NIST) as a computing model that provides easy access to a shared pool of resources such as computing facilities, storage devices, and software applications through Internet anywhere and anytime [1, 2]. Advancement of cloud computing has been promoting the development of video surveillance and has formed a totally new cloud computing service model called video surveillance as a service [3]. According to the statistics by IMS Research, the demand for video surveillance service based on cloud computing is growing at an annual rate of 20% to 30% [4, 5]. Because of high reliability and availability, cloud video surveillance (CVS) services can help solve the problems encountered with traditional video surveillance technologies including high maintenance costs of communication and storage devices and low level of data security. For example, in [3], the architecture of typical CVS for commercial service based on virtual machine technology is designed to provide users with a kind of “plug and play” monitoring service

with less cost and better convenience, and a task-oriented scheduling method is also proposed subsequently to reduce the energy consumption of the CVS. In [6], a model of video surveillance system based on the network coding distributed file system was established to save storage space.

It is estimated that global mobile data traffic will increase 13-fold between 2012 and 2017, and the video data will be two-thirds of such explosive traffic [7, 8]. Mobile devices such as laptops, tablet PCs, and smartphones are connected by networks that serve end users to get access to large service applications. The CVS system becomes a kind of mobile video application because of the wide use of mobile devices which make it convenient for users to browse, play back, or monitor in real-time video surveillance. The exponential growth of mobile video data, widespread adoption of mobile devices, and demands for video surveillance make the CVS system based on mobile networks a future direction of development.

Given its role of the main carrier of all surveillance video and multimedia data in a CVS system, the video data center is largely a contributing factor of the system performance. Some studies show that the average utilization rate of servers in data

centers is only 20% to 30% [9]. The energy consumption of idle devices is more than 50% of that in a full load [10]. This indicates that currently the power utilization rate of video data centers is low and there is room for improvement and optimization. Low energy efficiency of the data centers also translates to billions of wasted dollars every year [2]. An improvement of technology on data center energy efficiency will bring in significant saving [11]. Different from [6] whose purpose is to save storage space, we address the energy saving problem in this paper.

When users want to use the video service system, they are required to sign a service level agreement (SLA) with the providers. The SLA is used as a contract for users to agree on the cloud service items including video quality, monitoring time, volume of data storage, and pricing. An optimal resource provisioning in the SLA will help reduce the providers operational cost [10]. Our goal in this paper is to develop an energy saving storage method based on an improved SLA for CVS systems. Unlike the task-oriented scheduling method presented in another paper of ours [3], the method presented in this paper focuses on storage node scheduling for energy saving. In practice, both of our methods can be applied to the same data center.

The rest of the paper is organized as follows. Section 2 describes the CVS system and its characteristics of storage energy consumption. Section 3 presents the proposed method together with corresponding deployment and operation strategy. Experiments and conclusions are given in Sections 4 and 5.

2. Related Work

Energy saving of data centers is one of the most challenging problems in the resource-constrained networking field. This problem has attracted a lot of attention in the last few years. Currently, there are two main solutions to the data center energy consumption problem: hardware energy saving technology and software energy saving technology [12]. The hardware energy saving technology mainly reduces the energy consumption through the use of low power devices or adoption of specific hardware and device architecture [12–14]. Obviously, the hardware technology is not flexible and has low portability. In a large-scale commercial cloud video data system, there are numerous different devices. Replacing all devices that are already in use is too expensive and inconvenient. Therefore, the limitation of hardware technology approaches makes it difficult to be implemented widely.

On the other hand, the software energy saving technology focuses on utilizing software strategies to turn off some nodes or put them into sleep state with minimum performance degradation [15]. Comparing with the hardware technology, the software approach is more flexible, much cheaper, and more convenient, receiving more attention. Primarily, there are two storage management methods: dynamic data placement and static data placement.

The dynamic data placement method leverages some strategies or factors to dynamically adjust the location of data

from node to node. For example, the data with a high access frequency will be migrated to some nodes together. Thus, the rest of the nodes will be set idle and can be put into low energy consumption state. In [16], dynamic consolidation of VMs which use live migration is improved. The method is the resource management strategy on “energy-performance tradeoff,” that is, minimizing energy consumption while meeting the SLAs. The proposed deterministic algorithms and adaptive heuristics are claimed to have high efficiency. In [17], an energy saving algorithm is developed based on the storage structure configurations of data blocks. The frame area of data blocks is divided into active and sleep. Each data block is moved to corresponding area according to the access frequency and the blocks in sleep area are switched off to save energy. The approach is flexible and efficient. However, both methods involve the process of data migration which takes a lot of network bandwidth and results in migration delay [18]. The cost of migration will be significant especially when the methods are applied in data intensive applications. Therefore, the dynamic data placement method is not suitable for storing streaming media data and is difficult to implement in CVS system.

On the contrary, the static data placement method does not rely on dynamic resource management such as data migration. It shuts down or suspends redundant nodes by reasonable deployment according to improved schedules or policies. Usually, nodes are turned off in certain time under the premise of having a certain degree of fault tolerance or performance lost [19, 20]. A research group at Stanford University has successfully shut down more idle nodes in Hadoop by improving the strategy of replica placement [19]. But its data availability is to be verified for the CVS system. An energy conservation replica placement method is proposed in [20]. DataNodes are separated into three logical zones according to the forecast results: hot, warm, and cold. Different energy management strategies are adopted for different zones to achieve energy saving. Once the activity frequency of nodes does not match the current zone, the nodes should be transferred to the suitable zone. However, the transfer can only be done step by step among zones, without taking into account the high randomness of user access in a mobile video networking system.

In this paper, we propose an energy saving storage method based on SLA classification for CVS systems. Our approach will divide VMs and data storage nodes into different classifications according to the modified SLA with an extra parameter of access time period. VMs and nodes in different classifications will operate separately at different time period under the guidance of designed schedule. Each time period has the proper number of operating VMs and nodes while redundant ones are shut down to save energy. No data migration is required to consolidate resource. For the aforementioned reasons, our method can be classified as a static data placement method. Unlike hardware energy saving technology, our proposed method is easy to implement and more suitable for mobile video applications than dynamic data placement methods. The effectiveness of our method can be shown in comparison with other static data placement methods.

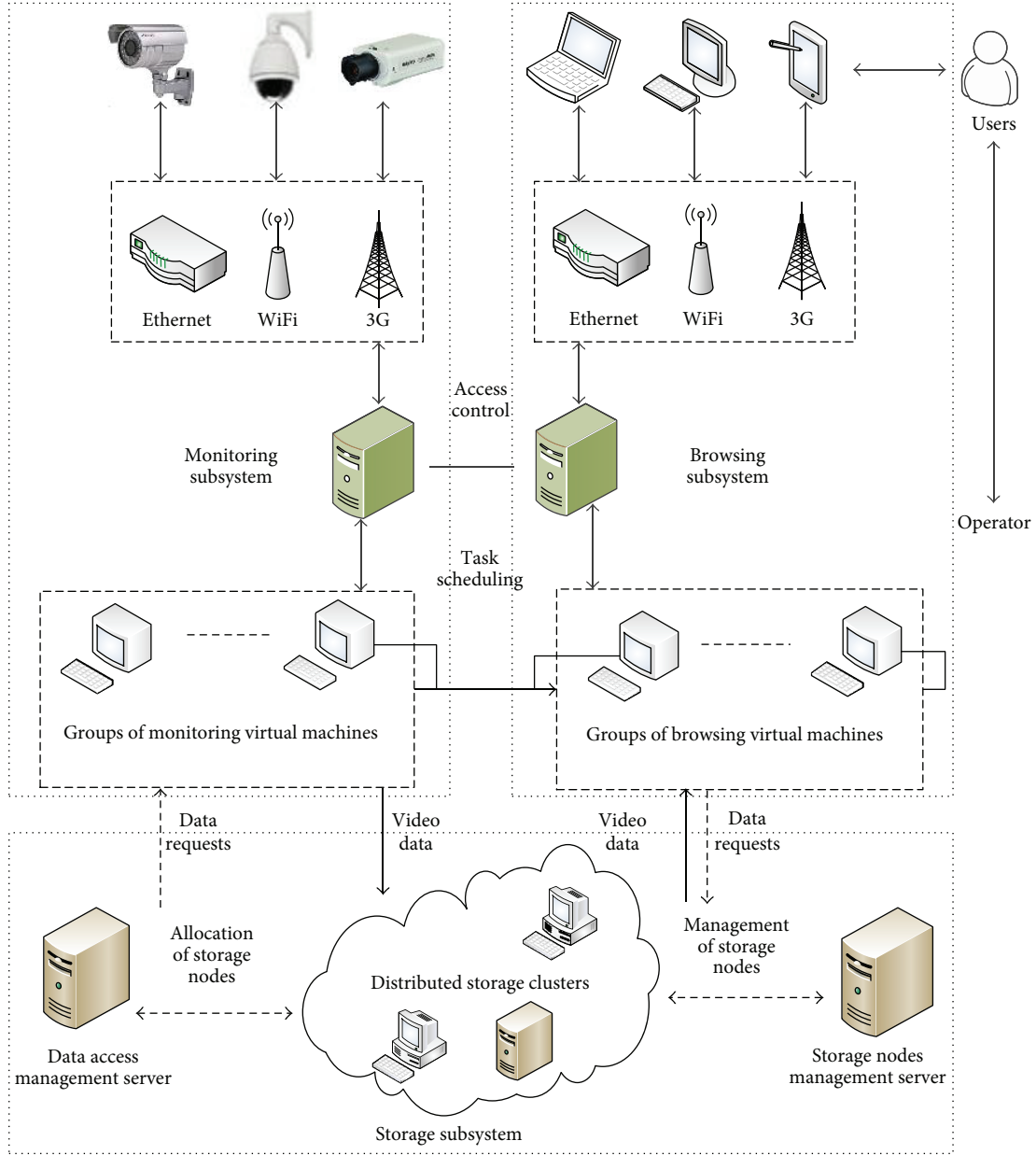


FIGURE 1: Architecture diagram of CVS system.

3. Cloud Video Surveillance System

The video surveillance system has been widely used in military, transportation, production, healthcare, service, and daily life. However, problems exist for an ordinary network video surveillance system: high investment and maintenance costs, limited system scalability from the space, poor security, and low reliability on stored data. These limitations have greatly led to the development of the CVS system, a video surveillance system based on cloud computing.

3.1. System Structure and Operation Mechanism. The CVS system, that is, video surveillance system based on cloud computing, consists of three parts: CVS terminals, a cloud

video data center, and user browsing terminals [21, 22]. The architecture diagram of a CVS system is shown in Figure 1 which is from our prior work [23]. The monitoring subsystem, storage subsystem, and browsing subsystem are corresponding to the CVS terminal, cloud video data center, and user browsing terminal, respectively. In the figure, the webcams in the monitoring subsystem serve as surveillance terminals to obtain surveillance video data. In the browsing subsystem, user can access video data through a browser installed on desktop computers, tablet PCs, and mobile phones for obtaining video data and performing real-time monitoring.

The storage subsystem, as a data center, mainly consists of distributed storage clusters, data access management server,

and storage nodes management server. It is used to store and manage the historical surveillance video data. To make a CVS system operational, the monitoring terminals have to be installed in the areas a user wants to monitor and be registered in the CVS center.

An SLA is an agreement between two service providers or between a service provider and a service consumer. This is an important way to ensure the service quality in various complex network environments so far. An SLA includes a collection of items such as service description, priority, responsibility, guarantee and service level, the availability of services, reaction speed, price, and penalty for agreement violation. It is a document that describes the minimum performance requirements that can be met by service providers when they provide the services.

Different SLA templates are used for different demands [24]. This means that the items in SLA can be arbitrarily expanded according to the negotiation of both sides. Solutions have been developed for services with no matching SLA templates to achieve agreements [25]. There are some negotiable items such as the price and the maximum number of users. Different prices are corresponding to different SLA parameters that provide a range of choices at each level to meet users' customized requirements.

The basic structure and specification of typical SLA are shown as follows:

Property:

Tenant.

Validity Period.

Service description:

Service parameters.

Service level objectives (SLO).

SLA parameters.

Metrics.

Function.

Operand.

Measurement service.

Expression.

Threshold.

Violation Handling:

Predicate.

Action.

Parameters of a typical SLA can be classified into three classes: property, service description, and Violation Handling. Parameters of the property mainly include the Tenant and Validity Period. The Tenant includes the information of ID, the maximum number of users, and the average number of active users. The Validity Period refers to the Validity Period of SLA specifications. Service descriptions of an SLA describe the service quality level or value range, including performance parameters of the SLA, calculation algorithm,

and some necessary information about service provisions. The Violation Handling includes Predicate and Action. The Predicate specifies the operation of each SLA parameter to be calculated. The Action is what is to be taken for SLA violations.

Based on the registered information and SLA, first, the CVS center performs the video monitoring tasks through VMs that are controlled by the access management server, and later the center acquires monitoring data and stores the data in the storage subsystem. When real-time monitoring is needed, the system will directly connect the monitoring VM with the browsing VM. When monitoring historical video is needed, the system will establish a connection between the browsing VM and the storage subsystem to execute the transmission of data and commands. A user can achieve the monitoring purpose through a browser to access the CVS center.

3.2. Analysis of Storage Energy Consumption in CVS System.

In the CVS system, the amount of video data increases linearly with time. Currently, the regular hardware system is unable to meet the needs for the space capacity and storage performance. Meanwhile, it is necessary for the system to deal with huge access requests from users and provide simultaneous services. Therefore, the CVS system must have a strong data storage technology that is of characteristics of scalability, high throughput capacity, high transmission rate, high reliability, and so forth.

The distributed storage technology adopts the scalable topology data organization that is of effective fault tolerance and high resistance to damage. It also includes the efficient distributed router, file cache system, and load balancing algorithm. It can avoid or alleviate the influence of the dynamic and unpredictable nature of network environment and provide consistent high performance storage services to every user in the distributed storage system. Therefore, the video data are generally stored using distributed storage method in the CVS system and the reliability is ensured by the redundancy copies.

The commonly used distributed file systems mainly include HDFS of Apache, OceanStore of UC Berkeley, CFS of MIT, and GFS of Google [26–29]. These storage systems have the characteristics of petabyte storage capacity, easier expandability, high service bandwidth, and so forth. They are effective ways for massive data storage and can meet the storage and access request needs for a CVS system.

With the growth of storage capacity, a large number of CPUs, disk arrays, and numerous cluster servers are continuously running and consuming a lot of energy in the distributed storage system. Power consumption has become the biggest expense in the continuous-operating service. The main reasons that make a CVS distributed storage system inefficient are as follows.

(1) The energy consumption at the idle state is still about 50% of that at the full-load state. An idle state means that there is no storage and access load in the system. In the CVS system, monitoring demands or video data storage may not be needed all day in some scenarios, for example, when videos are being

watched by some people or storage operation is triggered by capturing an object's movement. Numerous nodes are still wasting energy while they have no task.

(2) Redundant configuration in the system designed to meet the needs of peak application will consume a large amount of energy. In the distributed storage, copies of storage data are often added to improve the probability of successful data access. To maintain the accessibility of these data copies, some redundant nodes are required to be run at the normal state that will consume extra energy power. In the distributed storage of cloud video data, the extra energy consumption problem is even worse because more storage space and storage nodes are needed due to the large number of users and larger video data capacity.

(3) The system provided service capability is greater than what the actual applications need. In many cases, distributed systems generally have a redundancy design to ensure the system service's quality. That is, extra nodes in the system are considered as available storage nodes and used, yielding some unwanted energy consumption.

Having considered the above reasons, we add parameters such as access time period and video data storage space into the standard SLA. And a low energy consumption storage method is proposed based on the classification of access time specified in the SLA. The storage method in the CVS data center is optimized to shut down nodes not being used. The system power consumption is reduced at the premise of ensuring the SLA.

4. Low Energy Consumption Storage Method Based on SLA Classification

In addition to general parameters, the SLA of a CVS system should include characteristic parameters of a video surveillance system such as video quality, video data storage space, and scalability. Even though the video quality and storage space parameters are not specifically discussed in the paper, they are a part of the CVS SLA and can be used as reference values to other related designs of CVS. An access time period parameter, what we will mainly discuss, is added in order to limit the time range of the user's monitoring tasks.

4.1. Design of SLA with Access Time Period for Classification. The studies on CVS systems are still at the stage of exploration and trial operation. To the best of our knowledge, there is not a mature SLA for the CVS system at the moment. By analyzing the operation flow of CVS systems, the design of SLA for CVS systems proposed in the paper is shown in Figure 2. Compared to a typical cloud computing SLA shown in Section 3.1, there are the following differences.

An access time period parameter is added to the service description parameters. It is the time period for users to access the surveillance video data. In the CVS system, there is large randomness for users to access the video data because of their needs or preferences. The approximate time range of users' access to video data can be arranged with the access time parameter. According to the parameter, the 24 hours per day can be divided into several common time periods

that can be chosen: "0~4 a.m." as class A, "8 a.m.~12 a.m." as class B, and so forth. When the real-time monitoring is needed, the "all-day" class can be chosen for the access time period. According to the classification, the storage nodes can be labeled correspondingly. In practice, not all users need to perform monitoring twenty-four hours a day, and users can purchase the service of corresponding access time periods according to their needs. During the purchased time period, the system can provide stable and reliable video surveillance services. Therefore, an SLA with access time period not only can, by regulating users' access behavior, make the unified management of resources easier, but also provides users with a more reasonable service with more combinations of selected access time periods.

The video quality and video storage space parameters are added to the service description parameters as well. Video quality can be classified into three grades as high (H), medium (M), and low (L) according to different video resolutions. Since different applications have their own needs such as a user's requirement and network bandwidth, it is not necessary to use the same video quality for all. Therefore, users can get a reasonable choice of the video quality level at the premise of their needs. For example, a high level may be chosen for the building monitoring, a medium level may be chosen for enterprise monitoring that requires a good video quality but is limited by the budget for service, and a low level may be chosen for the urban traffic monitoring service.

It is costly to operate the cloud data center for service providers especially when there is no restriction to users for the storage allocation. The greater the allocated storage space is, the longer the monitoring time of the historical video will be. The video storage space, in a unit of gigabyte, is a very important parameter for both CVS users and service providers. Based on the coding standard and parameters such as video quality and access time period, the storage space for a task can be approximately calculated [3]. That is, the video storage space is a reference parameter provided by the service provider to provision the minimum storage space for a user according to the access time period. It is commonly set at the maximum value for redundancy. Therefore, users cannot directly decide its value but can preset a minimum storage space according to their needs. If the storage space is not preset, it will get the default value calculated by the system. When an adjustment of minimum storage space is required by users, the parameter can be updated. However, the change is not immediate but will take effect only at the next operation time. Once users set the storage space parameter when the SLA is initialized, the overall space for tasks will be calculated. Therefore, purchasing storage space according to the need can not only control users' cost more accurately and reasonably, but also help service providers schedule resources reasonably well based on the storage space parameter in the SLA. This will improve the resource utilization and thus save the energy.

The scalability in SLA parameters is further enhanced. Although scalability is one of the basic characteristics of cloud computing and has its common specifications, it should also include the ability for users to increase or decrease the number of monitoring sites, the monitoring time range, and the video data storage space according to the needs.

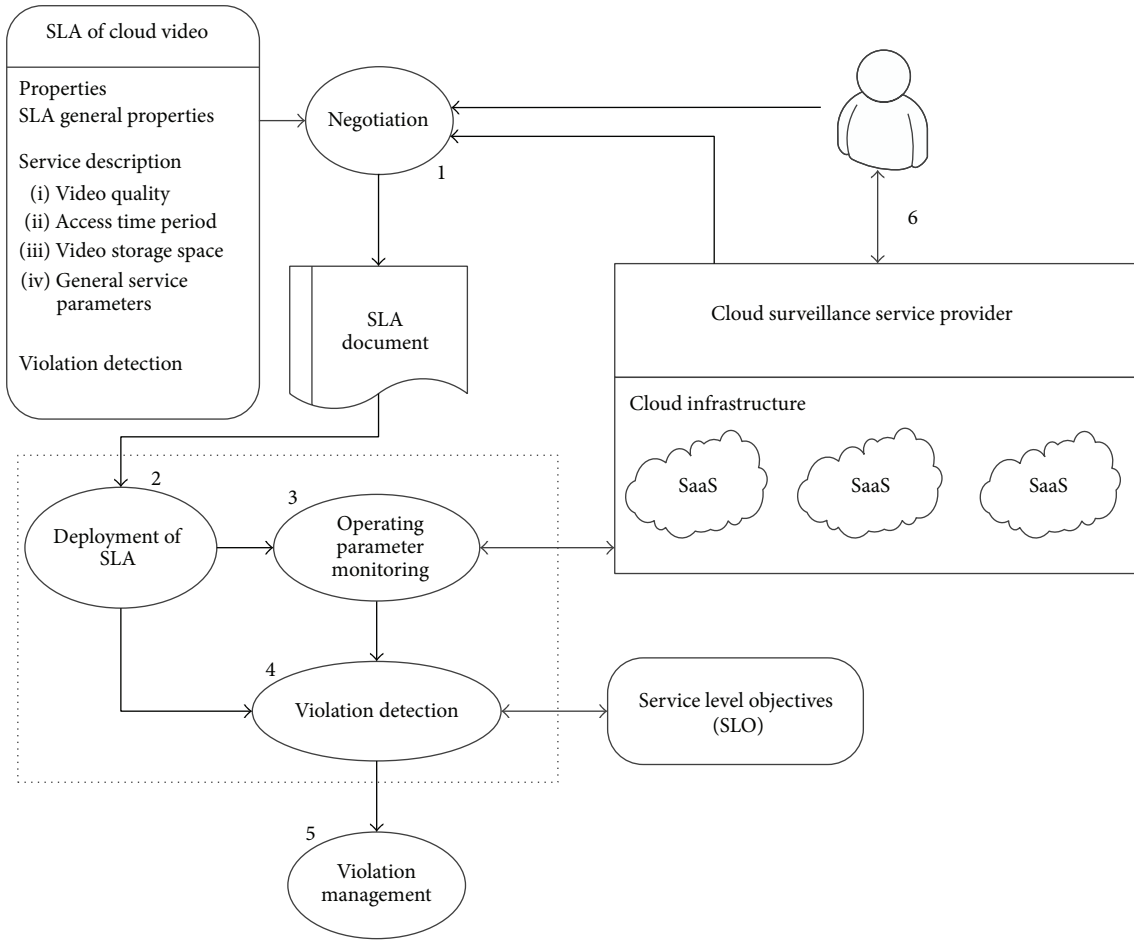


FIGURE 2: SLA framework of CVS.

The SLA deployment and operation management process mainly consists of the following steps.

Step 1 (negotiation). Parameters such as the monitoring time period, level of service quality, cost, and performance should be discussed between CVS service providers and users.

Step 2 (deployment). After the negotiation, SLA should be finalized and deployed to the CVS center and then the SLA monitoring, operation parameter monitoring, and violation management are activated.

Step 3 (operating parameter monitoring). Parameters in the SLA are managed and calculated in real time in the operation process of the CVS system.

Step 4 (violation check). The parameters in SLA are checked against the data measured in Step 3 to make sure that they have met the specified service level objectives.

Step 5 (violation management). The corresponding Action will be taken based on the results of violation check. Either the penalty payment as specified or replacement of other services in the SLA will be requested if violations occur in the process of the service.

Step 6 (termination of service). When the SLA valid period is over, the related instructions are sent to stop the monitoring and storage tasks and the service will be terminated as specified in the SLA.

4.2. Deployment and Resource Management for Storage Method Based on SLA Classification. In the low energy consumption storage method, the distributed storage cluster in a CVS data center is deployed to store and manage the historical video files at first. Later, different servers are deployed in the CVS center to manage the VMs and storage nodes. They include the SLA information management server, resource management server, data access management server, and storage nodes management server. Figure 3 shows the overall structure of the low energy consumption storage method.

The SLA information management server contains all users' access time period information as agreed in the SLA. On the one hand, it gets SLA service demand parameters and other information from the user registration information database server in the cloud surveillance center. On the other hand, it processes the information and passes the processed results as parameters to the resource management server, data access management server, and storage nodes management server at the same time.

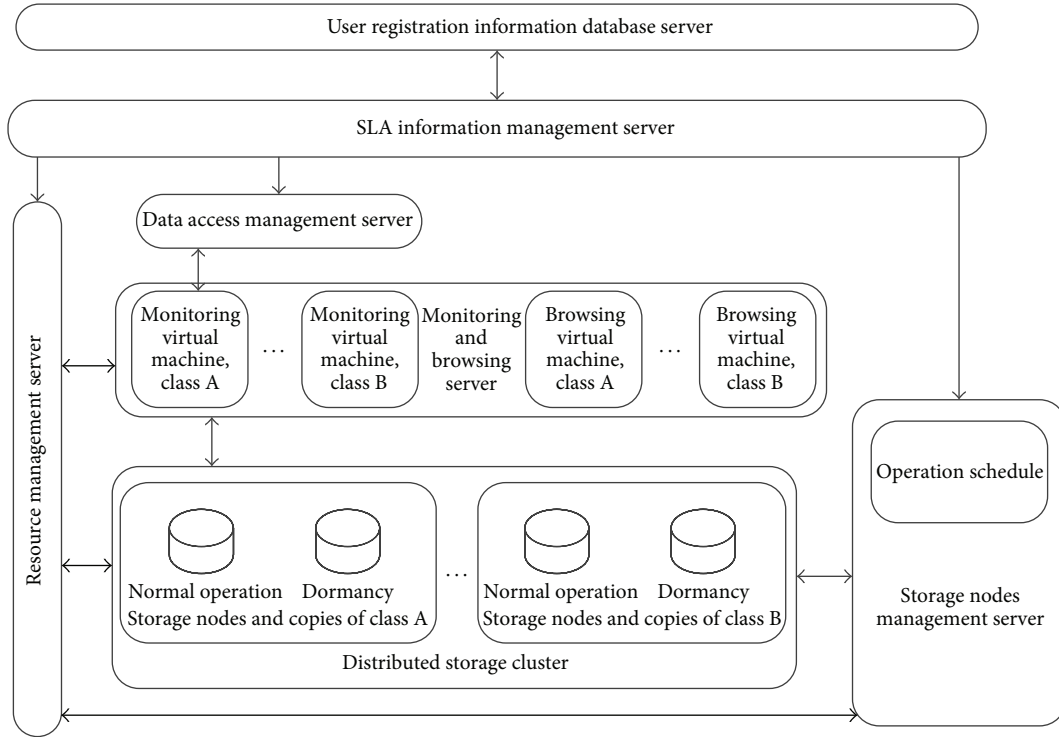


FIGURE 3: Overall structure of the low energy consumption storage method.

The resource management server counts the number of VMs, storage nodes, and remaining storage space in the CVS system. It receives the parameters from the SLA information management server. Afterwards, based on the parameters and the schedule from storage nodes management server, it integrates, manages, and classifies the VMs and the storage nodes in the CVS center.

The data access management server is one of the most important parts in the CVS storage subsystem. It is mainly responsible for sending data storage requests to monitoring VMs and to allocate storage nodes to the corresponding monitoring VMs.

The storage nodes management server is the key to the low energy consumption storage management method because it provides the operation schedule. The schedule is initialized by access time period parameters which are the output of the SLA information management server and the classification information from the resource management server.

Regarding the management of storage nodes, it is obviously improper to classify and manage each storage node alone by the resource management server. Each storage node can be managed through the master node in the distributed storage cluster as long as each master node is associated with the storage nodes in the same subcluster. Then, the resource management server can classify all storage nodes according to the IP address of the master node and the associated storage nodes in the subcluster. The IP address of each master node in each subcluster is combined with the same class of VMs' IP addresses as a subset to be controlled. In this way, all storage

nodes can be managed through those master nodes. For each class of VMs and storage nodes after classification, further fine-grained scheduling management is possible through other management strategies that many research efforts are dedicated to. In this paper, we mainly focus on coarse-grained resource management and the further fine-grained management will be studied and developed in the future study.

4.3. Low Energy Consumption Operation Strategy. Once monitoring or browsing tasks are requested during an access time period, the access control program of the CVS center will select the classified category corresponding to current time period so that the monitoring or browsing tasks can be executed through VMs in the same category. For the historical storage data, the data access management server will periodically send requests to the monitoring VM for data storage access and adjust the categories of storage nodes based on the SLA classification information. The video data with the length up to one or two days (real-time data) are stored in the virtual disks of VMs. The minimum capacity of virtual disks can be estimated based on the maximum bandwidth, video transmission rate, and storage time. However, the historical data are stored in the specific nodes in the network distributed storage so that the static classification method can be applied for unified management.

The operation schedule of the storage nodes in the server is deployed to control the operation mode of data storage nodes in the distributed storage cluster. Based on

```

...
(1) Initial(); //system starts
(2) for  $i$  from 1 to  $r$  do //  $i$ :  $i$ th class based on access time periods
(3)   chooseRandom( $U_i$ ); //  $U_i$ : number of users in each time period.
   end for
...
(4) While() //main loop of CVS system
{
  ...
(5)  If(update) //information updated
  {
(6)   if(current) //new user asks for current service
  {
(7)    New VMs/Nodes added;
(8)    Validity Period recorded;
  }
(9)  else if(! current) //new user does not ask for current service
  {
(10)   for  $i$  from 1 to  $r$  do
(11)     $f_{SLA}(i, U_i, \dots)$ ; //classify VMs and nodes based on updated information.
   end for
  }
}
(12) if(request) //task to access
(13)  if(request ==  $i$ ) //task is relevant to  $i$ th class.
(14)   for  $j$  from 1 to  $dn_i$  do
(15)    Node $_j$  = normal; //Node $_j$ : Nodes of  $i$ th class.
   end for
  else
(16)   other nodes = sleep;
  end if
end if
...
}

```

ALGORITHM 1: Pseudocode of operation strategy.

the analysis in Section 3.2, the pseudocode of operation strategy is illustrated in Algorithm 1.

At first, the service parameters including access time period and number of users in each access time period are treated as inputs to initialize the resource scheduling and allocation program. Thereafter, the VMs and storage nodes are classified into several classes such as A, B, C, by the resource management server in advance. Therefore, users cannot decide when the classification starts. The storage nodes and their replicas in each class of monitoring VMs are assigned to the same category as well.

Once a new user is added and registered in the CVS center, if its access time does not cover the current time, the information management server will update the classification information based on the access time period. Based on the information and the number of users in each time period, the resource management server then reclassifies the monitoring VMs, browsing VMs, and storage nodes. Further, the allocated VMs and storage nodes are adjusted based on the number of users in each class. In addition, the IP addresses of VMs in each category are bound as a group and saved in a storage file on the resource management server.

If the newly initiated user asks for an immediate service, the task will be requested at once. The number of tasks increases in the current access time period while other tasks are still in operation. New VMs and storage nodes in the same category will be open and information is recorded in the parameter of Validity Period. The cost will be different from that of a whole access time period service and adjusted according to the Validity Period parameter in the SLA specification.

An operation schedule is designed for all storage nodes and their replicas in the data center. The storage nodes are defined in two modes: normal and sleep. For example, the storage nodes of class A, between 8 a.m. and 12 a.m., are normal while the storage nodes of class B are in sleep mode.

When a task is requested, the task access control program will send the task request to monitoring or browsing VMs corresponding to the current time period. Then, control commands are sent to the storage nodes according to the operation schedule. Only nodes corresponding to the agreed access time period are set to the normal mode and others are set to the sleep mode.

Once a browsing VM receives a request in the current access time period, data from the storage nodes in the normal

mode are transmitted to the same class of browsing VMs. The service parameters and the operation schedule are updated whenever the access time period information in the SLA is changed.

Suppose the total number of VMs in the CVS center is VM, the total number of storage nodes is DN, and the total number of classifications based on access time periods is r . The number of users in each classification of access time periods is set to U_i , $i \in \{1, 2, 3, \dots, r\}$. The fraction of users in each classification of access time period is set to P_i , $i \in \{1, 2, 3, \dots, r\}$. The number of VMs and the number of storage nodes in each classification of access time period are vm_i and dn_i , respectively, $i \in \{1, 2, 3, \dots, r\}$. Then,

$$P_i = \frac{U_i}{\sum_{i=1}^r U_i}, \quad i \in \{1, 2, 3, \dots, r\},$$

$$vm_i = VM \times P_i = VM \times \frac{U_i}{\sum_{i=1}^r U_i}, \quad (P_i \geq 0), \quad (1)$$

$$dn_i = DN \times P_i = DN \times \frac{U_i}{\sum_{i=1}^r U_i}, \quad (P_i \geq 0).$$

Since there are no monitoring or browsing tasks in VMs before the CVS system starts, the resource management server can randomly select vm_i VMs and dn_i nodes as one class through function ChooseRandom() according to U_i (the number of users in each class) and continue in this way until the classification is completed. Function $f_{SLA}()$ is used to classify VMs and storage nodes according to updated parameters.

The pseudocode is an outline for the system operation. The running time of the algorithm is mainly contributed by the loops from Lines (2) to (3), (10) to (11), and (14) to (15) in the pseudocode, and the computational cost is $O(n)$. Clearly, the code is concise and easy to understand.

There are other details necessary but not shown in the code. When some storage nodes in a classification stop working due to fault or other accidents during operation, the system can still guarantee the availability of data and keep working. The distributed storage system can later recover the invalid replica by duplication and the number of duplications in each category is adjusted according to the number of users in this class.

Since the storage nodes in the sleep mode cannot store historical data, a system process program is needed to receive the access time period parameter from the SLA information management center. Every day, when each class of the storage nodes is running in the corresponding access time period, the process program will initiate data storage requests to the monitoring VMs of the same class. Later, it will store the video data in the virtual disk of the distributed storage system and empty the virtual disk space.

Thus, the video surveillance or browsing tasks can be assigned to the appropriate VMs to execute. The all-day cloud video services will be divided into services in different time periods that can be handled by classification.

5. Experiments and Results

In this paper, the low energy consumption storage method is proposed to improve the power utilization rate in a data center. Therefore, experiments are mainly conducted to investigate the percentage of running storage nodes in the storage system and the average running time of the storage nodes.

For the experiment platform, we use the Hadoop Distributed File System (HDFS) to build the required storage clusters. In the traditional HDFS method, all nodes are located in one HDFS cluster and the default storage management policy, which maintains all data replicas to ensure data availability and prevent data inaccessible problem from unexpected downtime, is used [19, 30]. A low energy consumption storage management method proposed by Stanford University also uses one HDFS cluster but with an improved storage management strategy [19]. In this study, a HDFS cluster is divided into several subclusters whose operation states are managed according to the operation schedule. The storage management in each subcluster still uses the default policy. The results are compared with those from the traditional HDFS method and Stanford University's method.

5.1. Design of Experiment Platform. First of all, the hardware environment and software environment are prepared for the deployment of the storage clusters. The hardware environment includes forty-eight PCs as storage nodes and seven PCs as management nodes called "NameNode." Each node is configured in a Dell Vostro computer (OS: Ubuntu 11.04, RAM: 1 GB, and disk size: 320 GB) with software applications of Hadoop 2.1.0, jdk 1.7, Eclipse 3.7, SSH, and so forth.

The experimental system consists of seven HDFS clusters named A through G. As the reference group, G cluster is used to operate the traditional HDFS method and the Stanford method. All nodes in G cluster use one HDFS system, one NameNode, and twenty-four DataNodes (including replicas).

Clusters A to F are the experimental group that adopts the proposed method. Each cluster has 4 DataNodes (see Table 1). Each node works by following the operation schedule in the storage nodes management server. For the ease of comparison, the number of DataNodes in G and the total DataNodes' number in A~F are the same. The six clusters from A to F are classified according to the SLA access time period. The SLA access time periods are divided into 6 classes for the 24-hour period with 4 hours for each class: class A for the access time between 0 a.m. and 4 a.m., class B for the access time between 4 a.m. and 8 a.m., and so forth.

It should be noticed that, in order to make a storage node react in time, each cluster has a NameNode running continuously. Since the energy consumption of the NameNode is negligible when a large number of storage nodes are involved in an actual data center's distributed storage cluster, only the energy consumption from the running storage nodes is counted in the experiment.

The structure of the experiment platform is shown in Figure 4. The replica number of DataNodes in each cluster is set to 2. We experiment with multiple groups of video storage data (5 G, 10 G, and 15 G) with different time periods.

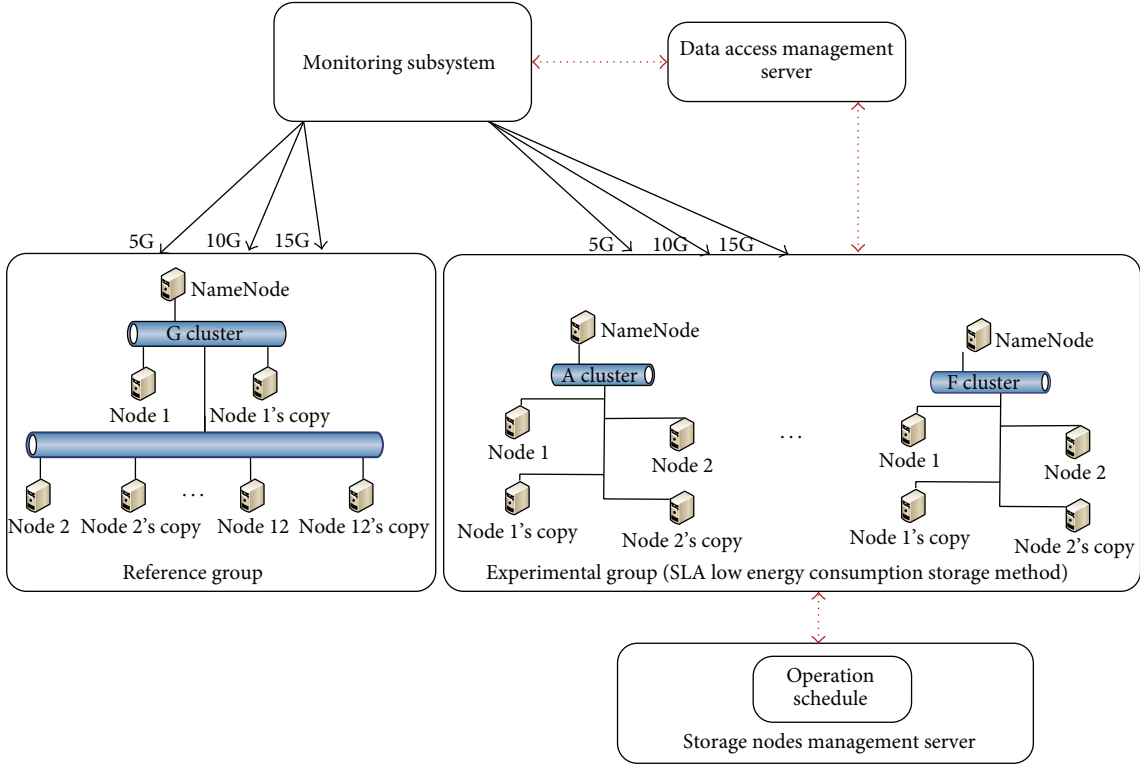


FIGURE 4: Structure of the experimental platform.

TABLE 1: Configuration number of DataNodes in clusters.

	Cluster	DataNodes
Experimental group	A (0:00–4:00)	4
	B (4:00–8:00)	4
	C (8:00–12:00)	4
	D (12:00–16:00)	4
	E (16:00–20:00)	4
	F (20:00–24:00)	4
Reference group	G	24

Reference group is used to operate the traditional HDFS method and the Stanford method while the experimental group uses the proposed method. Only experimental group is divided.

The number of the running storage nodes (DataNode) and the running time in the multiple experiments are compared and analyzed. To facilitate the immediate verification of experimental results, one day is simulated in a three-hour experiment. In other words, the experiment time is compressed to one-eighth of the actual time, and each access time period in classification is reduced to thirty minutes. Meanwhile, we vary the SLA classification number to verify the energy saving effect at the premise of ensuring the equal number of total nodes in the experimental groups (A~F) and in the reference group (G).

5.2. Experiment Results. In the lab environment, the HDFS storage cluster is built to simulate a CVS data center.

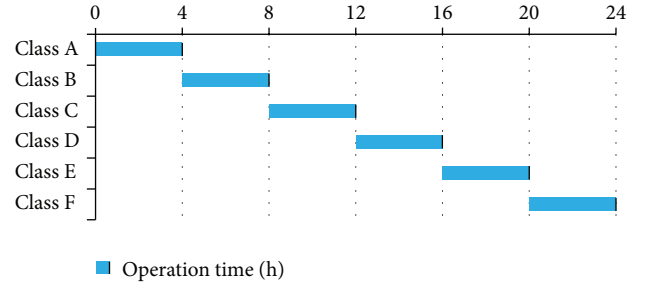


FIGURE 5: Operation schedule of the storage nodes.

The number of storage nodes in the data center and running time of the storage nodes are recorded. Figure 5 shows the operation schedule of the storage nodes for the experiments.

It is clear to see, in Figure 5, the running state of the storage nodes from the operation schedule for the whole 24 hours. At any given time, only the class of storage nodes in access time period is in normal operation state (blue stripe) while other storage nodes not in the access time period are all in the energy saving state.

(1) *Tests for Proportion of Nodes in Operating State.* In the simulated experiments, with the same amount of data and loads, the initial number of storage nodes in the reference group and that in the experimental group are both set to twenty-four. All nodes are in the normal operation state. With six classifications, a day is divided into six time periods for

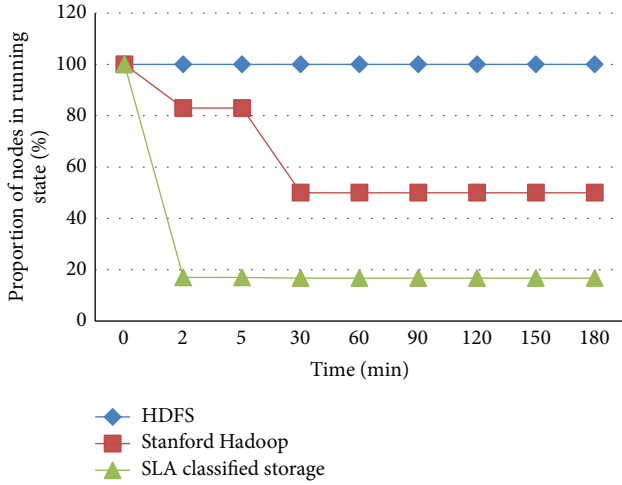


FIGURE 6: Comparison of storage nodes' operating rate.

data access. The storage nodes' running state changes with the execution of the storage operation as shown in Figure 6.

As illustrated in Figure 6, 100% of the storage nodes in HDFS tests (blue curve) are in the running state all the time. This proves the characteristics of the default storage management policy of the traditional HDFS method. In the experiments using the Hadoop cluster with Stanford University's method, about 50% of the storage nodes after the initial 30 minutes are in the running state (red curve). This is because only proper number of nodes is set to run according to the dynamic load analysis. Therefore, to some extent, the number of nodes in running state is reduced. Moreover, the maximum number of sleep nodes is limited to no more than 50% of the total number in order to guarantee the performance and the availability of data.

Although the low energy consumption storage method proposed in this paper needs to maintain the copies as well, the storage nodes and their copies in the data center are divided into a number of categories according to the access time period and only one class of storage nodes and their copies are running at any time. Therefore, it greatly reduces the number of storage nodes in the running state. It can be seen in Figure 6 that after the first 2 minutes only 17% of the storage nodes are needed to run (green curve), significantly lower than the other two methods. In short, our method with classifications results in significant saving in energy consumption.

(2) *Tests for Average Time of Operating Storage Nodes.* By varying the number of SLA classifications, we can test the effects of different SLA classification numbers on energy saving. Figure 7 shows the statistical average operating time of storage nodes versus number of classifications for the previously discussed three methods.

As one can see in Figure 7, nodes in HDFS tests (blue bar) keep running in full capacity regardless of the number of the SLA classifications, and the Stanford Hadoop follows the same trend, maintaining a constant average operating time. With the limitation of maximum number of sleep

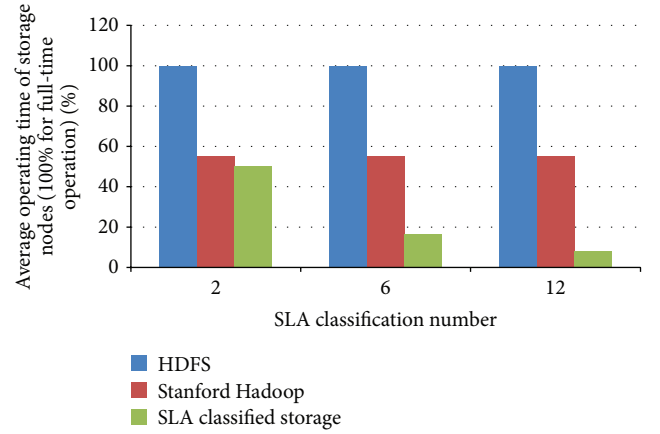


FIGURE 7: Average operating time of storage nodes.

nodes not exceeding 50% of the total, Stanford Hadoop just makes some nodes sleep dynamically depending on different loads. With this in mind, it can be seen that the operating time of the Stanford Hadoop method keeps at about 55% for different SLA classification numbers. Our proposed low energy consumption storage method divides the whole day's tasks into small tasks and operates in batches at different time periods. With the increase of the number of SLA classifications, the operating time of storage nodes in different time period is reduced. Thus, the average operating time of all nodes in the data center is reduced too. As shown in Figure 7, the percentages of average operating time for the proposed method (green bar) are 50%, 17%, and 8.3% for classifications of 2, 6, and 12, respectively. In the experiment, the best case in energy saving is when twelve classifications are used, yielding 41.7% more in saving than the Hadoop method.

The higher the classification number is, the shorter the operating time of each node will be. In practice, if the operating time is too short, it is hard to ensure the completion of a task. For the aforementioned reason, the low energy consumption storage method will have a bottleneck after the classification number reaches a certain threshold. However, its energy efficiency is significantly higher than the other two methods and can be further improved.

6. Conclusion

We have proposed, in this paper, a low energy consumption storage method based on the SLA classification after carefully analyzing the structure of a CVS system, the operating mechanism, and the energy consumption problems in the distributed storage of video data. By designing an SLA with the addition of access time periods, the VMs and the storage nodes for monitoring operation and browsing tasks can be reasonably classified. Therefore, all-day CVS services are divided into various tasks associated with different time periods and processed separately. By designing an operation schedule for all storage nodes in the data center, the operation state of storage nodes can be effectively controlled to achieve energy saving. The deployment process and operation

strategy for the low energy consumption storage method are described in detail in the paper.

Experimental results show that the current method only uses 17% of the storage nodes in normal state and takes 8.3% average operating time with twelve classifications, resulting in more energy efficiency than the full load of traditional HDFS method and about half load of the Stanford improved HDFS method. In addition, the method is easy to implement and definitely offers certain application value. Our method demonstrates the following advantages:

- (i) It is specially designed for a CVS system and is easy to adopt widely and deploy for other systems including cloud computing, cloud disk, or other mobile applications such as mobile UGC (user generated content) and mobile advertisement.
- (ii) It can meet the performance requirement while minimizing energy consumption.
- (iii) Compared with traditional and Stanford improved HDFS methods, the proposed approach proves its superiority and effectiveness.

In the future work, we will mainly aim at developing and improving a fine-grained scheduling management strategy for each storage node.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

This work was supported in part by the National Nature Science Foundation of China under Grant no. 61202340, the International Postdoctoral Exchange Fellowship Program under Grant no. 20140011, and the Hubei Provincial Natural Science Foundation of China under Grant no. 2015CFA010.

References

- [1] G. Aceto, A. Botta, W. De Donato, and A. Pescapè, "Cloud monitoring: a survey," *Computer Networks*, vol. 57, no. 9, pp. 2093–2115, 2013.
- [2] T. Mastelic, A. Oleksiak, H. Claussen et al., "Cloud computing: survey on energy efficiency," *ACM Computing Surveys (CSUR)*, vol. 47, no. 2, p. 33, 2014.
- [3] Y. Xiong, S. Wan, J. She, M. Wu, Y. He, and K. Jiang, "An energy-optimization-based method of task scheduling for a cloud video surveillance center," *Journal of Network and Computer Applications*, vol. 59, pp. 63–73, 2016.
- [4] P. Simoens, F. De Turck, B. Dhoedt, and P. Demeester, "Remote display solutions for mobile cloud computing," *Computer*, vol. 44, no. 8, pp. 46–53, 2011.
- [5] 2013, <http://www.ctiforum.com/news/guandian/379338.html>.
- [6] Q. Zhao, K. Li, H. Li, Y. Wang, and H. You, "Model of cloud video surveillance based on network coding cloud storage system," *International Journal of Future Computer and Communication*, vol. 4, no. 4, pp. 286–289, 2015.
- [7] T. De Pessemier, L. Martens, and W. Joseph, "Dynamic optimization of the quality of experience during mobile video watching," in *Proceedings of the 10th IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB '15)*, pp. 1–6, IEEE, Ghent, Belgium, June 2015.
- [8] J. Wu, B. Cheng, C. Yuen, Y. Shang, and J. Chen, "Distortion-aware concurrent multipath transfer for mobile video streaming in heterogeneous wireless networks," *IEEE Transactions on Mobile Computing*, vol. 14, no. 4, pp. 688–701, 2015.
- [9] L. Na, M. Zhang, K. Li et al., "Green task scheduling algorithms with speeds optimization on heterogeneous cloud servers," in *Proceedings of the IEEE/ACM International Conference on Green Computing and Communications & International Conference on Cyber, Physical and Social Computing*, pp. 76–80, Hangzhou, China, December 2010.
- [10] H. Goudarzi, M. Ghasemazar, and M. Pedram, "SLA-based optimization of power and migration cost in cloud computing," in *Proceedings of the 12th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid '12)*, pp. 172–179, IEEE, Ottawa, Canada, May 2012.
- [11] Q. Zhang, M. F. Zhani, S. Zhang, Q. Zhu, R. Boutaba, and J. L. Hellerstein, "Dynamic energy-aware capacity provisioning for cloud computing environments," in *Proceedings of the 9th ACM International Conference on Autonomic Computing (ICAC '12)*, pp. 145–154, ACM, San Jose, Calif, USA, September 2012.
- [12] Y.-J. Wang, W.-D. Sun, S. Zhou, X.-Q. Pei, and X.-Y. Li, "Key technologies of distributed storage for cloud computing," *Journal of Software*, vol. 23, no. 4, pp. 962–986, 2012.
- [13] X. Q. Luo, Y. Qi, L. Wang, and Q. Wang, "Dynamic power dissipation control method for real-time processors based on hardware multithreading," *China Communications*, vol. 10, no. 5, Article ID 6520948, pp. 156–166, 2013.
- [14] H. Shim, J.-S. Kim, and S. Maeng, "BEST: best-effort energy saving techniques for NAND flash-based hybrid storage," *IEEE Transactions on Consumer Electronics*, vol. 58, no. 3, pp. 841–848, 2012.
- [15] H. Goudarzi and M. Pedram, "Energy-efficient virtual machine replication and placement in a cloud computing system," in *Proceedings of the IEEE 5th International Conference on Cloud Computing (CLOUD '12)*, pp. 750–757, IEEE, Honolulu, Hawaii, USA, June 2012.
- [16] A. Beloglazov and R. Buyya, "Optimal online deterministic algorithms and adaptive heuristics for energy and performance efficient dynamic consolidation of virtual machines in cloud data centers," *Concurrency Computation Practice and Experience*, vol. 24, no. 13, pp. 1397–1420, 2012.
- [17] B. Liao, J. Yu, T. Zhang, and X.-Y. Yang, "Energy-efficient algorithms for distributed file system HDFS," *Chinese Journal of Computers*, vol. 36, no. 5, pp. 1047–1064, 2013.
- [18] D. Kliazovich, P. Bouvry, and S. U. Khan, "DENS: data center energy-efficient network-aware scheduling," *Cluster Computing*, vol. 16, no. 1, pp. 65–75, 2013.
- [19] J. B. Leverich, *Future Scaling of Datacenter Power-Efficiency*, Stanford University, 2014.
- [20] X.-Y. Zhao and L. Wang, "An activity-based replica placement method of energy-conservation," in *Proceedings of the IEEE International Conference on Fuzzy Theory and Its Applications (iFUZZY '13)*, pp. 446–451, Taipei, Taiwan, December 2013.
- [21] M. S. Hossain, M. M. Hassan, M. A. Qurishi, and A. Alghamdi, "Resource allocation for service composition in cloud-based

- video surveillance platform,” in *Proceedings of the IEEE International Conference on Multimedia and Expo Workshops (ICMEW '12)*, pp. 408–412, Melbourne, Australia, July 2012.
- [22] June 2013, <http://www.cloudsurveillance.com>.
- [23] Y.-H. Xiong, S.-Y. Wan, Y. He, and D. Su, “Design and implementation of a prototype cloud video surveillance system,” *Journal of Advanced Computational Intelligence and Intelligent Informatics*, vol. 18, no. 1, pp. 40–47, 2014.
- [24] S. Yangui, I.-J. Marshall, J.-P. Laisne, and S. Tata, “CompatibleOne: the open source cloud broker,” *Journal of Grid Computing*, vol. 12, no. 1, pp. 93–109, 2014.
- [25] A. Kertesz, G. Kecskemeti, and I. Brandic, “An interoperable and self-adaptive approach for SLA-based service virtualization in heterogeneous Cloud environments,” *Future Generation Computer Systems*, vol. 32, no. 1, pp. 54–68, 2014.
- [26] Y. N. Wang, S. D. Zhang, and H. Liu, “The design of distributed file system based on HDFS,” *Applied Mechanics and Materials*, vol. 423–426, pp. 2733–2736, 2013.
- [27] A. Dandoush, S. Alouf, and P. Nain, “Lifetime and availability of data stored on a P2P system: evaluation of redundancy and recovery schemes,” *Computer Networks*, vol. 64, pp. 243–260, 2014.
- [28] D. Wang, Z. Hu, J. Zhang, L. Zhang, and Z. Wen, “CFS: the design and implementation of a cluster file system service on Inspur AS3000,” in *Proceedings of the International Conference on Computational and Information Sciences (ICCIS '11)*, pp. 847–849, Chengdu, China, October 2011.
- [29] X. Y. Zhou and L.-Y. He, “A virtualized hybrid distributed file system,” in *Proceedings of the 5th International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC '13)*, pp. 202–205, Beijing, China, October 2013.
- [30] D. Borthakur, “The hadoop distributed file system: architecture and design,” *Hadoop Project Website*, vol. 11, no. 2007, article 21, 2007.

