# Optimal Scheduling Model of Cloud Storage Resources Based on Double Threshold Load Balancing Control

**Liuchun Zhan[1,a], Changjiang Huang[2,b], Mei Lin[3,c]**

[1]Huali College Guangdong University of Technology,Guangzhou,Guangdong
[2]Guangzhou University Sontan College,Guangzhou,Guangdong
[3]Nanfang Vocational College,Jiangmen Guangdong
[a] zhanlc@163.com,[b]huangchangjiang8@163.com,[c]linmei@163.com

**Abstract**. Effective resource scheduling design in cloud storage system can integrate cloud computing central resources and improve resource optimal allocation ability. An optimal scheduling method for cloud storage resources is proposed based on two-threshold load balancing control. The spatial distribution of storage resources is optimized by using grid region collocation method, and the model of cloud storage resource distribution information flow is constructed by nonlinear time series reorganization method, and the auto-correlation feature matching of cloud storage resource information flow is carried out. The association rule feature quantity of cloud storage resource is mined, the phase space of cloud storage system is reconstructed according to the distribution attribute of association rule feature, the priority attribute list of resource scheduling is constructed, and segment interpolation method is adopted to allocate cloud resource. The threshold value of resource scheduling is determined by adaptive optimization of cloud resource allocation threshold using piecewise interpolation method, and load balancing control of cloud resource allocation is carried out by double threshold control method. The partition scheduling of cloud storage resources is implemented according to the priority list. The simulation results show that the proposed method has good load balance, high precision and high throughput, and it has good application value in resource optimization.

## 1. Introduction

In the cloud computing environment, in order to improve the efficiency of data storage scheduling and provide users with more data search and utilization experience, it is necessary to optimize the design of cloud storage resource scheduling. Task scheduling and data cloud storage, cloud scheduling and cloud management technology are used to coordinate and integrate computer information resources and computing tasks. Data cloud storage is the basis for implementing cloud computing. In the cloud storage framework model, resources are integrated and adaptive scheduling to improve the optimal allocation and storage capacity of resources. The scheduling model has a good application value in the design of cloud storage system and data optimization management[1].

Cloud computing technology has accumulated a large amount of data resources for a long time, so it reduces the effectiveness of people's search, and also reduces the real-time performance of data search, in order to improve the efficiency of data storage scheduling[2]. It is necessary to establish a cloud storage resource scheduling model based on intelligent hierarchical storage and design a new data storage scheduling algorithm to improve resource scheduling ability. In the design of cloud storage resource scheduling algorithm, cluster energy consumption scheduling algorithm, grid computing method and priority scheduling method are used to realize cloud storage resource scheduling with cyclic stack control[3]. In reference [4], a cloud storage resource scheduling algorithm based on feature scale decomposition is proposed, which extracts the temporal scale features of cloud storage resources, recombines feature vectors in phase space, and divides the time axis of cloud storage resource scheduling into uniform grid space. Resource scheduling in grid space can improve the throughput performance of scheduling, but the computational overhead of this method is high, and the real-time performance of cloud resource scheduling is not good. In reference [5], a cloud storage resource scheduling algorithm based on adaptive energy recharge and cyclic stack control is proposed, which is based on linear coding control method. This method has some problems such as weak anti-jamming ability and poor control ability of adaptive equalization.

In order to solve the above problems, an optimal scheduling method for cloud storage resources is proposed based on double threshold load balancing control. The spatial distribution of storage resources is optimized by using grid region collocation method, and the model of cloud storage resource distribution information flow is constructed by nonlinear time series reorganization method, and the auto-correlation feature matching of cloud storage resource information flow is carried out. The threshold value of resource scheduling is determined by adaptive optimization of cloud resource allocation threshold using piecewise interpolation method, and load balancing control of cloud resource allocation is carried out by double threshold control method. The partition scheduling of cloud storage resources is implemented according to the priority list. Finally, a simulation experiment is carried out to demonstrate the superior performance of the proposed method in improving the scheduling capability of cloud storage resources.

## 2. Information flow model construction and feature extraction of cloud storage resource scheduling

### 2.1. Distribution structure model of Cloud Storage Resources

In order to realize cloud storage resource scheduling under big data environment, firstly, the kernel structure and resource storage structure model of the network are analyzed, and the information flow model and time series analysis of cloud storage resource are carried out. The ontology model of cloud storage resource distribution is expressed as $|d_{n-\max} - d_{n-\min}| \cdot (1/K)$. By using adaptive equalization control method, the cloud storage space is divided into several $k$ data subsets $A_k$, which is the feature distribution set of association rules for cloud storage resources under big data environment[6]. The association rule set of cloud storage resource flow meets $A_1 \cup A_2 \cup, \cdots, \cup A_k = A$. According to the distribution attribute of cloud storage resource, the grid model of resource distribution interval is constructed, which makes the initial quantized feature set of cloud storage resource distribution satisfy $A_i \cap A_j = \Omega$, where $i, j = 1, \cdots, m$, $i \neq j$. According to the scale scaling and time translation characteristics of cloud storage resource data flow, the frequent itemset of resource distribution are obtained as follows:

$$P = \{p_1, p_2, \cdots p_m\}, m \in N \tag{1}$$

For a cloud storage system, the data representing the $v_i$ task is provided by the resource node represented by the value $v_i$, where $\beta$ is the average information coefficient of the source[7]. Under the condition of fixed $\theta$, the upper expression is derived from the vector $\zeta$. P and Q denote two probability distribution functions and make them zero, as:

$$\zeta = P\beta Q^{+}(\theta) wQ \tag{2}$$

For the storage overhead of computing resource scheduling in multiple resource scheduling task flows, in a cloud storage embedded system[8], a workstation is equivalent to a processor, and the task flow to be assigned is:

$$flow_k = \{n_1, n_2, \cdots, n_q\}, q \in N \tag{3}$$

IOP Conf. Series: Materials Science and Engineering **392** (2018) 062205 doi:10.1088/1757-899X/392/6/062205

In the above formula, $q$ represents the node position of multiple task flow sets, $n_q$ represents the data sequence of task information flow, and $N$ represents the total number of tasks. In the above model of cloud storage resource distribution structure, the optimal allocation of resources is carried out, and the optimization design of resource scheduling is carried out by combining the methods of feature extraction and information fusion[9].

### 2.2. Information flow structure and feature matching of cloud storage resources

A cloud storage resource distribution information flow model with $m$ input control parameters is constructed by nonlinear time series reorganization method[10]. The $n$ output parameter cluster cloud storage resource scheduling model is expressed as follows:

$$\begin{cases} x=(x_1,x_2,...,x_n) \\ y=F(x)=(f_1(x),f_1(x),...,f_m(x))^T \end{cases} \tag{4}$$

Where, $x=(x_1,x_2,...,x_n)$ is the set of assigned nodes for resource scheduling in cloud storage system. Assuming the type of resource scheduling vector set $n_i$ is $y=F(x)$, then the resource allocation set $P(n_i)=\{p_k \mid pr_{kj}=1, k=1,2,\cdots,m\}$ represents the priority attribute of task flow, and defines RTT(Round-Trip time) as the active factor of cluster head node $RTT_s$. The activity coefficient of degree task set is:

$$RTT_s=(1-\alpha)\times RTT_s+\alpha\times RTT \tag{5}$$

The priority control method is used for the adaptive scheduling of cloud storage resource scheduling. $M$ is a $d$ dimension cloud storage resource scheduling resource classification set. $F$ is expressed as a vector with smoothness. For $\Phi: M\rightarrow R^{2d+1}$, there are active factors of point $N$, combined with the method of feature decomposition, the activity coefficient of scheduling task set is obtained.

$$\Phi(z)=(h(z),h(\varphi_1(z)),\cdots,h(\varphi_{2d}(z)))^T \tag{6}$$

The flow of scheduling tasks to be assigned at this point is:

$$flow_k=\{n_1,n_2,\cdots,n_q\}, q\in N \tag{7}$$

In the formula, $q$ denotes the different characteristics of multiple task flows, and $n_q$ represents the data sequence of real-time task information flow in the neighbor hopping node[11]. When $\omega_n=0$, the vector quantization output of the cloud storage resource scheduling sequence $\{x_n\}_{n=1}^N$ of the cloud storage system is obtained as follows:

$$x_n=\left[x(0),x(1),\cdots,x(N-1)\right]^T \tag{8}$$

The transmission scheduling data set of resource scheduling is $X=\{x_1,x_2,\cdots,x_n\}$, and $n$ is the number of slot allocation nodes $X$. Each element in $X$ is a $P$-dimensional vector. In this way, auto-correlation feature matching of cloud storage resource information flow is realized[12].

## 3. Improved design of resource scheduling algorithm

### 3.1. Mining association rule feature quantities for cloud storage resources

The distributed information flow model of cloud storage resources is constructed by using nonlinear time series recombination method. Based on the auto-correlation feature matching design of cloud storage resource information flow, the improved cloud storage resource scheduling algorithm is designed. This paper presents an optimal scheduling method for cloud storage resources based on two-threshold load balancing control[13]. The cloud storage system is taken based on the information of the arrival frequency and execution time of each type of tasks in the past, and obtains the characteristic scale as:

$$\begin{aligned} \min \quad & F(x) = (f_1(x), f_2(x), ..., f_m(x))^T \\ s.t. \quad & g_i \le 0, \qquad i = 1,2,...,q \\ & h_j = 0, \qquad j = 1,2,...,p \end{aligned} \tag{9}$$

The eigenvector of scale equilibrium input by cloud storage resource scheduling processor is cluster cloud storage resource scheduling set[14]. The mining output of association rule feature quantity of cloud storage resource is expressed as follows:

$$S = (U, A, V, f),\ P \cap Q = f,\ A = P \cup Q \tag{10}$$

In the above formula, $U$ denotes the number of tasks in $T$, $A$ denotes the eigenscale vector space, $V$ denotes the length of idle time slice, and $f$ is the random weighted vector of the cluster cloud storage resource scheduling set S. According to the result of feature mining, the threshold control of scheduling is carried out to improve the ability of balanced allocation of resource scheduling.

### *3.2. Load balancing control of cloud resource allocation*

The priority attribute list of resource scheduling is constructed, and the load balance control of cloud resource allocation is carried out by using double threshold control method. The task node and resource integration cluster center are covered seamlessly, in which real-time task scheduling is carried out[15]. The Sink nodes in cloud computing are divided into rule sets and instance sets $S = \{s_i \mid i = 1,2,\cdots,N_S\}(S \subset C)$, each cloud computing node has a clock, and the multidimensional performance metrics between nodes are defined as:

$$E = [E_G, E_T, E_W, E_L] \tag{11}$$

Where, the scheduling state vector feature of each task transport node $C \subset S$ can cover k-cycle grid according to the communication time and scheduling period between the tasks. Through load balancing estimation, the balanced crossover of task $n_j$ performed by the cloud computing process management processor attenuates is shown as follows:

$$S_p(u) = \{F^p[s(t)]\}(u) = \int_{-\infty}^{\infty} K_p(t,u)s(t)dt \tag{12}$$

Thus, a list of priority attributes for resource scheduling is constructed as shown in figure 1.

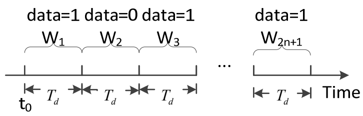


Figure 1 List of priority attributes for resource scheduling

According to the priority attribute list of resource scheduling, resource adaptive optimal configuration and equalization control are carried out to improve the balance of scheduling.

The length of each idle time slice of cloud storage resource scheduling is recorded as $x_1, x_2, \cdots, x_{m+1}$, and the following equations can be obtained:

$$x_1 + x_2 + \cdots + x_{m+1} = T + t - m \times t \tag{13}$$

The adaptive optimization of cloud resource allocation threshold is realized by piecewise interpolation method, and the association rule feature reconstruction of cloud storage resource information storage space in network environment is realized:

$$H_i(x) = \sum_{k=1}^{K} p_k \ln \frac{1}{p_k} = -\sum_{k=1}^{K} p_k \ln p_k \tag{14}$$

According to the reconstruction results of cloud storage resource information association rules, the dual threshold model of energy allocation can be expressed as:

$$\mathrm{AT}(v) = \max_{u \in FI(v)} [AT(u) + delay(v)] \tag{15}$$

$$\mathrm{RT}(v) = \min_{w \in FO(v)} [RT(w) - delay(v)] \tag{16}$$

Finally, according to the HList returned to IC Compiler for cloud storage resource scheduling,

IOP Conf. Series: Materials Science and Engineering **392** (2018) 062205 doi:10.1088/1757-899X/392/6/062205

according to the timing information of each node, the scheduling boundary of dual threshold scheduling for the task node of the whole cloud storage system is obtained as follows:

$$bnr_\beta\left(X\right)=R_\beta X-R_\beta X_1 \tag{17}$$

Therefore, under the condition of characteristic scale equilibrium, the threshold discriminant of cloud storage resource scheduling is obtained as follows:

$$\sum_{m=1}^{n}x_G^m\le E_G,\sum_{m=1}^{n}x_T^m\le E_T,\sum_{m=1}^{n}x_W^m\le E_W,\sum_{m=1}^{n}x_L^m\le E_L \tag{18}$$

By double threshold decision, the scheduling efficiency of the whole task can be improved at different nodes. The total efficiency optimization is: $\sum_{\sigma}\mu^{m\omega}T_\sigma^\omega$ , where $\omega\subseteq\{G,T,W,L\}$ , $m\in[1,n]$. The time axis is divided into an adjacent but not overlapping window $W_i$ , and the resources are configured in the window to realize the optimal scheduling of cloud resources.

## 4. Simulation experiment and result analysis

In order to test the performance of this algorithm in cloud computing environment, the simulation experiment is carried out. The experiment adopts Matlab 7 design, uses random function signal generator to generate 1024 real-time cloud storage resource scheduling information state set randomly, according to the task automatically adjusts the characteristic scale, constructs JDK 1.6, a1Gbps exchange network in the NS-2.27 cloud computing platform, it simulates 300 real-time task requests. The number of cloud storage resources is nearly 2 000. The data discrete sampling rate of cloud storage resource is $f_s=10*f_0\text{Hz}=10\text{KHz}$, and the length of monitoring data is 1000. The time slot allocation diagram of resource scheduling is shown in figure 2.

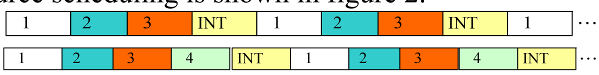


Figure 2 Slot allocation diagram of cloud storage resource scheduling

According to the above simulation environment and parameter setting, the cloud storage resource scheduling simulation is carried out, and the time domain waveform of cloud storage resource information sampling is obtained as shown in figure 3.

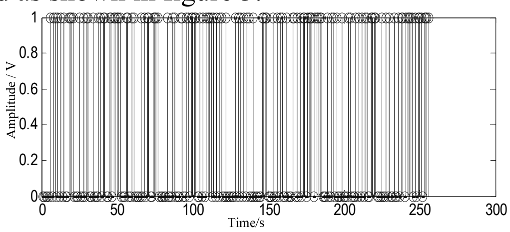


Figure 3 Time domain waveform of cloud storage resource sampling

The cloud storage resources in figure 3 is taken as the research object, and the load balancing control is used to optimize the resource allocation. The scheduling output is shown in figure 4.

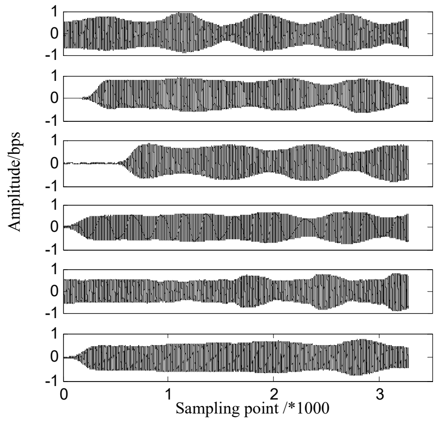

Figure 4 Balanced scheduling output of cloud storage resources

Figure 4 shows that the proposed method has a good balance and has a strong ability of multi-channel resource allocation. The accuracy of different methods for resource scheduling is tested. The comparison results are shown in figure 5. The analysis figure 5 shows that the accuracy of this method for resource scheduling in cloud storage is good, the precision of resource scheduling output is high, and the throughput is large.

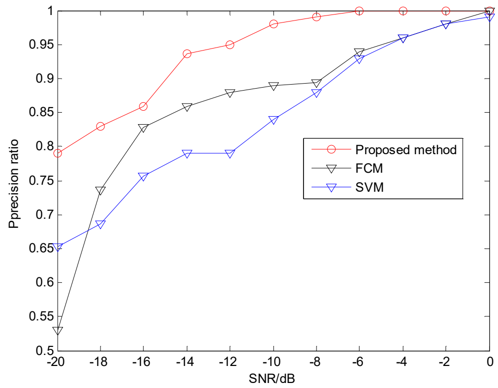

Figure 5 Performance comparison

## 5. Conclusions

In this paper, it presents an optimal scheduling method for cloud storage resources based on double threshold load balancing control. The spatial distribution of storage resources is optimized by using grid region collocation method, and the model of cloud storage resource distribution information flow is constructed by nonlinear time series reorganization method, and the auto-correlation feature matching of cloud storage resource information flow is carried out. The association rule feature quantity of cloud storage resource is mined, the phase space of cloud storage system is reconstructed according to the distribution attribute of association rule feature, the priority attribute list of resource scheduling is constructed, and segment interpolation method is adopted to allocate cloud resource. The threshold is adaptively optimized and the threshold of resource scheduling is determined. The dual threshold control method is used to control the load balance of cloud resource allocation and the partition scheduling of cloud storage resources is realized according to the priority list. The research shows that this method has good load balance, high precision rate and large throughput of resource scheduling output, and this method has good application value in resource optimization configuration.

## References

[1] XIAO Wen, HU Juan. Performance analysis of frequent itemset mining algorithms based on sparseness of dataset[J]. Journal of Computer Applications, 2018, 38(4): 995-1000.

IOP Conf. Series: Materials Science and Engineering **392** (2018) 062205 doi:10.1088/1757-899X/392/6/062205

[2] MAO W T, TIAN Y Y, WANG J W, et al. Granular extreme learning machine for sequential imbalanced data[J]. Control and Decision, 2016, 31(12):2147-2154.
[3] GU Q, YUAN L, NING B, et al. A noval classification algorithm for imbalanced datasets based on hybrid resampling strategy[J]. Computer Engineering and Science, 2012, 34(10):128-134.
[4] GUO Huaping, ZHOU Jun, WU Chang'an, FAN Ming. k-nearest neighbor classification method for class-imbalanced problem[J]. Journal of Computer Applications, 2018, 38(4): 955-959.
[5] CUN Yong-jun, ZHNAG Yong-hua. Linux System Dual Threshold Scheduling Algorithm Based on Characteristic Scale Equilibrium[J]. Computer Science, 2015,42(6):181-184.
[6] MA Yu CAI Yuan-li. Scheduled offline model predictive control based on multiple LPV models[J]. Control and Decision, 2016, 31(08): 1468-1474.
[7] WANG Cheng, ZHAO Bifang.Network Intrusion Detection Based on Fuzzy Data Mining and Genetic Algorithm[J].Computer Measurement & Control,2012; 20(3): 660-663.
[8] Wan Z, Kothare M V. Efficient scheduled stabilizing output feedback model predictive control for constrained nonlinear systems[J]. IEEE Trans on Automatic Control, 2004, 49(7): 1172-1177.
[9] XIAO Wen, HU Juan. Performance analysis of frequent itemset mining algorithms based on sparseness of dataset[J]. Journal of Computer Applications, 2018, 38(4): 995-1000.
[10] XU S H, SONG M L, XU C, et al. Training algorithm of process neural networks based on hybrid error gradient descent[J]. Journal of Northeast Petroleum University, 2014, 38(4):92-96.
[11] WEI J X, SUN Y H, SU X N. A novel particle swarm optimization based on immune selection[J]. Journal of Nanjing University (Natural Sciences), 2010,46(1):1-9.
[12] ZHOU Wei, LUO Jianjun, JIN Kai, WANG Kai. Particle swarm and differential evolution fusion algorithm based on fuzzy Gauss learning strategy[J]. Journal of Computer Applications, 2017, 37(9): 2536-2540.
[13] WU Zheng, YU Hongtao, LIU Shuxin, ZHU Yuhang. User identification across multiple social networks based on information entropy[J]. Journal of Computer Applications, 2017, 37(8): 2374-2380.
[14] YE N, ZHAO L, DONG L, et al. User identification based on multiple attribute decision making in social networks[J]. China Communications, 2013, 10(12):37-49.
[15] LI Yuezhi, ZHU Yuanyuan, ZHONG Ming. k-core filtered influence maximization algorithms in social networks[J]. Journal of Computer Applications, 2018, 38(2): 464-470.