Sensors & Transducers, Vol. 158, Issue 11, November 2013, pp. 65-69

www.sensorsportal.com

Sensors & Transducers

http://www.sensorsportal.com

A New Data Cleaning and Scheduling Algorithm for Wireless Sensor Networks Based on Internet of Things

[1] Zhang YONG-HENG, [1,2] Zhang FENG, [3] Xu DONG-SHENG, [1] Dong JIAN-GANG

[1] School of Information Engineering, Yulin University, 719000, Yulin, China

[2] School of automation, Northwestern Polytechnical University, Xi'an 710072, China

[3] School of Software, Xi'an University, Xi'an 710065, China

[1] Tel.: +8613509121314

[1] E-mail: 709863637@qq.com, tfnew21@sina.com

Received: 21 September 2013 /Accepted: 25 October 2013 /Published: 30 November 2013

Abstract: Wireless sensor networks data cleaning and scheduling is one of the important research issues in Internet of things theory. The existing data cleaning and scheduling algorithms assume all the data can be loaded into the main memory at one-time, which are infeasible for data center. To this end, based on the Internet of things data center, a data cleaning and scheduling algorithm is proposed in wireless sensor networks. It proposes a scheduling model based on replica firstly, and then by introducing wireless sensor networks mechanisms, it presents a data-driven adaptive replica strategy by considering geographic features, network status, and application characteristics. This dissertation focuses on Internet of things data center, policy of replica, and scheduling mechanism. The experimental results show that the wireless sensor networks environment data replica and scheduling algorithm is effective and feasible, and has better expansibility. *Copyright © 2013 IFSA.*

Keywords: Data center, Scheduling algorithms, Wireless sensor networks, Data cleaning, Internet of things.

1. Introduction

Internet of Things (IoT) is defined as: The radio frequency identification (RFID), infrared sensors [1], global positioning systems, laser scanners and other information sensing device, according to the agreement agreed to anything connected to the Internet, the information exchange and communication, in order to achieve intelligent identify, locate, track, monitor and manage a network.

IoT information including complex management object information, complex sensing device information and real-time information of three kinds of things [2]. Belong to the basic information of two types, the traditional exchange manner can be solved; and complex real-time information not only has the traditional characteristics of real-time information data – a large amount of data, real-time high, but also embodies the associated complex, fast growth – data exchange and query, high frequency characteristic of network information, in the technology on the exchange of information put forward higher requirements.

For complex real-time information exchange in the management process of the problems encountered, one can replace the server processing ability is stronger [3], larger storage capacity, on the other hand, can consider to use improved technology and appropriate strategies. The following, according to the different way of complex information exchange, respectively from the database real-time processing performance (temporary table), large data files to read and analyze performance (memory mapped file method), log management (file buffer strategy) were analyzed in three aspects, then simple statement on the exchange of query optimization strategy.

With the development of data center, especially the emergence of wireless sensor networks technology, data center has not only a simple server hosting [4], maintenance of the site, focus on high performance computers it has evolved into a large amount of data storage and computing [5]. Each IT manufacturer will be prior to the single unit for the server through a variety of ways to become more than one group mode, function based on the development of such as virtualization, IoT [6], cloud storage and a series of efficiency in the use, in order to improve the number of units within the server.

Next-generation data center is not a single discipline; it should be an integrated, standardized, optimized, virtualized, automated and adaptive infrastructure and high-availability computing environments [7].

Wireless sensor networks is to achieve the purpose of collaboration and resource sharing, and IoT diversity environment in every kind of machine resources reflect the heterogeneous and dynamic and user demand, make wireless sensor networks resource management environment is extremely complex. Scheduling problem has been a hot research resource management problem. Because of the own characteristics of IoT, data scheduling problems in the emergence of new challenges: massive heterogeneous resource utilization. IoT services depends largely on the success of the level of utilization of data center resources, and massive heterogeneous data center resource utilization and the level of user task scheduling mechanism have great relevance. Adopt efficient scheduling mechanism which can effectively improve the resource utilization is a IoT service providers need to be resolved.

This paper analyzes and summarizes the new characteristics of IoT data center, from the expansion of wireless sensor networks structure of data center and green energy-saving strategy, copy, starting and scheduling mechanisms, focusing on these three issues.

2. Data Scheduling Algorithm

This section using IoT market model, according to some user task budget less or users of task completion time of all tasks require a higher, resource provider cannot complete the user's situation, put forward a task scheduling algorithm for quality of service driven. This algorithm combines the Min-min algorithm for high throughput and linear planning the advantages of global optimization, not only takes into account all the task of the user, but also considering the higher priority task.

2.1. IoT Scheduling Model

IoT scheduling model primarily by the Client, Broker, resources, wireless sensor networks resource providers and the information server, whose architecture shown in Fig. 1.

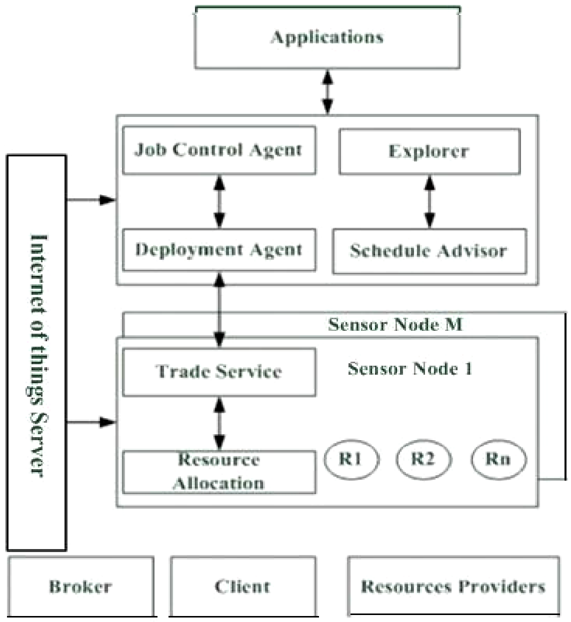

Fig. 1. IoT scheduling model.

Users need to perform tasks usually can be divided into serial applications, parallel applications [8], parameter scanning applications, collaborative applications. The system allows the user to set the resource requirements and parameters of preferences; users through paid use of resources. The agent is the intermediate interface between users and resources, resource discovery, resource selection, for receiving task, returns the scheduling results, exchange of user and resource information. Proxy support different scheduling policies, according to the needs of users to discover resources and scheduling tasks. Agents mainly by job control agent, schedule advisor, explorer, and trade manager and deployment agent components.

In the wireless sensor networks environment has a variety of computing resources, resources of computing performance and calculate the price attribute. Calculated using MIPS (Million Instructions Per Second) performance, property prices per unit time G$. In general, the better the performance of computing resources is calculated, then the higher the price.

2.2. Scheduling Problem Description

In the wireless sensor networks environment, the service provider can get by leasing out the benefits of computing resources [9], and users can buy resources to complete some of their own resources on quality of service requirements cannot afford to complete the task. We assume that the task budget and deadline time is limited; users want the total cost of services within budget, and task completion time prior to the deadline under the premise computing resources to complete the task as much as possible. Suppose a user-defined enough budget and deadline long enough, then the user tasks and deadlines will be in the budget completed within the time limit.

Because the user's budget less or more urgent task in itself, may be caused by cloud computing resources in the budget and deadline constraints cannot be completed tasks [10]. So, we mainly discuss how to maximize the number of tasks completed in this case. At the same time in all the tasks, we also considered in special cases, some more important and urgent task [11-12], users hope that resources can ensure that the shorter deadline to complete these tasks, while the remaining tasks are important, the user can accept this kind of task completion delay.

Now, users from all tasks in the separation of a class of tasks, such higher-priority task, be sure to be completed in a relatively short deadline, we use to represent the class HS task; remaining lower priority user tasks can be delayed a period of time until the last completed before the deadline, with the LS represents such tasks.

2.3. Task Scheduling Algorithm

Suppose in a wireless sensor networks environment, there are m machines $\{r_1, r_2, ..., r_n\}$ and some variables shown in Table 1. According to the user's requirements, HS task should be completed as possible before the HS; LS tasks are completed before LST as possible. To ensure high-priority tasks can demand complete HS, in the scheduling of such tasks, it should be a good allocation of resources to the task of calculating the length of the small, so you can get more in a period of time to complete the task number.

Table 1. Parameter Definition.

Variable	Definition
L_i	The length of the i-th task
k	Number of tasks
m	Number of computing resources
B	Budget data
R_{ej}	J-th performance data is calculated

So in the task scheduling algorithm, according to HS and LS two kinds of different priority tasks, to allocate resources to priority Min-min algorithm by HS. After performing HS task, because each different machine performance, while the size and number of assigned tasks are not the same, so the completion time of each machine and calculate the cost will be different. If using t_j and y_j to represent the j-th machine HS task completion time and cost, with b to represent all HS tasks executed by the computational cost, then $b = \sum_{j=1}^{k} b_j$.

However, if a limited budget or HS task deadline urgency, unable to complete the tasks of type HS all, it should be the type of task is added to the LS task set. In the implementation of LS task, we have driven through a linear programming to get an DM calculation module LS mapping between tasks and resources. We define the structure of each task has an attribute sort, when sort = 0, indicates that the task HS task; when sort = 1, indicates that the task LS tasks.

In the scheduling process, we assume that: in the course of the task cannot be preempted, once the machine is started, you can not interrupt the task instead perform other tasks, you must complete it before proceeding with the remaining tasks. For the same type of tasks without precedence, the user is concerned whether the deadline of resources to complete all tasks.

We can write:

$$b_{ij} = \begin{cases} 1 & i \in \\ 0 & i \notin \end{cases} \tag{1}$$

Because each task performed on a maximum of a resource, so we can write

$$\sum_{j=1}^{k} b_{ij} \le 1, i = 1, ..., k, b_i \in S_{LS} \tag{2}$$

When will execute LS task, due to the implementation of HS types of tasks, each resource has already been carried out for a period of time t_j, while all resources has also obtained the service cost b, according to the cut-off time T_{ls} and the rest of the available budget B-b, so we can write

$$\sum_{j=1}^{k} b_{ij} * L_j * \mathrm{R}_{ej} \le T_{ls} - t_j, j = 1, ..., m, b_j \in S_{ls} \tag{3}$$

In the LS task, the scheduling objective is to maximize the number of tasks, so have the formula (3), so we can write

$$\max = \sum_{i=1}^{k} \sum_{j=1}^{m} b_{ij}, e_i \in S_{LS} \tag{4}$$

As summary, we can get the maximum task for budget and deadline constraints over a number of scheduling algorithm, as shown in algorithm 1.

Algorithm 1. Maximum number of task completion scheduling algorithm:

1) S←the set of all tasks;

2) S_{HS} ←sort = 0 as the set of tasks;

S_{LS} ←sort = 1 as set of tasks;

3) Initialize S_{HS}, S_{LS}, t_j, b_j;

4) If ($S_{HS} = \varnothing$) then go to step 6; else

5) Using Min-min algorithm, HS task completion is expected to return to b, t_j;

6) $S_{LS} = S_{LS}$ □ {e_i | e_i not implemented□ $e_i \in S_{HS}$ };

7) for (i=1; i<=k; i++)
for (j=1; j<=n;j ++)
{

8) if (b_{ij} ==1 && $e_i \in S_{LS}$), Mapping of the i-th to the j-task on machine;

9) else continue;
}

10) According to the previous generation of computing tasks and resources mapping, the task scheduler to the appropriate resources to perform calculations;

11) To accept cloud resource scheduling service side results;

12) end;

3. Simulation Testing and Analysis

To validate our proposed algorithm, this section with two classical algorithms Min-min and BGD algorithm for comparison: 1) When the total task without HS task set when the budget for different tasks and deadlines combined application conditions, the total number of tasks to complete algorithm conditions; (2) When the total number of tasks HS tasks vary from 10 % to 80 %, the algorithm in the HS task completion rate and task completion rate of two aspects. Where HS task completion rate refers to the number of tasks completed HS accounted percentage of total number of tasks, task completion rate refers to the total number of tasks to complete (including the completion of the HS task) representing the percentage of the total number of user tasks. Simulation platform is Cloudsim, Cloudsim is a Java-based event-driven simulation toolkit for wireless sensor networks, and its main goal is to simulate wireless sensor networks environment to study the economic model based on the calculated effective resource allocation method. With 20 compute nodes, each node has 60 computing resources, a total of 600 computing resources. Each node in the calculation of rates and computing resources the same price per unit of time. All nodes in the computing resource performance are shown in Table 2.

Table 2. The time and space overhead for different task algorithm.

Task number	60	200	600	2000
Time overhead(ms)	12	10	23	56
Space overhead(K)	122	202	1024	2134

Fig. 2 respectively is the network performance test results of three kinds of network mode. We designed a simulation experiment; simulation experiments are set as follows: 2000 tasks, the size of the randomly distributed in 600MI (Million Instruction) to 15300MI between. User Tasks budget and deadline combinations of parameters are as follows: the first group (550000, 6000), the second group (900000, 8000), the third group (900000, 8000), the fourth group (1200000, 11000), fifth group (1350000, 14000), the sixth group (1300000, 17000), seventh group (1550000, 22000). Type of task the user does not define HS, HS task set is empty, so there are 2000 LS task set task.

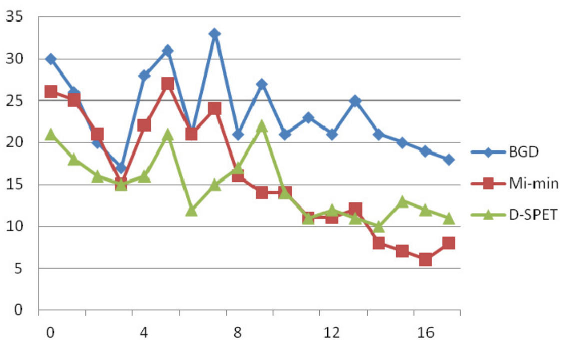

Fig. 2. Each group of budget and deadline corresponding task number.

6. Conclusions

In this paper, first introduces the concept of wireless sensor networks and data center, then aiming at the deficiency of traditional data center, emphatically introduces the features of IoT data center; a detailed analysis of the in wireless sensor networks resources in large scale, three important problems of heterogeneity and produced: scalability and green energy saving, data center network structure copy strategy, and scheduling mechanism research, research content of this thesis, the organizational structure and the contribution of this paper.

For the scheduling problem of this study is the meta task scheduling, did not consider the complex task of further decomposition or directed acyclic graph tasks. The scheduling of these tasks to complete the need to consider the constraints sequentially, and the network communication overhead and the process waiting problem, meta task scheduling is more complicated than. Further research is needed to solve these complex task scheduling.

Acknowledgements

This work is partially supported by the Funding Project for Department of Yulin University in 2013 #13YK48, Funding Project for Department of Education of Shaanxi Province in 2013 #2013JK1167, Research and Cooperation Project for Department of Yulin city in 2011 #2011SKJ07, 2013 #2012cxy3-26, #2013ky0271, #2013ky0148. Thanks for the help.

References

[1]. Yenumula B. Reddy, Secure packet transfer in wireless sensor networks, *Sensors & Transducers*, Vol. 14, Special Issue, March 2012, pp. 308-320.

[2]. Ilkka Kivelä, Chao Gao, Jari Luomala, Jukka Ihalainen, Ismo Hakala, Design of networked low-cost wireless noise measurement sensors, *Sensors & Transducers*, Vol. 10, Special Issue, February 2011, pp. 171-190.

[3]. Xiaoli Wang, Energy-efficient task scheduling model based on MapReduce for cloud computing using genetic algorithm, *Journal of Computers (Finland)*, Vol. 7, No. 12, 2012, pp. 2962-2970.

[4]. Lijuan Wu, Realization of quadrilateral mesh partition and optimization algorithm based on cloud data, *Journal of Computers*, Vol. 6, No. 12, 2012, pp. 2519-2525.

[5]. Tong Yang, Mass data analysis and forecasting based on cloud computing, *Journal of Software*, Vol. 7, No. 10, 2012, pp. 2189-2195.

[6]. Gang Zhang, Data dependant learners ensemble pruning, *Journal of Software*, Vol. 7, No. 4, 2012, pp. 919-926.

[7]. An Junxiu, The demonstration of cloud retrieval system mode, *Journal of Software*, Vol. 6, No. 2, 2011, pp. 249-256.

[8]. Changqing Ji, Big data processing in cloud computing environments, in *Proceedings of the International Symposium on Pervasive Systems, Algorithms, and Networks, I-SPAN 2012*, 2012, pp. 17-23.

[9]. Hongbing Cheng, Identity based encryption and biometric authentication scheme for secure data access in cloud computing, *Chinese Journal of Electronics*, Vol. 21, No. 2, April 2012, pp. 254-259.

[10]. Divyakant Agrawal, Big data and cloud computing: New wine or just new bottles, in *Proceedings of the VLDB Endowment*, Vol. 3, No. 2, September 2010, pp. 1647-1648.

[11]. Edmund Kohlwey, Leveraging the cloud for big data biometrics: Meeting the performance requirements of the next generation biometric systems, in *Proceedings of the IEEE World Congress on Services, SERVICES 2011*, 2011, pp. 597-601.

[12]. Hong-Wei Gao, Yan-Jun Yu, Feng Chai, Shu-Kang Cheng, Position sensorless control of interior permanent magnet synchronous motor based on carrier frequency component method, in *Proceedings of the Chinese Society of Electrical Engineering*, Vol. 30, No. 18, 2010, pp. 91-96.