Data-Intensive Scientific Discovery and Applications

Proceedings of International Workshop on Data-Intensive Scientific Discovery and Applications (2013)

数据密集型的科学发现与应用

——数据密集型的科学发现与应用国际研讨会论文集（2013）

Editors-in-chief :

Jiang Xie

Ronald Cools

Craig Douglas

Wu Zhang

Shanghai University Press

上海大学出版社

ntensive Scientific
ry and Applications

Proceedings of International Workshop on Data-Intensive Scientific Discovery and Applications (2013)

数据密集型的科学发现与应用

——数据密集型的科学发现与应用国际研讨会论文集(2013)

Editors-in-chief Jiang Xie
Ronald Cools
Craig Douglas
Wu Zhang

Shanghai University Press
上海大学出版社

图书在版编目（CIP）数据

数据密集型的科学发现与应用：数据密集型的科学发现与应用国际研讨会论文集(2013)：英文/谢江，罗纳德·库尔斯，克雷格·道格拉斯，张武，主编. —上海：上海大学出版社，2014.8
ISBN 978-7-5671-1371-8

Ⅰ.①数… Ⅱ.①谢… ②罗… ③克… ④张… Ⅲ.①数据采集-国际学术会议-文集-英文 Ⅳ.①TP274-53

中国版本图书馆 CIP 数据核字（2014）第 133616 号

责任编辑 王悦生
封面设计 柯国富
技术编辑 章 斐

数据密集型的科学发现与应用
——数据密集型的科学发现与应用国际研讨会论文集（2013）
谢 江 罗纳德·库尔斯
克雷格·道格拉斯 张 武 主编
*
上海大学出版社出版发行
（上海市上大路 99 号 邮政编码 200444）
（http:/www.shangdapress.com 发行热线 021-66135112）
出版人：郭纯生
*
北京京华虎彩印刷有限公司印刷 各地新华书店经销
开本 787×1092 1/16 印张 12.25 字数 305 千字
2014 年 8 月第 1 版 2014 年 8 月第 1 次印刷

ISBN 978-7-5671-1371-8/TP·061 定价：90.00 元

Preface

Jiang Xie
Chair of Organization Committee,
Shanghai University, China, jiangx@shu.edu.cn

Sponsored by Shanghai University, China, University of Calgary, Canada and International Federation for Information Processing (IFIP) Working Group 2.5, the International Workshop on Data-Intensive Scientific Discovery(DISD) and Annual Meeting of IFIP WG 2.5 (Working Group of Numerical Software) were held successfully in Shanghai during August 1-5, 2013, to address the research issues in designing, building, managing and evaluating advanced data-intensive systems and applications. More than 50 scientists and researchers came from the world exchanged cutting-edge ideas in algorithms, tools, software, and experiences on data-intensive applications as well as the development and availability of sound numerical software. More than 10 sessions were held during the conference.

Underwent several rounds of reviewing DISD2013, 23 papers that received the best review reports finally were accepted for publication on this proceedings. Thematically, the papers showcase the diversity of research that is currently conducted within the domain of data science and technology, that is, data and image processing from scientific and engineering computing, medicine and health, social network and management.

In their paper "A Novel Multi-Block Lattice Boltzmann Method", Wu Zhang et al. demonstrate how LBM simulates three-dimensional flows of high Reynolds number with a new model. Kawata et al. analyze the uncertainty in scientific computing and present new proposals toward an uncertainty inference or management in their paper "Uncertainty Inference for Scientific Computation". In the contribution "The Trade off between Bottom up and Top down Methods in Systems for Data Intensive Scientific Discovery", Randy Goebel takes a review on the fundamental challenges in the development of methods and systems to manage and interpret impossibly large and complex data streams, and the balance between bottom up pattern detection and top-down model-based inquiry. Kai Lu and Huiran Zhang et al. take a different look at the interesting issue of tooth-interstice reconstruction in their paper entitled "DATIR Method: A New Method of Automatic Tooth-Interstice Reconstruction from Dental CT Images". A new method is proposed for automatic tooth-interstice reconstruction. Filip Jeremic and Sanzheng Qiao describe a parallel Jacobi method for lattice basis reduction on GPU and then discuss scalability of the approach and performance issues in their paper entitled "A GPU Implementation of a Jacobi Method for Lattice Basis

Reduction". Craig C. Douglas and Li Deng et al. next discuss combing dynamic big data driven application systems and big data in the paper "A Note on Dynamic Big Data Driven Application Systems (DBDDAS) with Examples". Finally, the paper entitled "Parallel Triangular Solvers on GPU" from Zhangxin Chen et al. develops a new matrix format to address GPU-based parallel triangular solvers systematically.

In the end, I want to thank National Natural Science Foundation of China (NSFC), Science and Technology Commission of Shanghai Municipality (STCSM), IBM in China, Intel in China, and Lenovo China for their support. Thank Shanghai University Press, the responsible editor Yuesheng Wang for giving us the opportunity to realize the proceedings. Thank all the editors for their excellent service, and the anonymous reviewers for their patience and hard work reviewing the manuscripts and suggesting improvements. Thank Dr. Huiran Zhang for the great amount of meticulous work he has done publishing the proceedings.

Shanghai
May 6, 2014

Contents

A Data-Intensive System for Plant Cultivation in Closed Type Plant Factory

Hitohide Usami[1], Yasuhiko Manabe[2], Akira Nakai[3], Shigeo Kawata[4]

[1] Tamagawa University, 6-1-1 Tamagawa Gakuen,
Machida, Tokyo 194-8610, Japan, usami@lab.tamagawa.ac.jp
[2] Numazu National College of Technology, 3600 Ooka,
Numazu, Sizuoka 410-8501, Japan, manaba@numazu-ct.ac.jp
[3] Tateyama Kagaku Industry Co., Ltd, 1583 Oizumi,
Toyama, Toyama 930-8501, Japan, akira.nakai@gmail.com
[4] Utsunomiya University, 7-1-2 Yoto,
Utsunomiya, Tochigi 321-8585, Japan, kwt@cc.utsunomiya-u.ac.jp

Abstract. Recently, plant factory for crop production comes to a front of new Japanese industries. Plant factory has several advantages compared with typical outdoor cultivation. The most important advantage is optimizing control of the plant growth environment in a case of closed type plant factory. It conduces to reassurance, safety, high valued crop production, and year-round or demand season production. We describe a concept of the closed type plant factory and a data-intensive system for the plant cultivation in the plant factory. Our data-intensive system works as an intelligent problem solving environment, based on information and communication technology (ICT) and plant science technology. The plant science perceives exactly the plant growth status from the plant growth data. The plant environment data are easy to obtain by a sensor system, but the plant growth data are difficult to obtain using an image sensor system, because it contains complexity and uncertainty essentially. The knowledge for the plant cultivation, including mainly optimized plant growth environment control values represented as cultivation scenario data, should be progressed through cultivation cycles.

Keywords: data intensive system, sensor system, plant factory, plant cultivation, plant science, IPSE (Intelligent Problem Solving Environment)

1 Introduction

Japanese plant factory production level is still significantly lower than that of the Netherlands in view of information and communication technology (ICT) utilization. It is the most considerable matter to Japanese plant factory that optimal control of cultivate envelopment for the plant growth. We developed a data-intensive system with the view to optimization of the plant growth. This system consists of two parts: the main part is a data collecting subsystem using several types of sensors, and the other part is a scenario based control subsystem based on cultivation knowhow. The collecting data are the plant growth data (flesh weight of plant, plant height, number of leaves, etc) and the cultivation environment data (temperature, humidity, light intensity, CO_2 density, air flow, water solution temperature, etc.). Otherwise, the scenario based control subsystem is an advanced feature for realizing the optimized plant cultivation environments. The plant growth strongly depends on light quality and intensity, and it is varied each plant growth stages, because plant has red and blue photoreceptors (the red light affects the plant growth and the blue light affects the morphology in typical plants). The scenario based plant control system can change the light quality and the intensity according to an optimized cultivation scenario. We descried the data-intensive system for realizing of optimized plant cultivation in the closed type plant factory. Figure 1 shows the closed type plant factory using a part of this technology, and it named by "SciTech Farm TN Produce" (commercial level full controlled LED based facility). The restaurant "KEYAKI" (bricked building in behind of picture) serves flesh harvested vegetable salad as the factory outlet. This facility is two-story building equipped with a solar power generation system. Herbs and leaf lettuce can be grown 3,900 (currently 600) shares per day in near future.

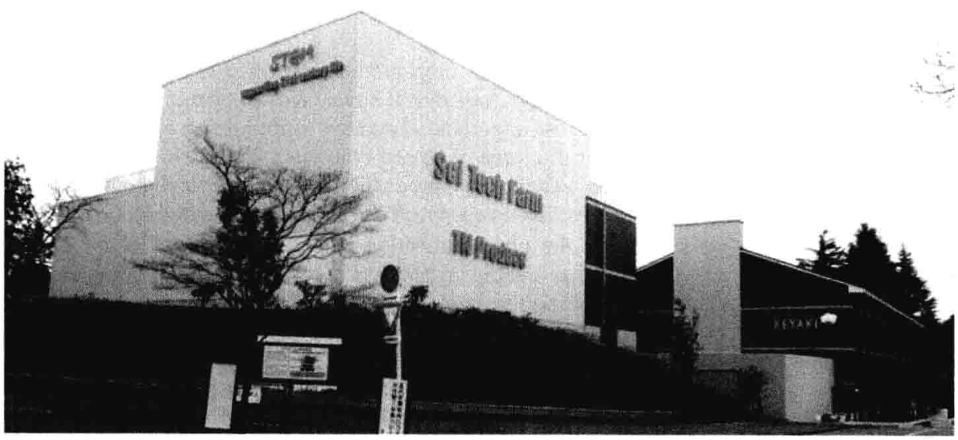

Fig. 1. Closed type commercial level plant factory in Tamagawa University.

2 Closed type plant factory

The closed type plant factory is a front-line plant production facility with the artificially controlled cultivation environments, and those major control factors shown as follows; light (color quality, intensity, duration), temperature, humidity, CO_2 density, liquid manure (temperature, EC, pH), and air flow.

2.1 Environment control

Our plant factory focuses on the artificial lighting environments for the plant growth. The lighting system can be expressed in any color of light by using LED panel consist of RGB (Red, Green, Blue) LED chips. These LED panels can control the color quality of light, the light intensity, and the lighting duration by the computer. The red and the blue light give the great affect on the plant growth, because plants have red and blue light photoreceptors. However the green light is a small affect on the plant growth, because plants don't have green light photoreceptor. In addition, the liquid manure management is also important factor on the realization of the optimum plant cultivation. We use automatic manure control machine called "rakuraku manure management machine", and it control temperature, EC (Electric Conductivity), and pH (potential Hydrogen) of the hydroponic solution. Room temperature and humidity are controlled by standard large scale air-conditioners. Inset, dust, and bacteria elimination are basic features for the closed type plant factory; we use high quality HEPA (High Efficiency Particulate Air) filter for keeping the higher clean level (class 10,000 clean room) in the cultivation room.

2.2 Advantages of the full controlled plant factory

There are many advantages to the closed type full controlled plant factory and those advantages shown as follows.

- Year-round and efficient crops production (improving productivity)
- Crops production on demand market (job shop type production)
- Reassurance, safety and delicious (none residual pesticides)
- High valued plants production (drug plants, genetically modified plants, etc)
- Crops production in urban and barren areas (city, desert, outer space production)

The plant factory can produce about two to four times faster growth and drastically increase metabolic substance in plants than typical outdoor cultivation in case of lettuce by the optimized cultivation environments. In addition, we use multi-shelf system (multistage cultivation shelves), and its easy expand to vertical direction called vertical farm. We can achieve the mass production of crops in a small space as the urban agriculture.

3 Data-intensive system

3.1 Evidence based plant cultivation

We are now cultivating the high valued plants (transgenic plants for drug production) for the purpose of examination of the optimized cultivate environments. The plant growth and the amount of metabolic substance are influenced by those cultivating environments (especially the light quality) and some kind of stress (e.g., UV-light, hydrogen peroxide, atmospheric pressure). The cultivation data transform to the evidence based plant cultivation knowledge which include the optimized environment parameter values derived from relations of the plant growth observations and the cultivation environments. The plant cultivation scenario is rewired from these evidence based plant cultivation knowledge to the time series of sequence control.

3.2 System overview

The evidence based plant cultivation use the plant cultivation knowledge as a formed the plant cultivation scenario. Figure 2 shows the whole of our data-intensive system conceptual diagram for the plant cultivation in the closed type plant factory.

This system achieves the scenario based plant cultivation to accommodate on-demand plant production in the plant factory. This scenario can describe the cultivation environment parameter values on each plant growth stages. The cultivation scenario is a kernel part in this system, and initial scenario makes following procedures. First, collect the plant growth data (flesh plant weight, plant height, number of leaves, etc) with the environmental data such as temperature and humidity. Next step, extract useful knowledge from the collected data used by advanced technologies, for example data mining, and then extracted knowledge are stored in the database. These multiple fragmented knowledge are synthesized as one cultivation scenario and it store in the scenario database for the next cultivation. These scenarios are modified or rewrote to the progress of the cultivation efficiency and quality through every cultivation cycles. We are developing the part of the red dashed box in Fig. 2 for the experimental cultivation and this paper describes this part. This system consist of two parts, one is the data acquisition include sensor system, and the other part is the environment control system based on the scenario operation. The data-intensive system for the plant cultivation works as follows. (1) The several type sensors monitor the cultivation environment and sending those data to the data acquisition equipment and log to the data server system. (2) If abnormal data come from the sensor system, the management system sending the alarm message to the operator by e-mail. (3) The operator easy checks the plant growth status used by Web camera before go to the next cultivation stage. (4) The operator modifies the scenario data for the optimized plant growth.

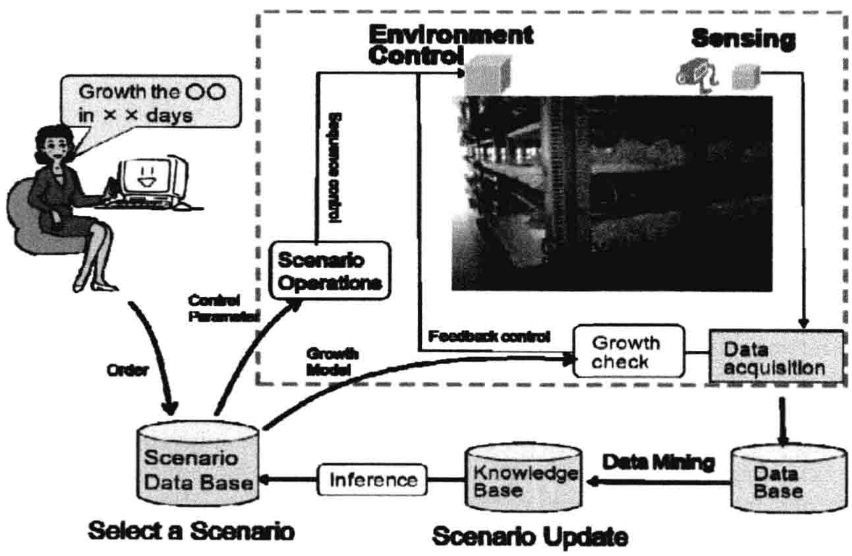

Fig. 2. Closed type commercial level plant factory in Tamagawa University.

3.3 Cultivation experiments

We have performed 6 times the cultivation experiments from January through May in 2013. Unfortunately, we couldn't obtain the complete results caused by equipment glitches in those cultivation experiments. This paper shows the No. 3 cultivation scenario as a typical case, because it's relatively stable experiment than others.

1) Cultivation scenario

Table 1 shows control values of the cultivation scenario on the 3rd cultivation experiment (No. 3 cultivation scenario) from February 25th to March 15th, 2013.

Table 1. Control values of cultivation scenario No.3.

Cultivation experiment	Plant Growth Stage			
	Stage 1 (5 days)	Stage 2 (5 days)	Stage 3 (4 days)	Stage 4 (4 days)
LED	R:200→B:200 →Dark	R:200→B:200 →Dark	R:200→B:200 →Dark	R:200→B:200 →Dark
CO2 concentration	1000PPM	1000PPM	1000PPM	1000PPM
Wind speed	0. 5m/s	0. 5m/s	0. 8m/s	0. 8m/s
Nutrient Solution Temperature	21℃	21℃	21℃	21℃
Mist interval	2 minute	2 minute	2 minute	2 minute

The LED right control is key issue in this experiment. First light quality is the red, power 200 μ mol · m^{-2} · s^{-1}, duration 15h, next the blue light, power 200 μ mol · m^{-2} · s^{-1}, duration 1h, and the remaining 8h is the dark period on each plant growth stages.

2) Cultivation results

We measured flesh weight of plants, light intensity, CO_2 density, temperature, humidity, air flow and water solution (EC, pH, etc) by the online system during experiment days. Figure 3 shows an experimental result graph of the three key data (CO_2 density, LED intensity, Flesh weight) in the cultivation scenario 3 of Table 1.

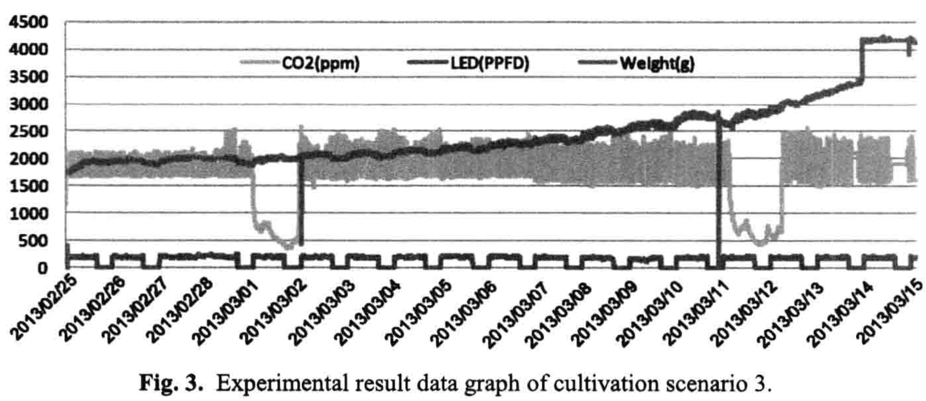

Fig. 3. Experimental result data graph of cultivation scenario 3.

The real-time flesh weight measured (blue line) by using distortion gage sensors and we had verified the exact plant weight by compared with harvested plant flesh weight. Supply of CO_2 gas (green line) had stopped twice during this experiment caused by gas empty. The LED control (red line) worked almost normal lightning.

Finally, we had done multiple regression analyses to find the most effective environment factor for the plant growth. The multiple regulation analysis results do not shows clear effective factor. However, the result of experiments data shows that the flesh weight strongly correlated to the LED intensity, but not to CO_2 density.

4 Discussions and conclusions

We have conducted researches and developments of a data-intensive system for the plant cultivation in the closed type plant factory, and performed experiments several times in scenario based plant cultivation. We obtained enormous amount of data using multi-factorial sensor based on a monitoring system and tried to apply an environment control using knowledge extracted from the analyzed data. The multiple regulation analysis results do not shows clear effective factor. However, the result of experiments data shows that the flesh weight strongly correlated to the LED intensity, but not to CO_2 density.

Acknowledgments

This work was partly supported by a grant from the Ministry of Agriculture, Forestry and Fisheries. We would like to express special thanks to Prof. Hiroyuki Watanabe, Prof. Keiko Ohashi, Dr. Eiichi Ono, Mr. Masayoshi Fuse (Tamagawa University), Mr. Kazuoki Saito, Mr. Takashi Uchibori (SENECOM), and Mr. Akifumi Horiguchi (Enzan Koubou) for their supports and valuable discussions on this work.

References

1. A. Ikeda, Y. Tanimura, K. Ezaki, Y. Kawai, S. Nakayama, K. Iwao, H. Kageyama, "Environmental control and operation monitoring in a plant factory using artificial light, Acta Hort. (ISHS) 304, pp. 151-158, March 1992.
2. T. Morimoto,T. Torii,Y. Hashimoto, "Optimal control of physiological processes of plants in a green plant factory", Vol. 3, No. 4, pp.505-511, April 1995.
3. Noboru Noguchi, "Intelligent bioproduction system towards enviroment protection", Environ. Control Biol.,Vol. 47, No.2, pp.111-120, June 2009.
4. Yasuhiko Manabe, Hitohide Usami and Shigeo Kawata, "A Problem Solving Environment for a Plant Factory", Proc. the Conference on Computational Engineering and Science, Vol.16, F-6-3, May 2011.
5. Yasuhiko Manabe, Hitohide Usami and Shigeo Kawata, "The Growth Simulation for a Plant Factory", In proceedings of the Conference on Computational Engineering and Science, Vol.17, H-4-1, May 2012.
6. J. Vos, L. F. M. Marcelis, P. H. B. De Visser, P. C. Struik, J. B. Evers, Functional-Structural Plant Modelling in Crop Production, Springer, 2007.
7. CyberGreen: http://www.enzan-green.com/cyber_green/ .

A Double-Sided Combinational Auction Strategy in the Federated Clouds

Xing Wu[1,2], Shaojian Zhuo[1], Haitao Lv[1], Hanyu Wang[1]

[1] School of Computer Engineering and Science, Shanghai University, Shanghai, China
[2] State Key Laboratory of Software Engineering, Wuhan University Wuhan, China

xingwuvip@gmail.com

Abstract. In federated clouds, people can share and trade cloud resources even they are in different regions. However, it is difficult to allocate cloud resources efficiently and dynamically to achieve maximum revenue. A double-sided combinational auction strategy is proposed to address this problem. In the proposed method, every user can play the role both as cloud resource user and as resource vendor. Simulation experiments are performed to compare the proposed method with a traditional method. Experimental results reveal that our proposed method has a better performance both in revenue and in success rate.

Keywords: double-sided auction, resource allocation, federated clouds, winner determination

1 Introduction

Cloud computing is a hot topic in recent years. Some business applications have been already occurred, for example Amazon EC2, Google AppEngine, Microsoft's Azure and so on. In cloud computing, there are three parts: infrastructure as a service (IaaS), platform as a service (PaaS) and software as a service (SaaS). Using clouds, users can share and trade cloud resources [1]. Efficient allocation cloud resources and pricing strategy are of great importance in the development of cloud computing. Four types of approaches have been proposed by some researchers: (1) cloud resource vendors adopt fixed pricing strategy [2]; (2) cloud resource vendors adopt dynamic pricing strategy according to supply-demand relationship [3]; (3) cloud resource vendors adopt auction strategy to achieve maximum revenue [4]; (4) negotiation strategy is adopted by vendors and buyers to determine resource allocation and price [5].

In federated clouds, users can get cloud resources from different resource vendors [6]. Since there are a lot of cloud resource vendors, users may have multiple choices, such as they can choose the vendors having lowest price or the vendors serving the best service. However it takes users' time to select the most suitable vendors, which may decrease the system efficiency. There have been some researches concerned with the cloud pricing strategy and resource allocation. Arun Anandasivam in [7] presented his understanding of bid price control and dynamic pricing in clouds and shows that bid-price model and dynamic-price model outperform static-price model in cloud computing. Sharrukh Zaman in [8] studied the combinatorial auction-based allocation of virtual machine instances in clouds and shows that it can significantly improve the allocation efficiency while generating higher revenue for the cloud vendors. Xin Sui in [9] proposed an adaptive bidding for combinational auction which perform very well even without any prior knowledge about the market. Ori Regev [10] implemented several market mechanisms of the single-sided and double-sided auction to match buyers and sellers in CPU time market.

Inspired by auction-based resource allocation protocols in grid [11] and dynamic resource pricing of multiple resource types on federated clouds [12], we proposed a double-sided combinational auction strategy to help cloud users choose the most suitable vendors and boost the efficiently of cloud market. In the proposed method, cloud users can submit combinational request for the cloud resources and the third party will provide the most suitable vendors automatically with a Service Level Agreement (SLA) selection mechanism [13].

The remainder of this paper is organized as follows. Section 2 introduces system model. Section 3 discusses our propose strategy. Section 4 evaluates our strategy. And Section 5 concludes our paper.

2 System model

A federated cloud is that people can share and trade their resources no matter where they are. When a user needs some resources, for instance one computing unit and two storage units, to finish work, he/she may submit a request to the cloud. If a cloud resource vendor has suitable resources, the user and the vendor may negotiate about the price. Furthermore if the price is acceptable by the user, the transaction will be carried out. After the user finished the work and return the resource, the vendor can sell it to another user. The whole process can be shown in Figure 1.

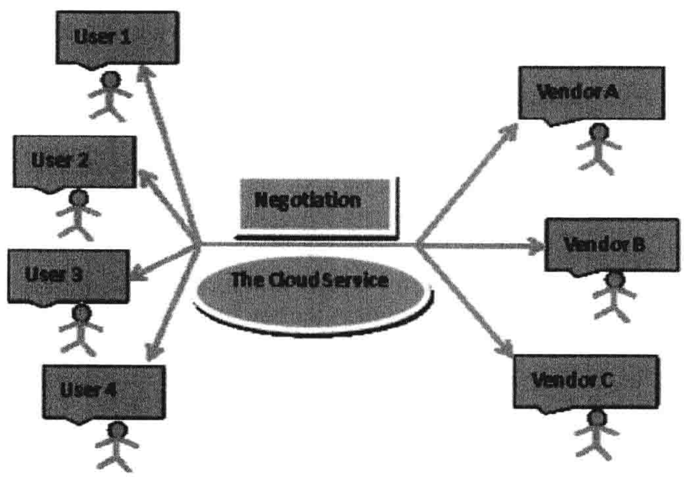

Fig. 1. System model for federated cloud.

However, there are some problems in federated cloud. On the one hand, if a cloud user has a low budget for the required cloud resources, he/she will have to negotiate many times to find an appropriate cloud vendor. This will reduce the system efficiency. On the other hand, different vendors has different service level and price. Therefore it will take users plenty of time to choose suitable vendors. As a result, a new mechanism is needed to provide better user experience.

3 Double-sided combinational auction strategy

3.1 A simple double-sided auction

The simple double auction can be described as two priority queues, one for the cloud users and the other for the cloud vendors. The cloud users queue is ordered according to their bids. A cloud users with a higher bid is positioned higher in the users' queue. The cloud vendors queue is ordered by the vendors ask prices. A vendor that asks for a lower price is positioned higher in the vendors' queue. In every round of the repeated auction, the highest user is matched to the lowest "vendor" (unless the vendor's ask is higher than the user's bid). The winning user pays to the winning vendor an amount that is equal to the average of their offers. The users' and vendors' offers may change across rounds. The mechanism is simple and easy to implement.

3.2 Double-sided auction in federated cloud

A double-sided auction strategy is proposed for the federated clouds. In federated clouds, there are *M* vendors who want to sell their resources, *N* users who want to buy cloud resource to fulfill their computing missions. There are *T* different types of resources provided by vendors such as computing unit, storage unit, bandwidth and so on. Before an auction, users submit their requests and vendors submit their resources to the third party which will run a winner determination program to match the requests and resources. After a pre-defined period *t*, such as 10 minutes, 15 minutes or etc., the third party begins to run winner determination program and selects the winners who get resources. The process is shown in Figure 2.

A user's request can be represented as a vector :

B[user_id,
item_1, price_1,time_1,
item_2, price_2,time_2,
....,
item_n,price_n,time_n,
SLA], n<=T

and in the request, "item_1" means the number of item#1, "price_1" means the max-price of per item#1 and "time_1" means the timeslot needs to be used. (In the winner determination program, a user is a winner only if he/she gets all requested resources).

A vendor's resources can also be showed as a vector:

P[vendor_id,
item_1, price_1,

item_2, price_2,

....,

item_n, price_n].

in which "item_1" means the number of available item#1 and "price_1" means the min-price of per item#1.

The buyers' requests define a demand curve, and the vendors' resources define a supply curve. Then the "Marshallian cross [14]" is used to decide on a price and quantity for trading. The intersection of the two curves defines the trading price, p_0. Users who bid more than p_0 trade with vendors who ask for less than p_0 and the market is cleared. Most of the characteristics of the simple double-sided auction apply here, too. The exceptions are the price determination rule and the fact that more than one user-vendor pair may trade in each round.

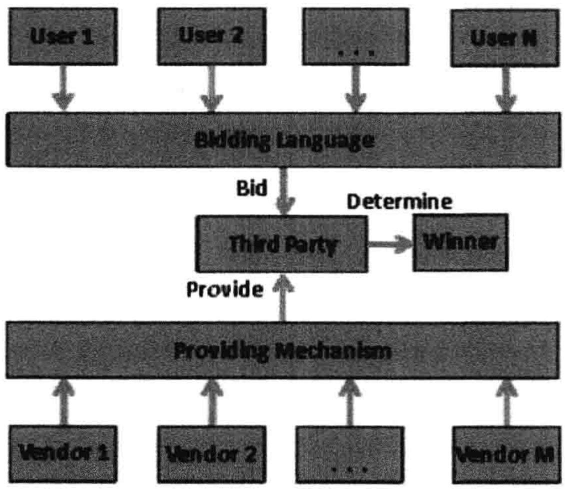

Fig. 2. Double-sided auction strategy federated cloud.

3.3 Winner determination

The main idea is that when allocate a type of item, a user who has a higher price can get resources first, and choose the vendor who has a lower price first. The clearing price p_0 in each round is determined as follows. The users' requests are written in descending order, $V_{(1)} \geq V_{(2)} \geq \ldots \geq V_{(n)}$. The sellers ask-prices are written in ascending order, $A_{(1)} \leq A_{(2)} \leq \ldots \leq A_{(m)}$. We add a fictitious bid, $V_{n+1} = 0$, and a fictitious ask, $A_{m+1} = \infty$. The mechanism then looks for the first pair $(V_{(j+1)}, A_{(j+1)})$ satisfying $A_{(j+1)} > V_{(j+1)}$. The traded quantity is then j, and the clearing price is $p_0 = (V_j + A_j)/2$, which is shown in Figure 3.