Proceedings of

2014 International Conference on Cloud Computing and Internet of things

CCIOT 2014

Dec 13-14, 2014

Changchun, China

Organizer:

Changchun Normal University

Co-organizers:

Northeast Normal University

Liaoning Normal University

Technical co-sponsor:

IEEE

2014 International Conference on Cloud Computing and Internet of things

IEEE Catalog Number: CFP1424Y-CDR
ISBN: 978-1-4799-4765-2

CCIOT 2014 Organizing Committee

Honorary Chair	
ZHAO Ji-min	Changchun Normal University, China
General Chair	
LIU Bao-yuan	Changchun Normal University, China
General Co-Chair	
YU Fan-hua	Changchun Normal University, China
Organizing Chair	
LIU Guang-jie	Changchun Normal University, China
Organizing Co-Chair	
ZHOU Xiao-ying	Changchun Normal University, China
Technical Program Committee Chair	
WANG Wen-yong	Northeast Normal University, China
Technical Program Committee Co-Chair	
YU Fan-hua	Changchun Normal University, China
Members of Technical Program Committees	
ZHOU Chun-guang	Jilin University
ZHAO Hong-wei	Jilin University
YANG Hua-min	Changchun University of Science and Technology, China
WANG Li-min	Jilin University of Finance and Economics, China
WANG Hong-zhi	Changchun University of Technology, China
LI Shi-jun	Jilin Agricultural University, China
MA Ming	Beihua University, China
SONG Shao-zhong	Jilin Business and Technology College, China

Table of Contents

Cloud-EDA:a PaaS Platform Architecture and Application Development for IC Design&Test

Chenlong Man[1], Zaifeng Shi*[1], Zehao Xu[1], Yong Zong[1], Ke Pang[1]
[1.]School of Electronic Information Engineering, Tianjin University, Tianjin, P.R. China
Email: shizaifeng@tju.edu.cn

Yuzhao Li[2]
[2.] Science&Technology Parks Corporation Hong Kong, P.R. China

***Abstract*—IC industry has made great progress in the recent years. However, restricted by the funds and budget, resources of IC design and test are imbalanced. This paper describes an architecture of Cloud-EDA Platform based on Cloud Computing, which is the most popular topic of the internet and IT industry for its high flexibility and low cost. According to the proposed architecture, we design the prototype of Cloud-EDA Platform and develop an application of test-pattern-conversion, which demonstrate that cloud computing can be combined with IC industry to enable ordinary users, app developers and EDA vendors to achieve a win-win-win result.**

Keywords—Cloud Platform;IC Design;Auto Test Equipment; PaaS

I. INTRODUCTION

With the development of Integrated Circuits (IC), the chip size is much bigger and the logic is much more complex. The lackness of essential equipment gradually become the bottleneck in IC design, especially for small and medium-sized companies who have no ability to build strong servers or hardware acceleration simulation equipment. Besides, the existing commercial Electronic Design Automation (EDA) software is various and expensive, the acquisition of EDA software is also a big restriction for small and medium-sized companies. In contrast, the hardware and software purchased by some big companies cannot be fully utilized, parts of equipment is vacant, which causes great waste of resources. In terms of IC test, the rapid growth of design scale and complexity causes an exponential increase in the workload of verification and test for these designs [1], which in turn is driving more investment in the hardware (e.g., automatic test equipment, servers), software (e.g., TetraMAX, EDA tools) and other cost, including hardware obsolescence, electricity cooling, management, networking, physical space, back-ups, etc.

During the process of IC test and verification, it is difficult for test engineers to generate a valid test pattern named automatic test pattern (atp) to perform in the Automatic Test Equipment (ATE), including scan and functional test. Usually, many IC design simulation tools output Value Change Dump (VCD) format or Waveform Generation Language (WGL) format as functional simulation. However, the VCD format is not a good representation of the required test pattern for test equipment, because it is "time-based" in nature and is quite difficult to be converted into the "cycle-based" test pattern for chip testing purpose. The "tailoring" of VCD file is a common nightmare experience to test engineers. Often, engineers require some additional tools, for example TetraMAX, to generate the required test pattern from previous VCD files. TetraMAX is an automatic test pattern generation tool provided by Synopsys, and one of its function is to generate atp file from the simulation tools. However, the rent of TetraMAX is expensive, especially for college teachers and students.

Cloud computing is a service and information acquisition mode based on internet, which enjoys great popularity among the internet and Information Technology (IT) industry in recent years. Through cloud computing, the shared hardware and software resources can be offered to required individuals or companies. Generally, Cloud computing services can be divided into three different kinds of models: Infrastructure-as-a-Service (IaaS), Platform-as-a-Service (PaaS) and Software-as-a-Service (SaaS) [2]. With these available computing services, a cloud platform is capable of providing diverse services for users [3]. In terms of its characteristics, cloud computing is no longer a nomenclature in the field of IT, many governments and enterprises have already started to develop public clouds, private clouds and mixed clouds. Google cloud provides java / Python runtime platform and data storage interfaces [4]. Azure services platform released by Microsoft provides an online development, storage and service escrow environment based on Windows series products. Besides, cloud computing has been widely used in the field of communication [5], medical treatment [6], manufacturing industry [7], etc. In [8], an architecture of Web-EDA system based on cloud computing is proposed. However, this platform is a kind of SaaS platform which provides access to partial EDA servers and special applications to ordinary users without strong flexibility.

This paper describes a PaaS platform of Cloud-EDA for IC design and test, the rest of the paper is organized as follows. Section II describes the architecture and the hierarchical structure of this platform. Section III shows the development process of a test-pattern-conversion app based on the prototype of this platform and displays the workflow of chip test process. The results and analysis of this platform are presented in Section IV. And the Section V concludes this paper.

II. ARCHITECTURE OF CLOUD-EDA PLATFORM

The cloud mode brings a new method of delivering resources more efficient and more economical. The Cloud-EDA Platform is based on this model, on the top level, offering Platform-as-a-Service (PaaS) as the service delivery model to the clients, on the bottom level, connecting with several EDA servers, ATE, etc. and providing hardware to developers or users on demand via the Internet.

Fig.1. shows the architecture of the Cloud-EDA Platform consisting of ordinary users, app developers, ATE providers, EDA service providers. The manager of cloud resources negotiates with the EDA service providers or ATE providers, such as EDA vendors, some big companies and IC design industrial centers, to provide hardware equipment as infrastructure. The developers contain with small and medium companies, individual person, etc. And they have partial rights to gain access to application development environment. Their applications are hosted in the cloud platform, and can be executed in its runtime environment. Ordinary users can get these applications from the App Store with a pay-per-use charging scheme or get the rights to use hardware equipment or software in pay-as-you-go cost model. Besides, it is the manager's responsible to make user management, app runtime management, payment management, storage management, external hardware management, etc.

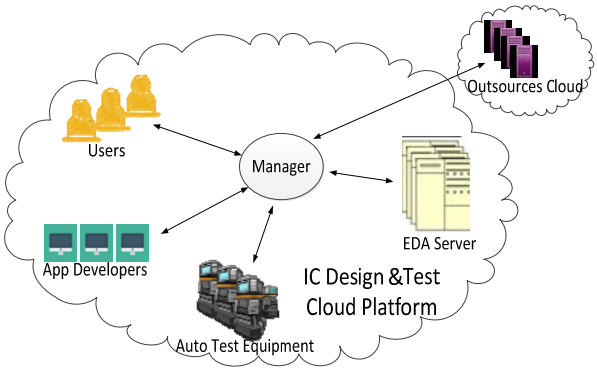


Fig. 1. Architecture of Cloud-EDA Platform

The hierarchy layered design of this Cloud Platform is shown in Fig.2. The Infrastructure-as-a-Service (IaaS) layer consists of various hardware and software, such as EDA servers, ATE, database, instruments and acceleration simulation equipment. Besides, the outsourced cloud resources will be employed if the local resources are insufficient. Above infrastructure level, there is a data center managing, configuring, and monitoring the hardware, database, and software resources virtually. To make the platform more generic and more efficient, The IaaS has to meet following requirements, such as practicability, high capacity, easy to configuration, security and so on.

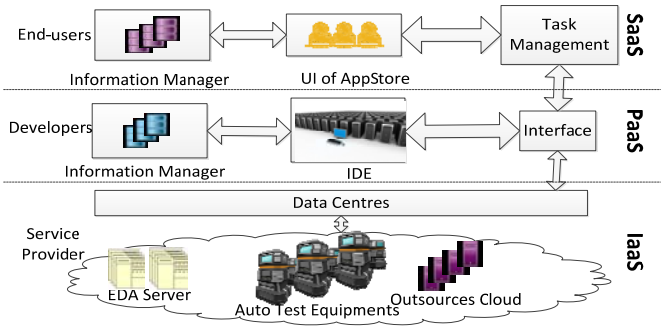


Fig. 2. The hierarchy layered disign of Cloud-EDA Platform

PaaS is situated on a higher level within the Cloud platform hierarchy. To avoid disturbing the developers with matters of allocation, the applications are executed in data center. Meanwhile, developers have to handle some constraints that the environment imposes on their application design at the cost of convenience.

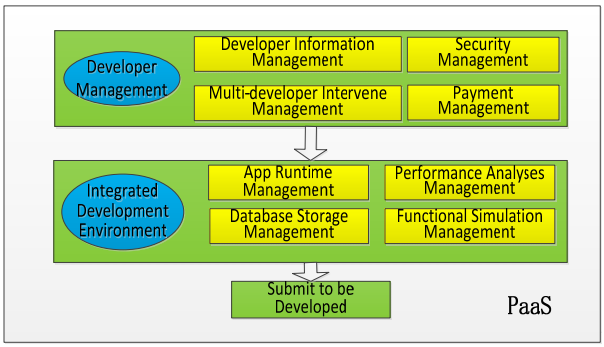


Fig. 3. Architecture of PaaS layer

The PaaS is a web application hosting service, allowing for development and deployment of web-based applications within a pre-defined runtime environment, as shown in Fig.3. The primary users of this layer are the developers of this platform, e.g. application develop companies and individual person fond of application development. It offers a developer management system to get access to Integrated Development Environment (IDE) to develop applications relating to IC design and test. The developer management part provides management of developer information, developer security, multi-developer intervene and payment. The IDE provides apps runtime management, database storage management, performance analyses and functional simulation. The runtime management mainly supplies some Application Programming Interface (API), such as API for apps to do I/O, API for interface to external hardware and software, API for user account enquiry, API for collaboration, etc.

The most widely used among ordinary users is the SaaS layer, users can directly visit the platform via browser to get service, such as the temporary rights to use EDA server, ATE or the apps provided by platform or developers in pay-as-you-go mode.

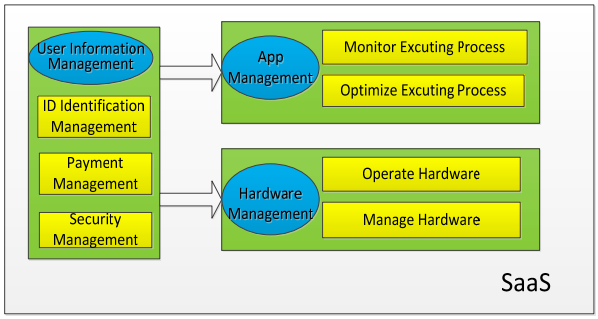


Fig. 4. Architecture of SaaS layer

Fig.4. shows the structure of SaaS layer, which consists of user information management, user app management and user hardware management. The user information management sub-system mainly contains user's ID identification management, payment management, security management, etc. The user app management provides users with authority to use the apps supported by other developers or platform itself, including monitoring and optimizing the process of apps' executing. The hardware management is responsible for operating an EDA service or ATE and sending back the results of hardware to the users and the managers of the platform.

III. Propotype of Cloud-EDA Platform and Application Development

According to the hierarchical structure discussed, we developed the prototype of this platform comprising of an EDA server, a set of ATE, an application server and a network server, as shown in Fig.5. The application server takes the task of executing application in its runtime environment, operating the EDA equipment, and returning feedback. The network communicates with users, and the communication contains presenting available apps for user, securing login for payment gateway and conveying results to end-users. The EDA server or ATE receives and executes commands from the main application server.

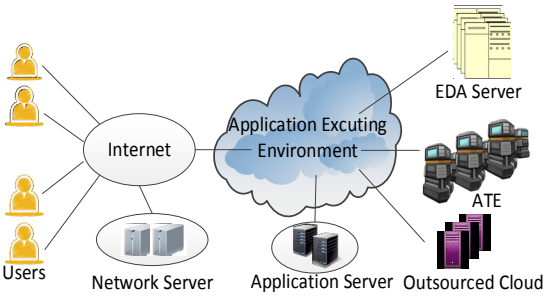


Fig. 5. Prototype architecture of IC Design&Test Cloud Platform

IC test and verification is gradually occupying more and more important position in IC produce flow. In the early stage, the scale of IC design is small and complexity is simple, the test of chips or sub-system is quickly and not the limiting factor. However, as the scale of IC gradually increasing and the complexity of chip's sub-system ascending, traditional test method is not applicative and the Automatic Test Equipment arises at the historic moment. The ATE is convenient and efficient in testing chips, and it needs configuration files named auto test pattern (atp) files to operate. Unfortunately, many IC design simulation tools output other formats (e.g., VCD, WGL) of file rather than atp file, so some additional tools are required to convert the original simulator outputs to required atp format.

Based on proposed PaaS platform, we developed the application of test-pattern-conversion. Firstly it extracts all information in the WGL file and store it in a dynamically created Waveform Database (WDB) with a propriety format. And the database is manageable in a web-server environment, the size of the test pattern, the use of CPU time and the allocation of storage space are prescribed in advance. Overall, the strategy of WDB is the key point of test-pattern-conversion process. The second step of the pattern conversion is extracting the WDB and re-generate the pattern according to the specification of the target ATE. During such conversion, we analyzed the restriction of different patterns and the format of atp, and pay much attention on the edge information of the waveform. The experiment have demonstrated that this app has ideal functionality of converting the WGL file to atp file.

The workflow of chip auto test based on EDA-Cloud platform is shown in Fig.6. After purchasing the test-pattern-conversion app supplied by the platform. Users provide the original IC Design simulation tools output, e.g. VCD, eVCD, WGL pattern files, then select the type of ATE (e.g. Teradyne J750 or Verigy 93000) and launch the test equipment. After that, the application server will execute the conversion program (the gray background of workflow) and verify the converted atp file until the result is accurate. After conversion verification, the atp file will be transmitted to ATE, in which it will experience the chip verification, the designed chips are proved to be eligible if the verification is success, otherwise the chips may have some defects.

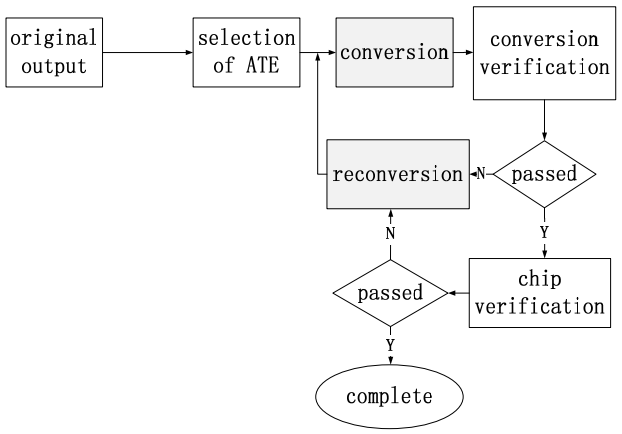


Fig. 6. Workflow of chip auto test system based on Cloud-EDA platform

IV. Implement for Cloud-EDA Platform and Results Analysis

Based on Google cloud computing platform, we implemented the User Interface (UI) of the Cloud-EDA App Store and deployed the test-pattern-conversion app in App Store.

The UI of App Store is shown in Fig.7. Users can register and login his personal account, change personal information, browse service information (including the price and function of

ATE, EDA service, and apps), recharge his account and get access to service. After authentication, the web server gets the user roles and user permission, and then dynamically load the accessible modules to users. In addition, the server monitors the executing conditions of hardware and apps to ensure the security of user's data and the platform.

Fig. 7. The UI of Cloud-EDA online App Store

With the development of IC industry, it gradually formed a contradiction, on the one hand, big companies own masses of software and hardware equipment with a low utilization rate limited by peaks and valleys of business, on the other hand, small and medium-sized companies don't have enough funds to purchase the expensive equipment or software licenses to support their design or test. So we combined the cloud's flexibility and pay-as-you-go model with the rapidly developing IC industry, and proposed the Cloud-EDA Platform. Through this mode, big companies or famous EDA vendors can serve their vacant equipment or products as infrastructure of this platform, avoiding waste and getting some profits. For app developers or small EDA vendors, they can advertise their EDA tools or apps through this platform instead of a marketing team. For ordinary users or small companies, they can get service on demand without massive investment to purchase infrastructure. Therefore, IC design and test Cloud-EDA Platform presents a new win-win-win mode of IC industry.

V. CONCLUSIONS

To solve the imbalanced situation of IC industry, we have proposed a Cloud-EDA Platform of IC design and test based on the characteristic of cloud computing in this paper.

Firstly, we introduced the dilemma of small and medium-sized companies lacking the essential equipment or software in IC design and test. Then we discussed the potential of combining cloud computing with IC design and test.

Secondly, we described the framework of Cloud-EDA platform and the hierarchy layered structure of the platform, including IaaS layer, PssS layer and SaaS layer.

Then we developed the test-pattern-conversion application employed in App Store based on the prototype of this platform. The workflow of a chip test process by this App was proposed.

Finally, an UI of App Store was proposed as a SaaS of this Cloud-EDA platform based on Google cloud platform, to demonstrate that the cloud platform can be used in rapidly developing IC industry with its flexible and pay-as-you-go merit. In the future, we will focus on multi-server parallel computing and the optimization of user UI.

ACKNOWLEDGMENT

This paper is supported by the funding of the International Science and Technology Cooperation Project of Tianjin (research and development on shared cloud platform for IC design and test), and the National High Technology Research&Development Program of China(2012AA012705)

REFERENCES

[1] H. Ranjan, "Cloud Computing And EDA: Is Cloud Technology Ready for Verification…"2011 International Symposium on VLSI Design, Automation and Test (VLSI-DAT). Hsinchu, pp. 1-2, April 2011.

[2] D. Agarwal, S.K. Prasad, "AzureBOT: A Framework for Bag-of-Tasks Applications on the Azure Cloud Platform" 2013 IEEE 27th International Parallel and Distributed Processing Symposium Workshops& PhD Forum (IPDPSW). Cambridge , pp. 2139-2146, May 2013.

[3] Rongheng Lin, Budan Wu, Fangchun Yang, Yao Zhao, "An efficient adaptive failure detection mechanism for cloud platform based on volterra series" Communications. China, vol. 11, pp. 1-12, April 2014

[4] Zeng Shu-Qing and Xu JieBin, "Google-Wide Profiling: A Continuous Profiling Infrastructure for Data Centers" Micro, vol. 30, pp. 65-79, August 2010

[5] P.Calyam, M.Sridharan, Yingxiao Xu, "Enabling performance intelligence for application adaptation in the Future Internet" Journal of Communication and Networks. vol. 13, pp. 591-601, Decemeber 2011

[6] Wen-Tsai Sung, Jui-Ho Chen, Kung-Wei Chang, "Mobile Physiological Measurement Platform With Cloud and Analysis Functions Implemented via IPSO" Sensors Journal, vol. 14, pp. 111- 123, January 2014

[7] Huifang Li, Lu Zhang, Rui Jiang, "Study of manufacturing cloud service matching algorithm based on OWL-S" The 26th Chinese Control and Decision Conference (2014 CCDC). Changsha, pp. 4155-4160, June 2014

[8] Xiaopeng Lin, Yiyang Li, Huaiyu Dai, Donghui Guo, "Architecture of Web-EDA System Based on Cloud Computing and Application for Project Management of IC Design" 2010 International Conference of Anti-Counterfeiting Security and Identification in Communication. Chengdu, pp. 150-153, July 2010

Task Allocation Strategy Based on Improved Quad-tree in the Cloud

Guobin Lan
Lab center, Jiujiang university, Jiujiang, China

***Abstract*—To make full use of the scattered resources in cloud network, this paper proposes an improved quad-tree based cloud computing task allocation strategy on the basis of traditional allocation strategy. The strategy makes effectively use of the temporary network nodes in cloud web by creating a parallel computing model based on improved quad-tree. In the model, a pointer to the next brother node is added to every intermediate node, effectively reducing the task allocations workload of the root node. The simulation results show that the proposed strategy presents good computing performance and high stability.**

***Index Terms*—task allocation strategy; quad-tree; cloud computing**

I. INTRODUCTION

Cloud computing is generally accepted as the main trend of calculation development. Through the full use of network computing and network storage, collaborative work and resource sharing of all the computing and storage resources over large areas is achieved by cloud computing. Cloud computing has the advantages of high efficiency and low cost and the increasing scale of cloud computing also makes the distributed computation of some very large scale data possible.

In recent yearstask allocation strategy in distributed environment has developed rapidly. Ref [1], [2], and [3] adopt ant colony algorithm and improved algorithm for the purpose of improving the performance of task allocation strategy through finding optimal resource allocation for each job in network environment. Ref [4], [5] and [6] adopt heuristic task scheduling strategy to implement priority scheduling through some specific approaches (such as standard variance); further more, Ref [5] also gives due consideration to the scheduling of low-priority work on the basis of priority scheduling strategy. Ref [7] and [8] use Directed Acyclic Graph (DAG) to describe the processor in the cloud and adopt hierarchical processing to the whole system, thus considerably alleviating the central processor's workload.

On the basis of DAG, this paper also uses an improved quad-tree structure to describe the processors in the cloud. Meanwhile, this paper constructs a cloud computing task allocation model based on improved quad-tree structure. For clouds of different scale, we can describe them by adjusting the depth of the quad-tree.

Here are the major contributions of this paper: (1)

1) This paper proposes an improved quad-tree structure and reduces the calculation amount of Root Note through inserting an indicator pointing to the next brother structure in the intermediate node of traditional quad-tree structure, thus improving the efficiency of Root Node.
2) This paper discusses the task allocation model based on improved quad-tree structure in detail.

II. IMPROVED QUADTREE STRUCTURE

Quad-tree structure is a commonly used data structure, which features good convergence speed, higher computational efficiency, etc. However, some problems still might arise from the use of qual-tree structure. First of all, an error occurring to one intermediate node of quad-tree may affect the whole quad-tree structure and result in the reframing of quad-tree model. Secondly, synchronous differences between the root node and intermediate node may cause such scenario: prior to receiving the data frames sent from the intermediate node, the root node might have assigned some new computing tasks to the intermediate node which is already in busy state. Thirdly, when a mistake occurs at the leaf node (including net mistakes and some irresistible natural error etc.), the task assumed by leaf node must be redistributed by the root node, and consequently affect the efficiency of the root node.

For the problem 1, it is agreed in this paper that temporary node can only be used as leaf node. When errors occur at temporary node for some reason, a new leaf node will be selected from other temporary nodes by its parent node. In this way, the possibility that the instability of leaf nodes may lead to the instability of the whole quad-tree structure will be eliminated.

For the problem 2, this paper inserts a pointer to the next brother node in the same layer in the intermediate node (i.e. the node apart from the root node and the leaf node), as is shown in Figure 1. When a new task is allocated to the intermediate node which is in busy condition, the task will be directly transferred to a brother node. In order to prevent the Root Node from entering infinite waiting caused by the endless loop generated during the transfer process, this paper uses the weighting method to automatically increase the priority value of the computing task which is transferred to the brother node from the intermediate node while it is in busy condition.

To address the problem 3, the strategy used in this paper is sending redundant computing tasks. In order to make full use of the computing resources in dedicated servers, the intermediate node will merely allocate the received redundant computing jobs to temporary node. Suppose that Q_1 is a random intermediate node and Q_{14} is Q_1's son node, as well as a temporary node and assume that the credibility of Q_1 is $p_{Q_{ij}}$. Then the specific process of sending redundancy computing tasks will be analyzed: (1)

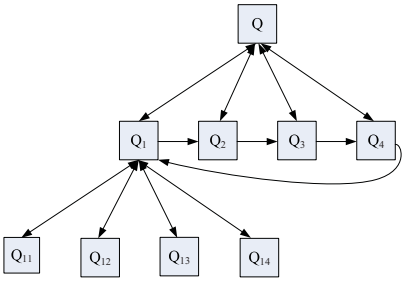


Fig. 1: Improved quad-tree for parallel computing

1) Q_1 sends job_i to Q_{14} and updates the temp attribute of job_i. Set it as $p_{Q_{ij}}$, namely set jobi.temp=$p_{Q_{ij}}$;
2) Judge whether job_i.temp is equal to or larger than $\Re$(The value of $\Re$ is based on past experience and it can be adjusted in accordance with need). If the result is yes, then finish the process of transferring redundant computing job_i; otherwise, turn to 3;
3) Send the redundant computing job (job_i') of job_i to his brother node (note it with Q_2);
4) Q_2 send job_i' to his son node (note it with Q_{24}) and update the temp attribute of the job_i'. Set it as the sum of and job_i'.temp, namely set job_i'.temp = job_i'.temp + $p_{Q_{24}}$;
5) Judge whether job_i'.temp is equal to or larger than $\Re$. If the result is yes, then finish the process of transferring redundant computing job; otherwise, turn to 3.

From the process of transferring redundant computing job, we draw a conclusion that the sum of the credibility of the son nodes in the process must achieve formula (1). n means the number of the son node. Therefore; in order to ensure that the redundant computing tasks in the transferring process won't fall into death cycle, the value of n should be reasonably as large as possible. This paper agreed that a temporary node can be a leaf node only when its credibility is equal to or larger than the threshold. Threshold values directly affect the efficiency of the system; if the value is too small, there will be too many redundant computing tasks to forward and the value of n will be too large; if the value is too large, the purpose of redundant calculation will not be achieved and the value of n will be too small. So the threshold value shall be adjusted by experience.

$$\sum_{k=1}^{n} p_{Q_k} \geq \Re, n = 1, 2, 3 \cdots . \tag{1}$$

III. CLOUD COMPUTING TASK ALLOCATION MODEL

The main goals for cloud computing task allocation are: (1) improve the efficiency of task allocation, (2) make an effective use of the scattered computing resources. In order to achieve these two goals, this paper presents an improved protection measure to maintain the stability of the quad-tree structure, i.e. to ensure that the first two son nodes in each branch of the quad-tree structure must be dedicated servers. Our whole model includes three parts, Root Node, Center Node and Leaf Node.

A. Related Concepts

Root Node: application server, located at the top of the quad-tree structure and is responsible for receiving request service from clients, returning the results of request to the client, work reception, work division and work distribution, etc.

Center Node: intermediate Node, made up of dedicated servers. It's main function is to receive task from the Root Node, allocate the task to Leaf Node, feed back the completed task information to Root Node, send and receive redundancy computing tasks, and maintain the network of his son Nodes, etc.;

Leaf Node: terminal node, including dedicated server nodes and computing nodes formed by temporary nodes and mainly responsible for calculation and returning the calculation results.

Task state table: the application server divides each of the received job into several jobs, each of which maintains a state table. The attributes of state table include task name, task id, son node, task priority, task completion status, etc.;

Node credibility: the credibility of the node is determined by Weibull distribution and the error rate within fixed time interval. The specific calculation is shown in figure (2). In the figure, λ is the scale parameter, α is the shape parameter, and t_0 is a fixed time interval.

$$p(j) = 1 - \sum_{j=0}^{n} \{\lambda^{\alpha} \alpha [(jt_0 + t_0)^{\alpha - 1} - jt_0^{\alpha - 1}]\}, j = 1, 2, 3 \cdots \tag{2}$$

Node attribute: Each node has seven attributes and they can be expressed as Node (Id, Com_M,Com_C,p,F,S,Next). Among them, Id is the only identification of each node in the cloud computing network; Com_M is the maximum calculation ability of the node and is determined by the hardware performance; Com_C is the current calculation ability of the node and is determined by current idle computing resources of the node; p refers to the credibility of the node. In this paper, it's agreed that the node is dedicated server when p equals 1. When p varies within the range of [p_0, 1), it indicates that the node is non-dedicated server, but can serve as leaf note; when p is in the range of [0,p_0), it indicates that the node is non-dedicated server, and can't serve as leaf note; F is the father node of the node; S means son node; Next refers to next brother node.

B. The Establishment of Cloud Computing Task Allocation Model

Cloud computing model must be established before calculation. This section establishes cloud computing model on the basis of quad-tree model. The depth of quad-tree is determined by the number of the nodes in the network, while for the convenience of discussion, the depth of quad-tree adopted in this paper is 3.

The establishment of quad-tree based task allocation model is the premise of task allocation. There are two questions needed to be considered while establishing the task allocation model. The first is to maintain the stability of the model, which means that changes in temporary node shall not affect the structure of quad-tree. The other is to try to make full use of the computing resources in temporary node. The following conventions are made for the abovementioned two problems: 1) among the four leaf nodes in the same branch, there must be two or more server nodes; 2) when intermediate nodes are allocating tasks to son nodes, redundant tasks should be allocated to the next node if the target for allocation is a temporary node.

The cloud computing task allocation model is established in the formation of cloud computing network and the specific establishment procedures are as follow:

Step 1 Root Node sends modeling broadcast to the cloud computing network; the nodes having received this broadcast should reply to Root Node;

Step 2 Root Node classifies all the received node packages into two tables according to the value of p, i.e. dedicated server and non-dedicated server, and then gives a descending order to each table according to the attribute of Com_C;

Step 3 select four values from the table of dedicated server from top to bottom to be the son nodes of Root Node and modify the attribute value of the four nodes;

Step 4 select nodes to be the son nodes of the above-mentioned nodes in turn according to the depth requirement of quad-tree;

Step 5 select two nodes from the table of dedicated server nodes to be leaf nodes and select enough nodes from the table of non-dedicated server nodes to be leaf node;

IV. Research on the Cloud Computing Task Allocation Based on Improved Quad-tree

The three components of task allocation model have different functions. In order to better demonstrate cloud computing task strategy that adopts quad-tree algorithm, we will give a detailed description on the function and algorithm of intermediate node. As a bridge between task transfer and task allocation, intermediate node plays an important role. Its functions include the following respects:

1) Network maintenance and network update; send network maintenance packets to parent node and son nodes regularly. When a leaf node changes, it should choose new leaf node from corresponding node table and send a packet for network update.

2) Task allocation. The tasks to be allocated include the tasks allocated from Root Node and the redundant computing tasks transferred from brother nodes. The former one has the priority to be allocated to dedicated servers; however, it needs to send redundant computing tasks if it is allocated to temporary nodes. The latter one has the priority to be allocated to temporary nodes, and it still needs to send redundant computing task if it doesn't meet the termination conditions.

3) Send redundant computing task, which includes two parts. One part is the tasks forwarded due to the busy condition of the node and the other part is redundant computing conducted on the tasks which are allocated to temporary nodes and fail to achieve termination conditions. For the detailed description of task allocation and sending redundant computing task, see algorithm 1, in which (1)

1) 1 intermediate node received a job;
2) 2-4, the task has the priority to be allocated to dedicated servers if the job comes from the father node;
3) 5-7, in case the task queues of dedicated servers is full, judge whether the queue of the temporary nodes is free and allocate tasks to them if they are;
4) 8-11, judge whether the termination conditions are met. If it is met, then task allocation is finished, otherwise it should send a redundant computing task;
5) 12-16, if the task queue of temporary node is full, then the task is weighed and sent to next brother node;
6) 17, the task comes from a brother node;
7) 18-19, if the task queue of temporary node is free, then allocate the task;
8) 20-23, update the temp attribute of the task if the termination conditions are not met, and send redundant computing task at the same time;
9) 24-27, if the task queue of the temporary node is full, then weigh the task and transfer it to next brother node;
10) 28-29, allocation finished

V. Experiment

In order to verify the superiority and feasibility of improved quad-tree task allocation strategy, there are two simulation experiments conducted in this section. The first experiment makes a comparison between improved quad-tree task allocation strategy and the similar algorithm, so as to verify the validity of the algorithm; the other one is to verify the algorithm's contribution on improving the working efficiency of the Root Node.

For the purpose of better comparison, the First Come First Served (FCFS) task allocation strategy and ordinary quad-tree based task allocation strategy (Quad-Tree) are introduced. The main idea of FCFS is to allocate task to each node according to the sequence of the task request. In ordinary quad-tree task allocation strategy, there's no pointer to the next brother node in the intermediate nodes. For the sake of better discussion, quad-tree adopts three-layered structure, with 16 leaf nodes.

In the simulation experiment, we first record the time required for the three allocation strategies to complete different task loads and the results are shown in figure 2. In the figure, horizontal ordinate refers to the number of allocated tasks and its unit is piece; the ordinate refers to the time required to complete corresponding tasks and the unit is ms.

Through analysis, the following conclusions are drawn: (1) in the case of small task load, the task allocation strategy using improved quad-tree has a bit poor performance mainly

Algorithm 1 mission assignment and redundancy

Input: job_i
Output: $Distribute(job_i), SendRredundancy(job_i)$

while (job_i) **do**
  if (Judge(job_i)) **then**
    if (Idle($Node_{DedicatedServer}$)) **then**
      Distribute(job_i)
    else
      if (idle($Node_{Temporary}$)) **then**
        Distribute(job_i)
        if $P_{Temporary} < \Re$ **then**
          Refresh($job_i.temp$)
          SendRedundancy(job_i)
        end if
      else
        UpPriority(job_i)
        SendToNext(job_i)
      end if
    end if
  else
    if (Idle($Node_{Temporary}$)) **then**
      Distribute(job_i)
      if ($P_{Temporary} + job_i.temp < \Re$) **then**
        $job_i.temp = P_{Temporary} + job_i.temp$
        SendRedundancy(job_i)
      end if
    else
      UpPriority(job_i)
      SendRedundancy(job_i)
    end if
  end if
end while

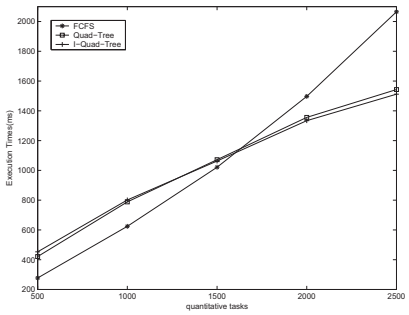


Fig. 2: The time needed for the complete quantitative tasks chart

because that the establishment of quad-tree model and the communication among the nodes will consume a considerable amount of computing resources; (2) in case of large task load, the task allocation strategy using improved quad-tree presents excellent computing performance.

VI. CONCLUSIONS

This paper presents an improved quad-tree task allocation strategy and gives a detailed introduction to improved quad-tree method and the use of the strategy. Experiment results show that task allocation strategy with improved quad-tree structure can effectively improve the efficiency of traditional task allocation and can effectively improve the working efficiency of the Root node.

But there are still some limitations in the proposed task allocation strategy:

1) The task allocation model is established on the basis of the stability of dedicated servers. If the dedicated server appears to be unstable, the performance of the strategy will be greatly reduced.

2) The superiority of the proposed task allocation strategy is prevented from being fully displayed when the task load to be processed is small. How to find a balance between the task allocation strategy and other distribution strategies is the focus of future work.

REFERENCES

[1] S. Lorpunmanee, M. N. Sap, A. H. Abdullah, C. Chompoo-inwai, An Ant Colony Optimization for Dynamic Job Scheduling in Grid Environment, International Journal of Computer and Information Science and Engineering. 2007, Vol. 1, 469-476

[2] S. Umarani, L. M. Nithya, A. Shanmugam, Efficient Multiple Ant Colony Algorithm for Job Scheduling In Grid Environment, International Journal of Computer Science and Information Technologies, 2012, Vol. 3(2), 3388-3393

[3] M. Venkataramana,N. R. S. Raghavan, Ant colony-based algorithms for scheduling parallel batch processors with incompatible job families, International Journal of Mathematics in Operational Research, 2010, Vol. 2(1), 73-98

[4] E. U. Munir, J. Li, S. Shi, Z. Zou, Q. Rasool, A new heuristic for task scheduling in heterogeneous computing environment, Journal of Zhejiang University. Science A. 2008, Vol.9(12), 1715-1723

[5] T. He, S. Chen, H. Kim, L. Tong, K. Lee, Scheduling Parallel Tasks onto Opportunistically Available Cloud Resources, Cloud Computing (CLOUD), 2012 IEEE 5th International Conference on, 2012, 180-187

[6] M. I. Daoud, N. Kharma, A hybrid heuristic-genetic algorithm for task scheduling in heterogeneous processor networks, Journal of Parallel and Distributed Computing, 2011, Vol.71(11), 1518-1531

[7] J. Kang, S. Ranka, Energy-Efficient Dynamic Scheduling on Parallel Machines, Lecture Notes in Computer Science, 2008, Vol.5374, 208-219

[8] N. A. Bahnasawy,F. Omara, M. A. Koutb, M. Mosa, Optimization procedure for algorithms of task scheduling in high performance heterogeneous distributed, Egyptian Informatics Journal, 2011, Vol.12(3), 219-229

[9] H. V. Singh, S. Rai, A. Mohan, S. P. Singh, Robust copyright marking using Weibull distribution, Computers & Electrical Engineering, 2011, Vol.37, No.5, 714-728

[10] W. Lee, A. G. L. Borthwick, P. H. Taylor, A fast adaptive quadtree scheme for a two-layer shallow water model, Journal of Computational Physics, 2011, Vol.230, No.12, 4848-4870

Power and Performance Analysis of the Graph 500 Benchmark on the Single-chip Cloud Computer

Zhiquan Lai*, King Tin Lam†, Cho-Li Wang†, Jinshu Su‡*
*College of Computer, National University of Defense Technology, Changsha, China
†Department of Computer Science, The University of Hong Kong, Hong Kong, China
‡National Key Laboratory of Parallel and Distributed Processing (PDL), Changsha, China
zqlai@nudt.edu.cn, {ktlam, clwang}@cs.hku.hk, sjs@nudt.edu.cn

***Abstract*—The concerns of data-intensiveness and energy awareness are actively reshaping the design of high-performance computing (HPC) systems nowadays. The Graph500 is a widely adopted benchmark for evaluating the performance of computing systems for data-intensive workloads. In this paper, we introduce a data-parallel implementation of Graph500 on the Intel Single-chip Cloud Computer (SCC). The SCC features a non-coherent many-core architecture and multi-domain on-chip DVFS support for dynamic power management. With our custom-made shared virtual memory programming library, memory sharing among threads is done efficiently via the shared physical memory (SPM) while the library has taken care of the coherence. We conduct an in-depth study on the power and performance characteristics of the Graph500 workloads running on this system with varying system scales and power states. Our experimental results are insightful for the design of energy-efficient many-core systems for data-intensive applications.**

***Keywords*—performance analysis; power efficiency; DVFS; Graph 500; data-intensive computing; many-core computing**

I. Introduction

A fundamental shift in supercomputing research focus is well underway. Apart from steering towards the best computational performance, much more attention has been drawn to data-intensiveness and energy awareness, both of which are playing increasingly important roles to influence the design of future HPC systems and data centers. The Graph500 [1] is a widely used benchmark for rating supercomputer systems for data-intensive workloads. Graphs are a core part of most data analytics. Compared with compute-intensive benchmarks like HPL (High Performance Linpack) which the Top500 List is based on, Graph500 models complex data problems by performing breadth-first search (BFS) on a large-scale graph. Apart from the data-intensiveness aspect, energy efficiency is another vital design constraint on HPC systems and data centers. From a "green computing" perspective, energy consumption is as important as performance [2]. A new definition of supercomputer ranking is one that sorts computer systems by performance per watt for data-intensive workloads.

There is rich literature on parallelization and optimization of Graph500 for various parallel architectures and programming models [3]–[5]. This paper, to the best of our knowledge, is the first attempt to tailor Graph500 to a (virtual) shared memory programming model on top of a non-coherent multicore architecture exemplified by the Intel Single-chip Cloud Computer (SCC) [6]. Fabricating a plethora of compute cores (a many-core approach) onto a single chip is the way to continue the performance leap for multiprocessors. This however also risks leading to a power-hungry chip eating up a great portion of the system power, so the state-of-the-art processors are typically equipped with DVFS (dynamic voltage and frequency scaling) components for dynamic power savings. The Intel SCC is a typical DVFS-enabled many-core system, but has not been evaluated with Graph500 for power and performance characterization.

To close this research gap, we introduce in this paper a data-level parallel implementation of the Graph500 on the SCC. By exploiting the shared physical memory (SPM) on the SCC, an efficient shared virtual memory (SVM) runtime is developed to support a shared memory programming model akin to POSIX-based multithreaded programming. We port Graph500 to this SVM runtime and perform a thorough evaluation on the power and performance characteristics of Graph500 using various scales of cores and power states. The experimental results provide some useful hints for energy-efficient programming and system design.

For the rest of this paper, Section II details the parallel Graph500 implementation on the SCC. Section III describes the evaluation methodology. Power/performance benchmarking results and analysis are presented in Section IV. Finally, we conclude the paper in Section V.

II. Graph500 on Intel SCC

A. Graph500

Researchers observed that data-intensive supercomputing applications are of growing importance to represent current HPC workloads, but existing benchmarks did not provide useful information for evaluating supercomputing systems for these applications. In order to guide the design of hardware architectures and software systems to support data-intensive workloads, the Graph500 was proposed and developed [1]. The workflow of Graph500 is described in Algorithm 1. Its kernel is basically a loop of breadth-first searches (BFSes) over a large-scale graph.

In the original Graph500, only step 2 and step 4.2 (i.e. kernel 1 and kernel 2) are timed and included in the performance information. However, the total execution time is also indicative of the performance. In this study, we inspect the power and performance characteristics in terms of both the entire execution and the two kernels alone.

B. SPM-based Shared Virtual Memory on the SCC

Today, hardware implementation of cache coherency and the use of a single operating system for managing all cores are the standard options for a traditional multicore system. However, a further growth of core count per chip implies a "coherence wall" problem—the chip complexity grows to a point beyond which the additional cores are not useful in a single parallel program. Therefore, researchers have been proposing a so-called "cluster-on-chip" architecture to make future many-core systems inherently scalable. The game-changing idea is to eschew hardware cache coherency support and to adopt a software-oriented message passing architecture instead. Intel's SCC is an example of such an architecture. The SCC processor consists of 48 non-coherent memory-coupled P54C cores. Each core has private 16KB L1 and 256KB L2 caches. All cores can access the shared physical memory (SPM) off the chip. The SCC can be configured to run one operating system instance per core, each of which is assigned a private memory region in the main memory. To parallel this architecture, a more scalable *multikernel* operating system (OS) design approach has been proposed and realized into the Barrelfish OS [7]. By directly mapping the shared virtual address space to some cacheable SPM region, all cores can share memory in a concurrent and efficient manner using virtual addresses (without the need of explicitly sending messages), provided that either the processor caches are disabled or the cache coherency is guaranteed by software.

In our previous work [8], we designed and implemented a shared virtual memory library called *Rhymes (Runtime with HYbrid MEmory Sharing)* for the Barrelfish OS. In Rhymes, there are two operational modes selectively assigned to each shared memory page—SPM mode or DSM mode. In the DSM (Distributed Shared Memory) mode, there exist a cached copy of each virtual memory page in the private memory space of each OS instance. The coherence of these cached copies is maintained via a scope consistency protocol employing traditional software DSM techniques like access trapping through page faults, twinning, diffing and passing of write notices. In contrast, the SPM mode represents an approach bearing much similarity to traditional shared-memory (SMP or multi-core) systems. In the SPM mode, there is only copy allocated in the shared DRAM region for each virtual memory page. SPM-mapped pages are allocated in the shared DRAM of the SCC and synchronized based on release consistency using proper invalidate and flush instructions.

The Graph500 (version 2.1.4) is ported to Rhymes and run in the SPM mode (i.e. all shared memory pages are assigned the SPM mode). We chose the OpenMP branch for porting since it resembles our programming model most closely. Our programming model exploits the data-level parallelism and follows the traditional SPMD (single program, multiple data) pattern. The master core is responsible for distributing the load to slave cores and collecting the computation results from them. The porting process involves several areas of modifications summarized below:

1) *Explicit load division*: Rhymes does not provide an loop-parallelizing compiler as in the case of OpenMP (omp). Rather than relying on omp compiler directives, the parallelization of a loop among parallel processes has to be done programmatically—a range of loop iterations formulated as a function of the process rank, is assigned to each core for handling.
2) *Shared memory allocation*: for buffers which are being shared across processes, they need to be allocated using our custom malloc API (`rhy_malloc`) rather than declaring them as shared variables in OpenMP.
3) *Synchronization*: we have to revamp all the code that originally rely on OpenMP synchronization constructs for protecting accesses against data races into a version that employs our synchronization routines, namely `rhy_lock`, `rhy_unlock` and `rhy_barrier` for achieving the same effect.

Algorithm 1: Algorithm of Graph500

Input:
$SCALE$: the vertices scale, 2^{SCALE} vertices
$EDGE$: the edge factor, $EDGE \cdot 2^{SCALE}$ edges

```
1 begin
2     Step 1: Generate the edge list.
3     Step 2: Construct a graph from the edge list. kernel 1
4     Step 3: Randomly sample 64 unique search keys.
5     Step 4: for each search key do
6         Step 4.1: Compute the parent array. kernel 2
7         Step 4.2: Verify the parent array.
8     Step 5: Compute and output performance information.
```

III. Evaluation Methodology

The problem size of Graph500 is set as follows: $SCALE$ = 18 (262,144 vertices) and $EDGE$ = 16 (4,194,304 edges). We use gcc version 4.4.3 to compile programs with the O3 optimization level. During the experiments, the clock frequencies of the network-on-chip (NoC) and memory controllers (MCs) are both fixed at 800MHz.

To avoid the overhead introduced by power measurement, we measure the power consumption of the SCC outside the chip. As shown in Fig. 1, we view the chip as a "black box" and read the current and voltage values from the board management controller (BMC) using the management console

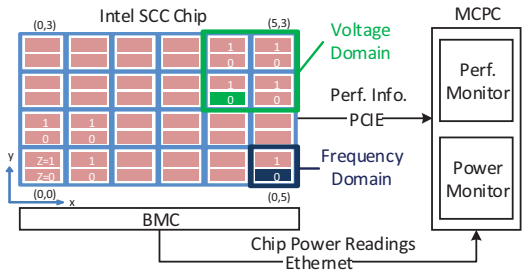


Fig. 1. Power domains of Intel SCC and power measurement setting

PC (MCPC) during the execution of each application. As shown in Fig. 1, the BMC is connected via Ethernet to the MCPC, from which the power values are read. The PCIE bus connects the SCC to the MCPC. The power states of the chip are recorded into a log file on the MCPC. A power sampling rate of about 3.3 samples per second has sufficed for our evaluation. The average power of an execution is estimated by taking arithmetic mean of all sampled power values.

The SCC CPU cores are set using the DVFS mechanism to four different power states (voltage/frequency settings), 800MHz/1.1V, 533MHz/0.9V, 400MHz/0.8V and 320MHz/0.8V. For each power state, we launch the benchmark on 1, 2, 4, 8, 16, 32 and 48 cores to inspect the power and performance characteristics.

IV. Experimental Results and Analysis

As an example, Fig. 2 presents the instantaneous power when Graph500 is executed on 48 cores at 800MHz/1.1V. P_{core}, P_{ddr} and P_{mc} denote the power consumptions of the CPU cores (including the network on-chip), the DDR3 memory modules and the four memory controllers respectively. P is then the sum of P_{core}, P_{ddr} and P_{mc}. From the figure, we can see that P_{core} and P_{ddr} are changing over time while P_{mc} is nearly static.

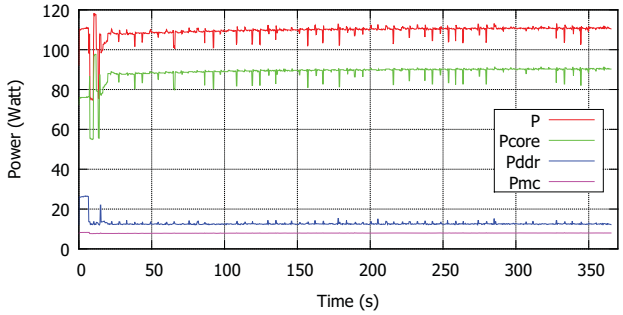


Fig. 2. The power characteristic of Intel SCC when Graph500 executes at 800MHz/1.1V on 48 cores

A. Performance Characteristics

To evaluate the scalability of the ported Graph500 benchmark, we show its runtime performance against the number of cores for the four power states in Fig. 4. We present the total execution time and the runtime of the two kernels of Graph500. For kernel 2, we present the average execution time of the 64 BFS loops.

From Fig. 4, we can observe that for a fixed number of CPU cores, the execution time generally becomes longer in a lower power state. For each particular power state, the performance increases with the number of cores. Compared with a single core, the execution on 48 cores achieves on average 1.69x, 4.10x and 3.99x maximum speedups for the whole benchmark, kernel 1 and kernel 2 respectively. However, the whole benchmark sometimes (in Fig.4(a)) cannot achieve the best runtime performance on 48 cores. For example, at 800MHz/1.1V and 320MHz/0.8, the total execution time on 48 cores is a little longer than that on 32 cores. This phenomenon is also observed in the kernel 1 and kernel 2 cases. The reasons behind might be the increase in parallel overhead of the SPM-based software coherence protocol or insufficient problem size when scaling up the number of cores. The whole benchmark achieves the best speedup (up to 1.72x) at 400MHz on 48 cores, 4.65x for kernel 1 at 320MHz on 32 cores and 4.23x for kernel 2 at 400MHz on 32 cores.

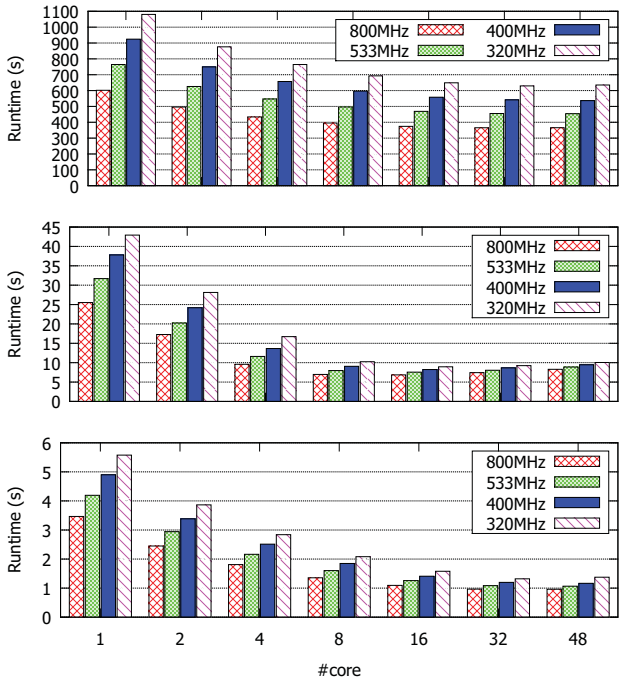
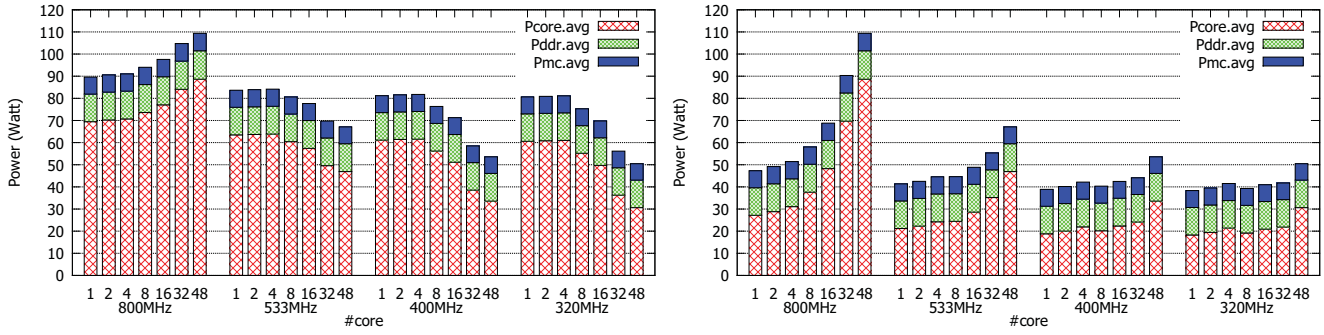


Fig. 4. Performance of Graph500 at different power states using a varying number of cores. From top to bottom, (a) Total execution time of Graph500; (b) Execution time of kernel 1; (c) Average execution time of kernel 2

B. Power Characteristics

Fig. 3 presents the power characteristics of Graph500 on a varying number of cores at different power states. Fig. 3(a) shows the measured power, we present the average values of P_{core}, P_{ddr} and P_{mc} during the execution. However, one might



(a) Measured power

(b) Estimated power excluding the standby dynamic power

Fig. 3. Power comparison of Graph500 at different power states on different number of cores

argue that the power of those non-activated cores (or standby cores) should not be counted in the measurements. Thus, we estimate the per-core dynamic powers of standby cores and exclude them from the measurements[1], deriving Fig. 3(b).

In both Fig. 3(a) and Fig. 3(b), the power at 800MHz is the most representative. The total power P almost linearly increases with the number of cores since a higher power state consumes more power. For the same number of core, e.g. 48 cores, a lower power state gives smaller P while some portions of the power (e.g. P_{ddr} and P_{mc}) remain the same as they are not affected by DVFS. The highest P could reach up to 110 watts at 800MHz on 48 cores.

C. Energy and EDP Characteristics

Energy consumption and *energy-delay product (EDP)* are two important power efficiency metrics [10] worthy studying here; their variations are shown in Fig. 5 and Fig. 6.

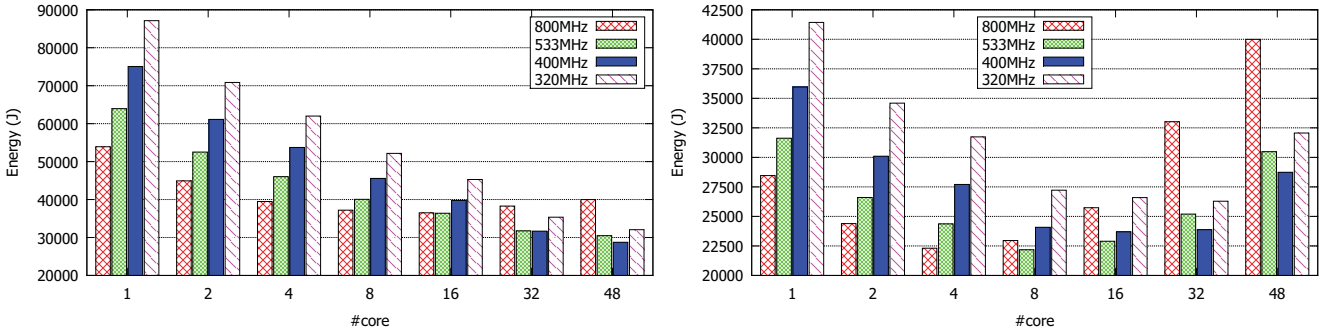
In Fig. 5(a), if we consider the standby power as the system power, we always get less energy on more cores for the same power state (800MHz is an exception). This is because the standby power is even larger than these three power states. Thus excluding the standby dynamic power is more meaningful and reasonable here. Fig. 5(b) presents the energy variation after excluding the standby dynamic power. For a fixed number of cores, the best power state for attaining the least energy usage is different. For example, 400MHz/0.8V is the best for 32 and 48 cores while 533MHz for 8 and 16 cores. For a fixed power state, the best number of cores to keep the energy consumption least also varies. A lower power state is always associated with a larger number of cores for the best result. Using 32 cores is the best for 320MHz, while four cores for 800MHz. Execution on 8 cores at 533MHz consumes the least energy, saving 46.48% compared with the largest energy usage on one core at 320MHz.

[1]As the standby cores are initialized to 533MHz/1.1, we first measure P_{core} at this power state (it is about 66.95 watts). Then we set all cores to the lowest power state (100MHz/0.8V) which is safe and supported by the SCC [9]. P_{core} is about 23.75 watts. Hence for each core, the dynamic standby power is (66.95-23.75)/48 = 0.90 watts.

For EDP, the situation is a bit different because the execution time in this case is weighted more heavily. As shown in Fig. 6(b), the power state for the least EDP is 533MHz for 32 and 48 cores; the number of cores giving the minimal EDP is 32 for 320MHz but becomes 8 for 800MHz. Execution on eight cores at 800MHz achieves the least EDP, saving 79.71% compared with the largest EDP in the single-core case at 320MHz. In short, the optimal power state and core count are dependent on the optimization goals, say targeting the least energy footprint or EDP instead.

V. Conclusion

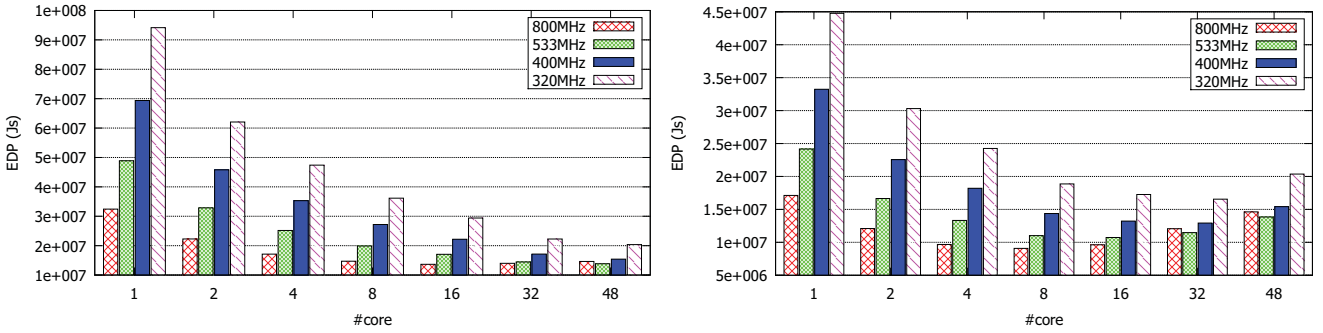
In this paper, we present a data-parallel implementation of Graph500 for the non-coherent memory-coupled many-core architecture exemplified by the Intel SCC. We port the benchmark to a shared virtual memory programming model which exploits the shared physical memory (SPM) of the SCC. We analyze the power and performance characteristics of Graph500 against the scaling of cores at different power states. Experimental results show that the kernels of Graph500 can achieve up to 4.65 times speedup; an appropriate scale of cores and voltage/frequency setting can bring about as much as 46.48% energy saving and 79.71% EDP reduction. We believe that this work has provided some insights into designing an energy-efficient many-core system for data-intensive applications. First, regarding the performance, we observe unsatisfactory speedup and scalability of the SCC system for data-intensive workloads, calling for better co-design of many-core hardware and software—perhaps future systems should have larger on-chip memory and write-combine buffers plus more efficient, fine-grained invalidate/flush instructions for curbing the memory wall effects of such workloads and the parallel overhead of software-managed coherence. Second, regarding the energy efficiency, while some recent study [11] (based on compiler analysis and high-level modeling) suggests diminishing returns from the use of DVFS, our work reaffirms the usefulness of DVFS for today's supercomputing systems because of the increasing number of cores per chip (taking up a great portion of the system power) and the considerable



(a) Energy with measured P

(b) Energy with estimated P excluding the standby dynamic power

Fig. 5. Energy comparison of Graph500 at different power states on different number of cores



(a) EDP with measured P

(b) EDP with estimated P excluding the standby dynamic power

Fig. 6. EDP comparison of Graph500 at different power states on different number of cores

power saving opportunities available in data-intensive workloads such as complex data graph computations. We should keep improving the DVFS and power modeling or prediction techniques for realizing green many-core computing systems.

ACKNOWLEDGMENT

This work is supported by Hong Kong RGC grant HKU 716712E, Program for Changjiang Scholars and Innovative Research Team in University (PCSIRT, No. IRT1012) and Aid Program for Science and Technology Innovative Research Team in Higher Educational Institutions of Hunan Province (No. 11JJ7003). Special thanks go to Intel China Center of Parallel Computing (ICCPC) and Beijing Soft Tech Technologies Co., Ltd. for providing us with support services of the SCC platform in their Wuxi data center.

REFERENCES

[1] "Graph 500 benchmark." [Online]. Available: http://www.graph500.org

[2] K. Asanovic, R. Bodik, J. Demmel, T. Keaveny, K. Keutzer, J. Kubiatowicz, N. Morgan, D. Patterson, K. Sen, J. Wawrzynek, D. Wessel, and K. Yelick, "A view of the parallel computing landscape," *Commun. ACM*, vol. 52, no. 10, pp. 56–67, 2009.

[3] F. Checconi and F. Petrini, "Massive data analytics: the graph 500 on IBM Blue Gene/Q," *IBM J. Res. Dev.*, vol. 57, no. 1, pp. 111–121, 2013.

[4] T. Gao, Y. Lu, and G. Suo, "Using MIC to accelerate a typical data-intensive application: The breadth-first search," in *Proc. 27th Int. Parallel and Distributed Process. Symp. Workshops & PhD Forum (IPDPSW)*, 2013, pp. 1117–1125.

[5] J. Jose, S. Potluri, K. Tomko, and D. Panda, "Designing scalable Graph500 benchmark with hybrid MPI+OpenSHMEM programming models," in *Proc. 28th Int. Supercomputing Conf. (ISC)*, vol. 7905, 2013, pp. 109–124.

[6] "SCC external architecture specification (EAS) (revision 0.94)," Intel Labs, Tech. Rep., 2010.

[7] A. Baumann, P. Barham, P.-E. Dagand, T. Harris, R. Isaacs, S. Peter, T. Roscoe, A. Schüpbach, and A. Singhania, "The multikernel: A new OS architecture for scalable multicore systems," in *Proc. ACM SIGOPS 22nd Symp. Operating Syst. Principles (SOSP'09)*, 2009, pp. 29–44.

[8] K. T. Lam, J. Shi, D. Hung, C.-L. Wang, Y. Yan, and W. Zhu, "Rhymes: A shared virtual memory system for non-coherent tiled many-core architectures," in *Proc. 20th IEEE Int. Conf. Parallel and Distributed Syst. (ICPADS'14)*, 2014, (in press).

[9] Z. Lai, K. T. Lam, C.-L. Wang, J. Su, Y. Yan, and W. Zhu, "Latency-aware dynamic voltage and frequency scaling on many-core architectures for data-intensive applications," in *Proc. Int. Conf. Cloud Computing and Big Data (CloudCom-Asia'13)*, 2013, pp. 78–83.

[10] V. W. Freeh, D. K. Lowenthal, F. Pan, N. Kappiah, R. Springer, B. L. Rountree, and M. E. Femal, "Analyzing the energy-time trade-off in high-performance computing applications," *IEEE Trans. Parallel Distrib. Syst.*, vol. 18, no. 6, pp. 835–848, 2007.

[11] T. Yuki and S. Rajopadhye, "Folklore confirmed: Compiling for speed = compiling for energy," in *Proc. 26th Int. Workshop on Languages and Compilers for Parallel Computing (LCPC)*, 2013, pp. 169–184.

Fully Homomorphic Symmetric Scheme without Bootstrapping

Nitesh Aggarwal[1], Dr CP Gupta[2], Iti Sharma[3]
Department of Computer Sciences and Engineering
University College of Engineering
Rajasthan Technical University
KOTA-324010, INDIA
[1]nit007agarwal@gmail.com [2]guptacp2@rediffmail.com [3]itisharma.uce@gmail.com

***Abstract*- Capability of operating over encrypted data makes Fully Homomorphic Encryption(FHE) the Holy Grail for secure data processing applications. Though many applications need only secret keys, FHE has not been achieved properly through symmetric cryptography. Major hurdle is the need to refresh noisy ciphertexts which essentially requires public key and bootstrapping. We introduce a refreshing procedure to make a somewhat homomorphic scheme, fully homomorphic without requiring bootstrapping. Our scheme uses symmetric keys and has performance superior to existing public-key schemes.**

Keywords: Symmetric fully homomorphic encryption, Refreshing, Cloud computing

I. INTRODUCTION

Need of cloud to manipulate and manage data is increasing rapidly for sharing resources. It is financially beneficial to store data with a third party, the cloud provider. However, storing data on third party infrastructure poses risks of data disclosure during retrieval. Therefore, the data is stored in encrypted form. Encryption alone is not sufficient, as it provides security but reduces usability. Major advantage to be drawn from cloud computing is due to delegation of computation, but encrypting data would require sharing of keys with the third party performing computation on it, thereby increasing vulnerability. Hence, there is a need of feasible homomorphic schemes that allow user to compute on encrypted data, to verify a computation done by third party, to search an encrypted database, and so on. Fully Homomorphic Encryption allows a third party to evaluate arbitrary functions over encrypted data without decrypting it.

Homomorphism is a property by which a problem in one algebraic system can be converted to a problem in another algebraic system, be solved and the solution later can also be translated back effectively. First achieved by Gentry [1], FHE has been a constant area of interest to researchers due to difficulties in making Gentry's scheme practically feasible. Problem of an efficient FHE scheme is still open together with it being application-specific.

In this paper, we propose a FHE scheme with symmetric keys based on linear algebra. The emphasis is on the fact that certain applications do not require public key cryptographic primitives and burden of long and multiple keys in public key cryptosystems can be reduced by using symmetric encryption, as pointed out in [2] .

The encryption proposed in [3] is Somewhat Homomorphic Encryption Scheme is, that is, it supports a limited number of operations to be performed on cipher texts homomorphically. To make it fully homomorphic, a refresh procedure is required which uses concept of bootstrapping [1] and public keys. We propose an alternative refresh procedure without the need of bootstrapping. Thereby adapting the scheme to a symmetric key setup. This achievement can be attributed to the special key structure employed in the proposed scheme.

Organization of the paper

Section II describes briefly the chronology of research work in the fields of FHE. Section III presents the proposed symmetric FHE scheme.

II. RELATED WORK

Idea of using homomorphism along with encryption introduced by Rivest et al [4] in 1978, the first homomorphic cryptosystem. Scheme [4] has only a multiplicative property that performs multiplications with encryption of messages without losing underlying information. Domingo [5] introduced another privacy homomorphism supporting both addition and multiplication operations on encrypted data. Authors in [5] showed an application that allows recovery of exact result at a classified level (user/delegator level) from unclassified computation (untrusted worker) on perturbed data. Though, the scheme in [5] was claimed to withstand a known-plaintext attack; Cheon and Nam [6] demonstrated how to attack [5] with d+1 known plaintexts in O(d3log2n).

A breakthrough work over Fully Homomorphic Encryption scheme was given by Gentry [1, 7] which is based on algebraic lattice theory. Gentry's scheme is not yet useful for practical applications due to its high computational complexity.

Schemes that followed the Gentry's blueprint [3, 8, 9] are inefficient and impractical because of the large key size and high per-gate evaluation time which is a bottleneck in practical deployment. Coron et al [10] showed an idea of reducing the public key size of somewhat homomorphic encryption down to $\tilde{O}(\lambda 7)$. Later in [11], the key size was

further reduced to λ3Õ() by using a probabilistic decryption algorithm. In 2012, Govinda and Vijaya Kumari [12] proposed an efficient public key fully homomorphic scheme with smaller key size in comparision to other schemes [3, 10] and was capable of encrypting integer plaintexts rather than single bit. Key size of the scheme [12] was Õ(n4) and overall computational complexity was Õ(n8). Security of [12] was based on partial approximate greatest common divisor. An easy and asymptotically faster FHE scheme technique called approximate eigenvector method was presented in [13]. Approximate eigenvector method is an improvement over [9] as [9] uses expensive step involving relinearization. In [13], authors showed a recent attribute-based encryption scheme [14] compiled into an attribute-based FHE scheme that permits encrypted data to be processed homomorphically under the same index. All the above schemes use public key in their process. Xiao et al [15] showed how security of a homomorphism can be based on hardness of large integer factorization using symmetric keys. Key Size and computation time has been reduced enough for practical deployment. Gupta and Sharma [2] proposed scheme that used symmetric keys of smaller size based on matrix operation and making it suitable for many data centric applications.

III. Proposed Symmetric FHE Scheme

Public cryptographic FHE schemes involve many keys that might not be even required for single-user applications. Moreover, secret keys impart higher level of security. We propose a symmetric FHE scheme based on linear algebra. The basic idea has been derived from the SWHE in [3].

The SWHE of [3] operates on bits only; we extend the scheme to operate on integers from message space Z_N. The major hindrance in making [3] fully homomorphic is the absence of any refresh procedure which could work with symmetric keys. The only way to make it fully homomorphic is to use the bootstrappability of the scheme, i.e. first decrypt the noisy ciphertext using encrypted decryption circuit, and then encrypt using another key. Such process needs a public key. We intend to re-encrypt the ciphertext without first decrypting, thereby eliminating the use of public key.

Deriving ideas from modular mathematics, we know that any number $y = x \bmod N$ is not affected if we perform $y = x \bmod M$, where M is any multiple of N. Thus, we design a refresh key as a multiple of secret key used. The factor by which to multiply is derived from other operations and parameters involved in the scheme. The special structure of the refresh key allows a re-encryption step to serve as refresh method, thereby not requiring any additional efforts to design a public key or an evaluation key. In summary, we have a secret key and a refresh key, thereby a symmetric FHE scheme.

The security parameter in the scheme used is λ. Both the keys are derived based on the security parameter. Encryption process randomizes the plaintext by extending it from λ bits to λ^2 bits. Decryption is simply the inverse of encryption. Refresh procedure involves re-encrypting the ciphertext using another key. The scheme provides following primitives:

KeyGen(λ,N)

1. *Select s_k such that it is a composite number i.e. product of two equal length primes (pq)*
   - ** Length of p or q is $\sqrt{\lambda}$ bits*
   - ** Length of s_k is λ bits, (s_k = secret key)*
2. *Suppose $\eta = \log_N \lambda$, pick randomly integer $k < N$, η bits long*
4. *Generate $r_k = k * s_k * N$, (r_k = refresh key)*

Encrypt *(m, s_k)*

1. *Pick $m' \equiv m \bmod N$, $m \in Z_N$*
2. *Pick random number r of length λ^2 bits*
3. *Output $c = m' + s_k * r$*

Decrypt*(c, s_k)*

1. *Output $m = c \bmod s_k \bmod N$*

Refresh*(c, r_k)*

1. *output $c' = c \bmod r_k$*

The correctness of scheme is direct, since the multiplication is inverted through modulo operations. The homomorphic operations are performed using the following two mathematically complete operations:

Add*(c_1, c_2)*

Output $c = c_1 + c_2$

Multiply*(c_1, c_2)*

Output $c = c_1 + c_2$

Thus, implementing a homomorphic equivalent of any circuit is easy and direct as the operations are straightforward integer addition and multiplication.

Complexity of proposed scheme is $O(\lambda^3)$ where λ is security parameter. Major operation involved in encryption is multiplication. Theoretically, when one number of λ^2 bit is multiplied with another number of λ bits, the total time required over bit level operations is $O(\lambda^3)$. On integer operations the complexity is $O(1)$. Decryption depends on the ciphertext size. Since Ciphertext size is of $O(\lambda^3)$ the bit-level operations in decryption algorithm are $O(\lambda^3)$. For integer operations the total complexity is $O(1)$. Refresh algorithm also depends on ciphertext size, thus the total complexity is $O(\lambda^3)$.

IV. Performance and Security Analysis

Security of the scheme is derived from the difficulty of factorizing large integers. Secret key is secure since an adversary can guess s_k only through hit-and-trial. Security of scheme against hit-and-trial guessing of key is of $O\left(\frac{1}{2^{\lambda+\log_N^2 \lambda}}\right)$.

Lemma 1: Secret key s_k can be guessed correctly with a negligible probability $\approx \frac{b \ln 2}{2^{b-1}}$, where $b=\sqrt{\lambda}$ and effort required to guess s_k is $O(\frac{2^{\sqrt{\lambda}}}{\sqrt{\lambda}})$.

Proof: s_k is of λ bit. To deduce s_k, one should guess two $\sqrt{\lambda}$ bit prime numbers since s_k is the product of these two prime numbers.

Let $b=\sqrt{\lambda}$. Then, there are $\frac{2^b}{\ln 2^b}$ primes of length of maximum b bits. Thus, total number of primes of length exactly b bits are $\frac{1}{\ln 2}\frac{b-2}{b(b-1)}2^{b-1}$.

Let $\frac{1}{\ln 2}\frac{b-2}{b(b-1)}2^{b-1} = N_{pb}$

Then Probability of guessing s_k correctly, $P_{sk} = \frac{1}{N_{pb}} * \frac{1}{N_{pb}}$

$$= \frac{1}{(N_{pb})^2} = (\ln 2 \frac{b(b-1)}{b-2}\frac{1}{2^{b-1}})^2 \approx \frac{b \ln 2}{2^{b-1}} \quad (1)$$

It can be easily noted P_{sk} is negligible for any $\lambda>64$.

Effort required to guess the secret key is $1/P_{sk}$. From equation1, this effort is $O(\frac{2^{\sqrt{\lambda}}}{\sqrt{\lambda}})$, since $b=\sqrt{\lambda}$

Since Refresh key is in public domain and is derived from secret key, we should consider this point of view. Again, s_k can be obtained from r_k only through factorization. The fact that s_k is composite number makes this guessing more difficult. Size of the factors is the only leaked information.

Lemma 2: Probability of correctly deriving s_k from r_k through factorization is $\frac{1}{2^{\lambda+\log_N^2 \lambda}}$ which is negligible for $\lambda>8$, and effort required to guess r_k is $O(2^{\lambda+\log_N^2 \lambda})$.

Proof: The refresh key r_k is product of three numbers.

Bit length of first multiple, $s_k = \lambda$ bits

Bit length of Second multiple, $k = \eta$ bit

Bit length of third multiple, $N = \eta$ bit.

Probability of success, $P_{rk} = \frac{1}{2^{\lambda}} * \frac{1}{2^{\eta}} * \frac{1}{2^{\eta}} = \frac{1}{2^{\lambda+\log_N^2 \lambda}}$

(using $\eta = \log_N \lambda$).

Effort required to guess r_k is $1/P_{rk}$, that is $O(2^{\lambda+\log_N^2 \lambda})$).

Experimental Results

Proposed scheme was implemented as a Java program on 2.00GHz Intel Core 2 Duo processor running Ubuntu 12.04. Experiment was performed with plaintext space Z_2. Since the proposed scheme is similar to DGHV [3], comparison of these two is presented in Table 1. The aspects selected for comparison directly affect the performance of the scheme.

TABLE I. COMPARISON OF PROPOSED SCHEME AND SCHEME OF [3]

Scheme	Size of key	Encryption	Decryption	Message Expansion
DGHV [3]	$O(\lambda^{10})$	$O(\lambda^{10})$	$O(\lambda^{10})$	$O(\lambda^5)$
Our Scheme	$O(\lambda^3)$	$O(\lambda^3)$	$O(\lambda^3)$	$O(\lambda^3/\log \lambda)$

Table 2 lists the runtimes for all primitives as recorded in the experiment. Growth of runtime for encryption, decryption and refresh operations can be observed with increasing values of the security parameter.

Also, it should be noted that runtimes for addition and multiplication remain constant with increasing value of security parameter. This can be explained as scheme operates on integers rather than bits. These implementation results suggest that the scheme is practically feasible.

TABLE II. RUNTIME OF VARIOUS PRIMITIVES

Parameter λ→ / **Primitives↓**	**16**	**32**	**64**
Encryption	76.66 ms	344.57 ms	25.74 s
Decryption	14.52 µs	26.68 µs	41.76 µs
Refresh	6.64 µs	20.25 µs	44.21 µs
Addition	13.06 µs	13.06 µs	13.06 µs
Multiplication	15.96 µs	15.96 µs	15.96 µs

V. CONCLUSION

Currently most of the feasible FHE schemes are based on public cryptosystems, while many applications could be handled by symmetric keys. Hence, our efforts have been to propose a symmetric key Fully Homomorphic Encryption. We have demonstrated how to construct keys so that there is no need of bootstrapping to refresh the noisy ciphertexts. The experimental and theoretical results show that the proposed scheme is better than DGHV. Also, it is practically feasible to be deployed in cloud-computing applications like PIR, e-voting, etc.

Implementing the scheme in cloud applications can be further research work.

REFERENCES

[1] Gentry, Craig. A fully homomorphic encryption scheme. Diss. Stanford University, 2009.

[2] Gupta, C.P.; Sharma, I., "A fully homomorphic encryption scheme with symmetric keys with application to private data processing in clouds," Network of the Future (NOF), 2013 Fourth International Conference on the , vol., no., pp.1,4, 23-25 Oct. 2013 doi: 10.1109/NOF.2013.6724526

[3] Van Dijk, Marten, et al. "Fully homomorphic encryption over the integers." Advances in Cryptology–EUROCRYPT 2010. Springer Berlin Heidelberg, 2010. 24-43.

[4] Rivest, Ronald L., Len Adleman, and Michael L. Dertouzos. "On data banks and privacy homomorphisms." Foundations of secure computation 4.11 (1978): 169-180.

[5] Ferrer, Josep Domingo I. "A new privacy homomorphism and applications." Information Processing Letters 60.5 (1996): 277-282.

[6] Cheon, Jung Hee, and Hyun Soo Nam. "A Cryptanalysis of the Original Domingo-Ferrer's Algebraic Privacy Homomophism." IACR Cryptology ePrint Archive 2003 (2003): 221.

[7] Gentry, Craig. "Fully homomorphic encryption using ideal lattices." STOC. Vol. 9. 2009.

[8] Smart, Nigel P., and Frederik Vercauteren. "Fully homomorphic encryption with relatively small key and ciphertext sizes."Public Key Cryptography–PKC 2010. Springer Berlin Heidelberg, 2010. 420-443.

[9] Brakerski, Zvika, and Vinod Vaikuntanathan. "Fully homomorphic encryption from ring-LWE and security for key dependent messages." Advances in Cryptology–CRYPTO 2011. Springer Berlin Heidelberg, 2011. 505-524.

[10] Coron, Jean-Sébastien, et al. "Fully homomorphic encryption over the integers with shorter public keys." Advances in Cryptology–CRYPTO 2011. Springer Berlin Heidelberg, 2011. 487-504.

[11] Stehlé, Damien, and Ron Steinfeld. "Faster fully homomorphic encryption." Advances in Cryptology-ASIACRYPT 2010. Springer Berlin Heidelberg, 2010. 377-394.

[12] Ramaiah, Y. Govinda, and G. Vijaya Kumari. "Efficient public key generation for homomorphic encryption over the integers." Advances in Communication, Network, and Computing. Springer Berlin Heidelberg, 2012. 262-268.

[13] Gentry, Craig, Amit Sahai, and Brent Waters. "Homomorphic encryption from learning with errors: Conceptually-simpler, asymptotically-faster, attribute-based." Advances in Cryptology–CRYPTO 2013. Springer Berlin Heidelberg, 2013. 75-92.

[14] Gorbunov, Sergey, Vinod Vaikuntanathan, and Hoeteck Wee. "Attribute-based encryption for circuits." Proceedings of the forty-fifth annual ACM symposium on Theory of computing. ACM, 2013

[15] Xiao, Liangliang, Osbert Bastani, and I-Ling Yen. "An Efficient Homomorphic Encryption Protocol for Multi-User Systems." IACR Cryptology ePrint Archive 2012 (2012): 193.

Secure Management of Key Distribution in Cloud Scenarios

Zongmin Cui*†, Hong Zhu*, Jing Yu†
*School of computer science and technology
Huazhong university of science and technology, Wuhan, China
Email: {cuizm01,whzhuhong,yujingellemma}@gmail.com
†Jiujiang university, Jiujiang, China

***Abstract*—Existing key distribution scheme based on key derivation has security default in cloud scenarios. The scheme distributes decryption keys to users via tags stored on cloud server. If the tag is destroyed by cloud server intentionally or unintentionally, the key distribution is destroyed too. Besides the above case, if all related tags are stored on client operated by user, the storage burden is high. To remove the insecurity of key distribution, we propose a novel solution based on tag derivation. In our scheme, each user needs to manage a single key and tag. Through the two information, the user can compute all authorized keys without using any information stored on cloud server. That is, our key distribution scheme is unrelated to cloud server to enhance the security of key distribution. The experiment results show that the performance of our method is better than a kind of existing methods in key distribution and query.**

***Index Terms*—key distribution; tag derivation; key derivation; cloud computing**

I. Introduction

As cloud server is untrusted, to prevent cloud server from accessing sensitive data, data owner should encrypt sensitive data before sending it to cloud. In the same time, the decryption key should be distributed to authorized user. If the number of data authorized to user is huge, the user may need to manage a huge number of decryption keys.

To decrease the number of keys managed by user, the existing key distribution methods [1-11] removed the issue by tags [12] stored on server. Via the tag and a few number of keys, user can compute all authorized keys in multiple data owner scenarios. That is, the scheme distributes decryption keys to users by tags stored on server. However, if the tag is destroyed by server intentionally or unintentionally, the key distribution is destroyed too. In another word, the user can not get the decryption keys with destroyed tags.

To enhance the security of key distribution, we propose a secure key distribution scheme SKD. In our scheme, each user needs to manage only a single key and a single tag. Via the two information, the user can compute all authorized decryption keys without using any information stored on server. Therefore, SKD removes the relationship between key distribution and tags to enhance the security.

Our contributions are illustrated as follows.
(1) We propose a secure key distribution scheme SKD, which removes the security default in key distribution for untrusted cloud.
(2) We evaluate SKD by experiments. The experiment results show that the performance of our method is better than a kind of existing methods in key distribution and query.

This paper is organized as follows. Section 2 states our problem. Section 3 illustrates the tag computation. Section 4 provides the key distribution. Section 5 evaluates our method through experiments. Section 6 concludes the paper.

II. Problem Statement

We use K to denote all keys in the system. Tag is the key element of key distribution based on key derivation [1-11]. Thus the definition of tag is shown as follow.

Definition 1 (Tag): For each $k_i, k_j \in K$, a token $t_{i,j}$ is specified as $t_{i,j} = k_j \bigoplus h(k_i, l_j)$, where l_j is a label associated with k_j, $\bigoplus$ is the bit-a-bit xor operator and h is a cryptographic hash function. As k_j is computable from k_i by $< l_j, t_{i,j} >$, k_j should remain secret. That is, derivated keys remain secret, while tokens and labels are public on the server. The binary-tuple $< label, token >$ is defined as a tag.

From definition 1, we find that if a user gets k_i and $T(k_i, k_j) =< l_j, t_{i,j} >$, he/her can compute k_j. However, if he/her gets k_j and $< l_j, t_{i,j} >$, he/her can not compute k_i. That is, key can not be derived reversely.

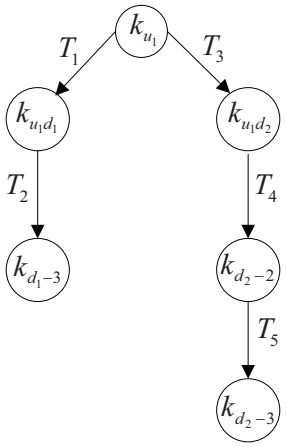


Fig. 1: An example of the insecure key distribution

To simplify the next illustration, we use T to denote all tags in the system. Meanwhile, we use the key distribution example shown in Figure 1 as an example to state our problem.

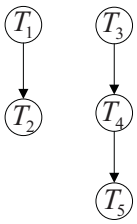


Fig. 2: Tag graph

Figure 1 shows only a single user u_1's key distribution, where each node is a key and each directed edge is a tag.

Figure 1 is generated according to an initial authorization list. Through the key distribution, each user needs to manage only a single key (access key [9]). For example, user u_1 needs to manage an access key k_{u_1}. Via k_{u_1} and tags stored on server, u_1 can compute all authorized decryption keys from different data owners. Meanwhile, u_1 can not compute the unauthorized decryption keys.

To simplify the next illustration, we define a single identifier for each tag: $\forall T_i \in T$, $i \in N^+$. If a tag is destroyed by server, the related keys can not be computed by user. With reference with Figure 1, if tag T_1 (from k_{u_1} to $k_{u_1 d_1}$) is destroyed, user u_1 can not compute $k_{u_1 d_1}$ by the access key k_{u_1}. In addition, he/her also can not compute k_{d_1-3} without $k_{u_1 d_1}$. That is, T_2 is useless with losing T_1. Therefore, the insecurity of key distribution is the problem existing in [1-11].

From the above illustration, we find that the security of key distribution is depended on the security of tags. Different tags have different use sequences. Ahead of derivation path is used firstly and bottom of derivation path is used last. For example, T_1 is used before T_2. Thus, we show the tag graph in Figure 2 corresponding to Figure 1.

Figure 2 shows the use sequences of tags, in which the first used tag is derived to latter used tag. In [1-11], the security of key distribution is the security of tag computation.

III. Tag Computation

To securely compute the tags, we propose a novel solution to remove this issue. In our method, each user needs only manage a tag and a single key, by which the user can compute all the authorized tags without using any other information. Thus our method does not need to store any information on the server, which lets the tags not to be destroyed by the server.

To facilitate the explanation of our algorithm, the necessary definitions are shown as follows.

Definition 2 (Leaf node): In the tag graph, leaf node is the node which only has in-degree and has not any out-degree.

Definition 3 (Parent node and child node): $\forall T_i, T_j \in T$, if T_j is directly computable from T_i, then T_i is the parent node of T_j and T_j is the child node of T_i.

Take the tag graph shown in Figure 2 as an example, the leaf nodes is $\{T_2, T_5\}$ and T_1 is the parent node of T_2.

Before we present the key computation algorithm, a sub-algorithm enforced by the data owner has to be presented firstly. Algorithm 1 Tag-computation takes a set of leaf nodes N_l as input and a set of corresponding parent nodes N_p (added the tags, by which the user can compute the child nodes) as out.

The core idea of Algorithm 1 is to reversely compute tag, which minimizes the number of tags in the system.

The algorithm firstly (Steps 1-6) generates the tag $T(n_p, n_l)$ to let the user can compute child node n_l by parent node n_p. And secondly (steps 8-13) it adds the corresponding tag $T(n_p, n_l)$ to the parent node n_p.

Algorithm 1 Tag-computation

Input: N_l
Output: N_p
1: **for all** $n_l \in N_l$ **do**
2: store n_l's parent nodes into N_p
3: **for all** $n_p \in N_p$ **do**
4: generate the tag $T(n_p, n_l)$
5: **end for**
6: **end for**
7: store N_l's parent nodes into N_p
8: **for all** $n_p \in N_p$ **do**
9: store n_p's child nodes into N_l
10: **for all** $n_l \in N_l$ **do**
11: $n_p = n_p \cup T(n_p, n_l)$
12: **end for**
13: **end for**
14: **return** N_p

With reference to Figure 2, $N_l = \{T_2, T_5\}$, and the generated tags are shown in Figure 3. First, $n_l = T_2$, thus we generates the tag $T_6 = T(T_1, T_2)$ and $T_7 = T(T_4, T_5)$. Second, $T_1' = T_1 + T_6$ and $T_4' = T_4 + T_7$.

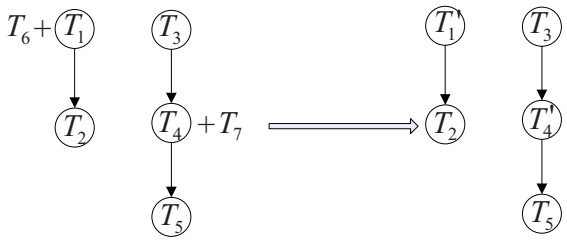


Fig. 3: An example of tag computation.

IV. Key Distribution

Based on the above sub-algorithm, Algorithm 2 Key-distribution regulates the key distribution progress, which takes a tag graph C and a user u as input and a single tag T_u as output. The algorithm is enforced by the data owner, which let the user need only to manage a single tag T_u and key k_u, by which the user can compute all authorized decryption keys without any other information.

The core idea of Algorithm 2 is to minimize the number of tags managed by the user via reversely computing tags.

The algorithm firstly calls the sub-algorithm Tag-computation(N_l) to get the new parent nodes N_p(Step 6). If the new parent node n_p has not a parent node, it means that all tags have been computed in this derivation path. Else, we takes n_p as a new leaf node stored in N_l for the next calling (Steps 8-14). Finally, we generates the tag $T_u = T(k_u, K^s)$ to connect access key with tags, by which the decryption keys have been securely distributed to the user.

Algorithm 2 Key-distribution

Input: C, u
Output: T_u
1: $K^s = \phi$, $N = \phi$, $N_l = \phi$
2: store all nodes of C into N
3: store all leaf nodes of C into N_l
4: **while** $N \neq \phi$ **do**
5: remove the nodes from N to make $N \cap (N_l \cup K^s) = \phi$
6: N_p=Tag-computation(N_l)
7: $N_l = \phi$
8: **for all** $n_p \in N_p$ **do**
9: **if** n_p has not a parent node **then**
10: $K^s = K^s \cup n_p$
11: **else**
12: $N_l = N_l \cup n_p$
13: **end if**
14: **end for**
15: **end while**
16: $T_u = T(k_u, K^s)$
17: **return** T_u

We take the tag graph shown in Figure 2 as an example to illustrate the key distribution progress, which is shown in Figure 4.

First, $N_l = \{T_2, T_5\}$, we call Tag-computation(N_l), then $N_p = \{T_1', T_4'\}$, where $T_1' = T_1 \cup T_6$ and $T_4' = T_4 \cup T_7$. As T_1' has not a parent node but T_4' has a parent node T_3, $K^s = \{T_1'\}$ and $N_l = \{T_4'\}$.

Second, $K^s = \{T_1', T_3'\}$, where $T_3' = T_3 \cup T_8$.

Finally, $T_{u_1} = T(k_{u_1}, K^s)$.

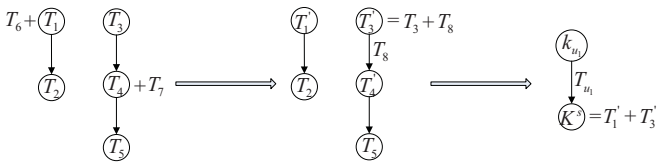


Fig. 4: Secure key distribution

Via k_{u_1} and T_{u_1}, user u_1 can compute $K^s = \{T_1', T_3'\}$. Then via T_1', the user can compute T_2. By the same way, the user can compute T_4 and T_5. In summary, via a single tag and key, the user can compute all authorized tags, by which the user can compute all authorized decryption keys.

V. Experiment Evaluation

To verify feasibility and practicality of SKD, this section will compare the performance of SKD and EBK[9] from the aspects of key distribution and query.

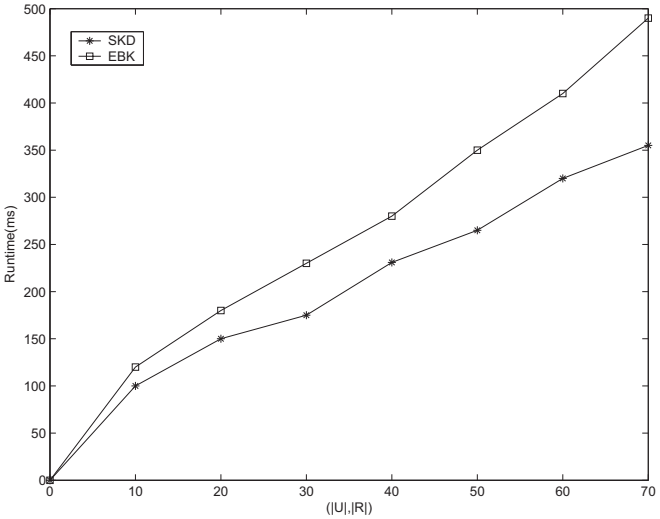


Fig. 5: The comparison of key distribution

A. Experiment Setup

The experiment environment is composed mainly of two machines: one for the server and the other for the visitors (data owners and users). The client and server machines are different. The server runs Windows Server 2007 and is equipped with Pentium Dual-Core 3.73-GHz CPU, 8-GB memory and one Hitachi HTS541616J9SA00 1-TB harddisk. The disks are formatted with 4-KB blocks, the default in New Technology File System. The client machine runs Windows 7 and has an Intel Core Duo 2.13-GHz CPU, 3-GB memory and one ST3160815AS 320-GB hard disk. To study the authentication mechanisms, we model the wide area network connection between the visitors and server as a queue with a capacity of 100 Mbps, corresponding to the 3.5-G data rate.

The size of the database (used to manage data stored on the server) is about 4 MB. The database has three tables: authorize (used to store the authorization-origin table), tag (used to store the set of tags), and datascale (used to store the experimental number of users $|U|$ and resources $|R|$). In our experiments, each resource is a file. The digest is the size of a hash function. The encryption method on the resource is Advanced Encryption Standard [13]. The programming tools are java1.6 and Eclipse_6.0. The networks and key distribution are actual implementations in our system, and the query operations are simulated components.

B. Key Distribution

Our key distribution experiments are to calculate the average runtime of distributing a single decryption key to a user. By the decryption key, the user can decrypt an authorized data.

Figure 5 shows the experimental results, where the X-axis represents the user number which is equal to data number. The Y-axis of Figure 5 represents the running time of single key distribution, where the unit is millisecond (ms).

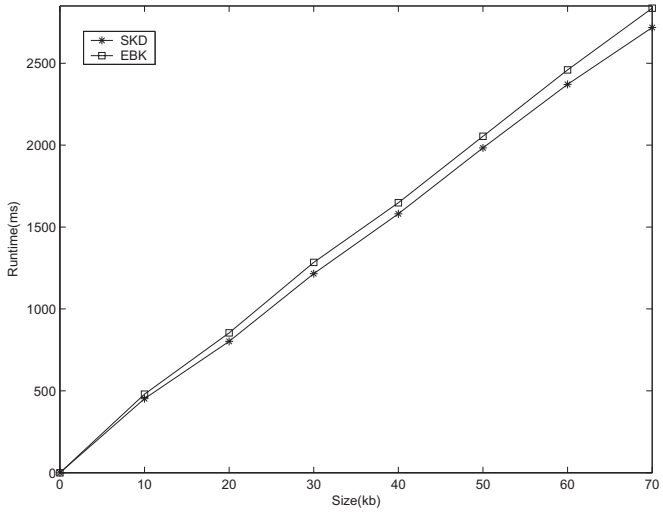


Fig. 6: The comparison of query

In our SKD method, user does not need to upload tags from cloud server. Thus, there is no communication burden in SKD. However, EBK [9] has the communication burden. Therefore from Figure 5, we can find that SKD is more efficient than EBK in key distribution.

C. Query

Based on key distribution, our query experiments are to calculate the runtime of querying a single data, which includes data decryption. Figure 6 shows the experimental results, where the X-axis represents the data size whose unit is kb. The Y-axis of Figure 6 represents the running time of single query, where the unit is millisecond (ms). In Figure 6, the numbers of users and data are 100 too, i.e. $|U| = |R| = 100$.

By the same reason, SKD decreases the communication burden than EBK in query. Therefore from Figure 6, we can find that SKD is efficient than EBK in query performance.

In summary, our method is efficient and secure than the existing methods for key distribution in cloud scenarios.

VI. Conclusion

To remove the insecurity default in key distribution for cloud data, we propose a secure key distribution scheme. The scheme is based on reversely computing tags, which removes the relationship between key distribution and tags to enhance the security. Experiment results show that our method is efficient and secure in cloud scenarios.

VII. Acknowledgement

This work is supported by the Jiangxi Natural Science Foundation: "Efficient key management for publish/subscribe system in cloud scenarios".

References

[1] E. Damiani, S. D. C. Di Vimercati, S. Foresti, et al. Key management for multi-user encrypted databases. In Proceedings of the ACM Workshop On Storage Security And Survivability, 2005; 74 83.

[2] E. Damiani, S. D. C. Di Vimercati, S. Foresti, et al. Metadata management in outsourced encrypted databases. In Proceedings of the Secure Data Management, 2005; 16 32.

[3] E. Damiani, S. D. C. Di Vimercati, S. Foresti, et al. Selective data encryption in outsourced dynamic environments. Electronic Notes in Theoretical Computer Science, 2007, 168: 127 142.

[4] S. D. C. Di Vimercati, S. Foresti, S. Jajodia, et al. Over-encryption: management of access control evolution on outsourced data. In Proceedings of the International Conference on Very Large Data Bases, 2007; 123 134.

[5] S. Liu, W. Li, L. Wang, Towards Efficient Over-Encryption in Outsourced Databases Using Secret Sharing. In Proceedings of the International Conference on New Technologies, Mobility and Security, 2008; 1 5.

[6] C. Blundo, S. Cimato, S. D. C. Di Vimercati, et al. Efficient key management for enforcing access control in outsourced scenarios. In Proceedings of the International Information Security Conference, 2009; 364 375.

[7] A. Singh, M. Srivatsa, L.Liu, Search-as-a-service: Outsourced search over outsourced storage. ACM Transactions on the Web, 2009, 3 (4): 1 13.

[8] S. D. C. Di Vimercati, S. Foresti, S. Jajodia, et al. Encryption policies for regulating access to outsourced data. ACM Transactions on Database Systems, 2010, 35 (2): 1 56.

[9] Z. Cui, H. Zhu, and L. Chi. Lightweight key management on sensitive data in the cloud. Security and communication networks, 2013, 6(10): 1290-1299.

[10] Z. Cui, H. Zhu, J. Shi, L. Chi, K. Yan. Lightweight Management of Authorization Update on Cloud Data. Crowd and Cloud Computing Workshop in conjunction with ICPADS, 2013; 456-461.

[11] Y. Chen, C. Chu, W. Tzeng, et al. CloudHKA: A Cryptographic Approach for Hierarchical Access Control in Cloud Computing. In Proceedings of the International Conference on Applied Cryptography and Network Security, 2013; 37 52.

[12] M. J. Atallah, K. B. Frikken, M. Blanton, Dynamic and efficient key management for access hierarchies. In Proceedings of the ACM Conference on Computer and Communications Security, 2005; 190 202.

[13] A. T. Hoang, T. Fujino, Intra-masking dual-rail memory on LUT implementation for tamper-resistant AES on FPGA. In Proceedings of the ACM/SIGDA international symposium on Field Programmable Gate Arrays, 2012; 1-10.

An Efficient and Robust One-TimeMessage Authentication CodeScheme using Feature Extraction of Iris in Cloud Computing

ZaidAmeen Abduljabbar[1,2], Hai Jin[1], Deqing Zou[1], Ali A.Yassin[2], ZaidAlaa Hussien[1,3],Mohammed Abdulridha Hussain[1,2]

[1]Cluster and Grid Computing Lab
Services Computing Technology and System Lab
School of Computer Science and Technology
Huazhong University of Science and Technology, Wuhan, 430074, China
Email: zaidalsulami@yahoo.com, hjin@hust.edu.cn
[2]University of Basrah, Basrah, Iraq
[3]Southern Technical University, Basrah, Iraq

***Abstract*—Cloud computing suffers from a number of problems in terms of security issues. Authentication and integrity play an important role in the security field and numerous concerns have been raised to recognizeor protect any tampering with exchanges of text between two entities (sender and receiver) within the cloud environment. Many schemes in this area can be vulnerable to well known methods of attack such as replay attack, forgery attack, dictionary, insider, and modification attacks. A robust scheme is therefore required to detect or prevent any modification or manipulation of a message during transmission. In this paper, we propose a new *message authentication code* (MAC) based on feature extraction of the user's iris in order to assure the integrity of the user's message. Features are extracted from the user's iris to generate a message code for each user's login and to prohibit malicious attacks such as replay, forgery and insider attacks. Our proposed scheme enjoysseveral important security attributes such as a user's one time bio-key, robust message anonymity, data integrity for a user's message, phase key agreement, bio-key management,and one time message code for each user's session. Finally, our security analysis and experimental results demonstrate and prove the invulnerability and efficiency of our proposed scheme.**

Keywords- Cloud Computing; Iris; Features Extraction; One Time Bio-key; One Tim Message Authentication Code; MAC.

I. INTRODUCTION

In recent yearsa huge volume of different types of data has been transferred over the Internet as a result of the rapidgrowth ofmodern information digitalization techniques such as cloud computing [1].Text is one of the most significant and most widely used mediums for transmitting data, along with image, audio, and video.Cloud computing is generally regarded as the next generation's computing infrastructureand asan effective way of enabling users to utilizelarge volume of resources and to provide anefficient and readily available on-demand service[2]. Its successful deployment depends on the existence of strong securitysafety techniques. Due to theessential need for message protectionwhen two parties are transmittingwithin the cloud environment, efficient and robust automatic methods are required to identify and validate the contents of text messages. In the others words, the protection of messagesagainstmalicious attacks such as replay, forgery and insider attacks is one of the most important security issuesin fields such as cloud computing and green computing.However, the issues of*message authentication code* (MAC) and integrity have been addressed as urgent matters and many achievements have been presented by researchers in recent years [3-7]. The advantage of the cryptography one-way hash function is that it is faster for authors to authenticatethan a digital signature [4]. Unfortunately,the major drawback of MAC is the fact it does not appear to be capable of ensuring the high level of security required when it is used alone as pure MAC.

In this paper, we propose an efficient and secure scheme for protection of text from being manipulated or tampered during transmission between users in the cloud environment. The algorithm integrates a crypto-hash function (SHA-1) [8, 10] with a biometric technique which involves the use of the robust features extracted by using 2-D Gabor filter [11-13] after an intersection between the sender's iris and the receiver's iris. These are usedtogether toprotect the user's message from being modified. Our proposed scheme is a well-organizedprocedure with respect to variousqueries and requires regular verification to decrease the audit costs per verification phase

The main contributions of our scheme tothe cloud environment in general, andtomessage authentication and integrity in particularare:(1) Our proposed scheme addresses all the previous weaknesses, creating a new robust message authentication scheme which uses the robust features extraction from shared biometric iris information and cryptography as a one-way hash function toprotect a message without losing integrity and authentication. (2)Both service providers and users can achieve authenticated phase keys. (3) It is computationallyefficient as well as provides simple integration with the available infrastructure, and it is easyfor deployment

and management. (4) Our scheme is very effective against many attacks such as replay attacks, insider attacks and reflection attacks. (5) The main idea behind our efficient schemeisto find the bestchoice of parameter value to reduce the computational cost of cloud audit services.

This paper is organizedas follows. In section II we describe the proposed scheme both interms of configuration phases and verification phases. Section III contains a security analysis with respect to the well-knownattacks. In section IVthe implementation and performance are described, while in section V wecompare the most significant and widespread text authentication solutions with our scheme.Finally, section VI concludes the paper.

II. Our Proposed Scheme

Our proposed scheme is composed of two phases, theconfiguration phase and verification phase.The configurationphase is performed only once. A bio-shared image and shared key are received by both thesender and the receiver.The verificationphase will be invoked every time a user wants to send anauthenticated message to another user. In the configuration phase, the main components (Cloud Service Provider, Sender, Receiver) also use RSA, a cryptographic hash function$h(.)$(SHA-1) and asymmetric key encryption/decryption$Enc(.)/Dec(.)$.It is important to emphasizethat they only need to run RSA for secure data transmissionamong*CSP*, *S*,and *R*over an insecure channel [8,9]. Therefore, such an operation is necessary only for the configuration phase and not for the later ones. Therefore, the*CSP* is not needed in the run time.The configuration phase performs the followingsteps:

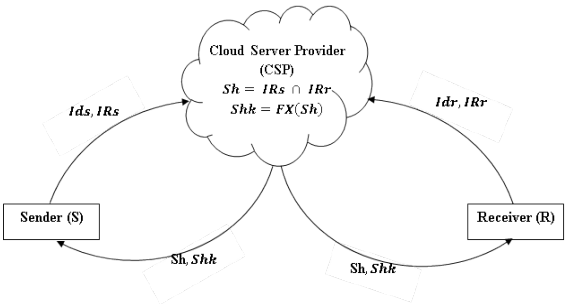


Fig. 1. Configuration phase

The RSA is run by*CSP*, *S*, and *R* in order to generate a public key and private key which will be used to secure irises transmissionfrom sender and receiver to*CSP*. Then, the*CSP* sendsthe public keyPU_{CSP} to both sender (*S*) and receiver (*R*) for encrypting their irises (*IRs*, *IRr*) andreturns them to the*CSP*.

Upon receivingthe encrypted(*IRs*, *IRr*), the *CSP*decrypts the received irises by using its private keyPR_{CSP}, saves (*IRs*, *IRr*), generatesa bio-shared image by intersection ($Sh = IRs \cap IRr$) and computesa shared key$Shk = FX(Sh)$as shown in Fig. 1, where *FX*refers to a function to extract features, it employs2-D Gabor filter to extract features from the normalized iris data. Having done this,*CSP*encrypts (*Sh*, *Shk*) by using (PU_S, PU_R) and transmits them both to the sender and receiver respectively. Finally, both the sender and receiver decrypt the received (*Sh*, *Shk*) by using their private key (PR_S, PR_R). After the configuration phase, the sender/receiver can use his or her bio-shared image to extract features, and then generate a one time, anonymous key and a bio-key for completing theverification phase. The verification phase is described as follows, as shown in Fig.2.

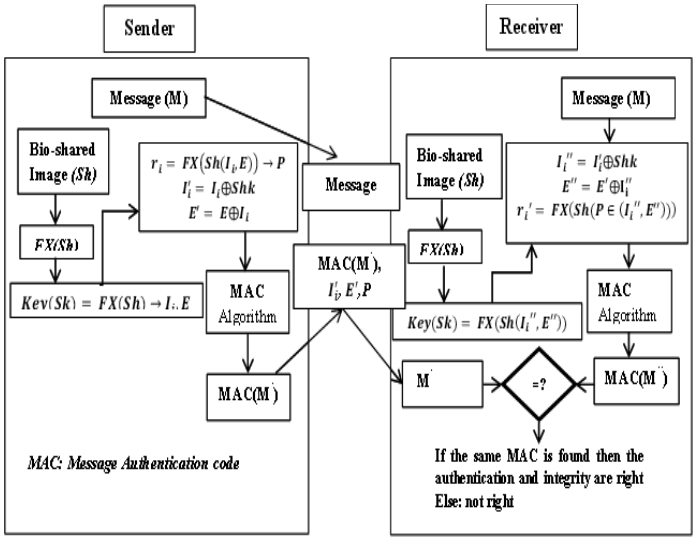


Fig. 2. Our proposed scheme

1. S→R: M, M', I_i', E', P.S performs the following steps:
- Assume sender's message is *M*.
- Generate one time salt-key$Sk=FX(Sh)\rightarrow I_i, E$:where *FX*represents a function to compute feature extraction,I_i and *E*are the start point and end point of the extracted features. BothI_i and *E* are selected randomly once, as shown in Fig.3. The*E* parameter must not exceed the length of the feature vector, which is 1024.
- Generate random number$r_i \epsilon FX(Sh)=FX(Sh(I_i, E))\rightarrow P$and compute a one-time anonymous message code,if the sender resends the same message to the receiver or vice versa.$M'= h(M||Sk||r_i)$.
- Compute $I_i' = I_i \oplus Shk$
- Compute $E' = E \oplus I_i$
- SendM, M', I_i', E', and*P*to*R*.

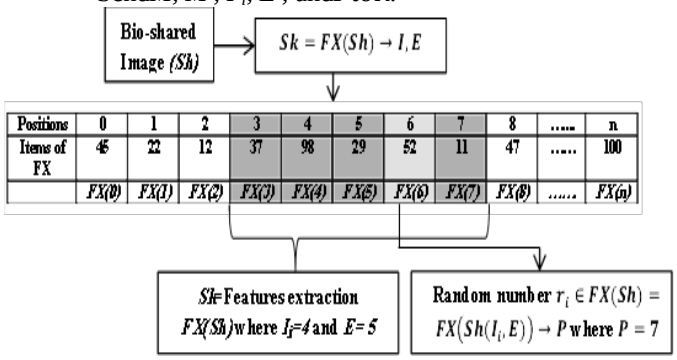


Fig. 3. The positions of *FX* vector's items, Salt-Key (*Sk*) and r_i generation extraction

2. *R*checks the integrity of receiver's message as follows:
- Compute $I_i'' = I_i' \oplus Shk$

- Compute $E'' = E' \oplus I_i''$
- Regenerate $Sk' = FX(Sh(I_i'', E))$ depending on the features extracted position (I_i'') and the end point of extracted features (E''). Extract random number $r'_i = FX(Sh(P \epsilon (I_i'', E'')))$. Then, *R* computes $M'' = h(M' || Sk' || r_i')$ if it matches M'' with M', the Receiver ensures the integrity of the message submitted by the sender. Otherwise, the verification phase terminates.

III. Security Analysis

Here, we argue that the proposed scheme can also withstand several threats to security such as replay attack and insider attack. Our proposed scheme has a number of merits and contains a one-time bio-key, a one-time anonymous message code and key agreement.

Theorem 1. Our proposed scheme can provide robust user messageanonymity.

Proof. Assuming a sender/receiver attempts to resend the same message which has been sent previously, if an adversary tries to eavesdrop on the sender's login request (M, M', I'_i, E', P), he cannot use the same sender's message authentication code ($M' = h(M||Sk||r_i)$) because the sender generates once for each sender's request (r_i and Sk). So, r_i and Sk have been extracted from the intersection of receiver's iris and the sender's iris $r_i \epsilon FX(Sh) = FX(Sh(I_i, E)) \rightarrow P$; $Sk = FX(Sh) \rightarrow I_i, E$; $Sh = IRs \cap IRr.FX$ is a function required to compute feature extraction, I_i and E are the start and end points of extracted features. Both I_i and E are selected randomly once, as shown in Fig. 3. Additionally, an adversary does not have the main keys (Sh, I'_i, E', P) to compute the crypto hash function M'. Hence, it is much harder for an adversary to disclose the sender's message authentication code. Clearly, our proposed scheme can support users' message anonymity.

Theorem 2. Our proposed scheme can providebiometric message authentication code.

Proof. The biometric operator can identifya person by means of particular physiological features such as iris recognition. Iris is the most effective form of security used in biometric topicsand can overcome well-known attacks. In the configuration phase, the sender (S) and receiver (R) send their irises (IRs, IRr) to the *CSP* through a secure channel. Then the *CSP* saves (IR_s, IR_r), generates a bio-shared image ($Sh = IRs \cap IRr$) and sends Sh to sender and receiver. During the verification phase, when the sender/receiver wishes to send message from one to other, a biometric-message authentication code ($M' = h(M||Sk||r_i)$) must be generated, based on salt-key $Sk = FX(Sh) \rightarrow I_i, E$ and a random number $r_i \epsilon FX(Sh) = FX(Sh(I_i, E)) \rightarrow P$. Clearly, our proposed scheme can support biometric message authentication codes.

Theorem 3. Our proposed scheme can providebiometric-keymanagement.

Proof. In our proposed scheme, when the sender sends a message (M) to the receiver or vice versa, a secret salt-key $Sk = FX(Sh) \rightarrow I_i, E$ is used to compute ($M' = h(M||Sk||r_i)$). Additionally, the mechanism of computing Sk is based on I_i and E, where I_i is the start point of the extracted features and E is the end point of extracted features. Both I_i and E are selected one time, randomly. As a result, an attacker cannot access the session keys, so is still unable to obtain the main operators (Sk, Sh) that generated at configuration phase by *CSP* and that generated (I_i, E, P) at verification phase by sender. Therefore, our work supports biometric-key management.

Theorem 4. Our scheme can prevent a replay attack.

Proof. An attacker performs a replay attack by eavesdropping the login message sent by a rightful sender to the receiver. While the interchange is over between sender and receiver, an attacker reuses this message to impersonate the valid user when he logsoff the system. In our proposed scheme, each new sender's longing request should be identical with *CSP*'s keys Sh, Shk, IR_s, IR_r. Therefore, an adversary cannot pass any replayed message to the receiver's verification. As a result, an adversary fails to apply this type of attack and our proposed scheme is much harder to replay attack.

Theorem 5. Our scheme can prevent a forgery attack or a parallel-session attack.

Proof. If any adversary is attempting impersonation, a valid session message M, M', I'_i, E', P can be accessed by using secret parameters Sh, Sk, Shk, r_i, I_i, E, P. An adversary does not have any information about Sh, Shk, IR_s, IR_r to compute M', I'_i, E', P. Lastly, an adversary will fail to forge a valid session message and therefore, cannot use a forgery attack. Our proposed scheme can thus prevent forgery attack.

IV. Implementation and Results

To evaluate the efficiency and accuracy of our proposed scheme, we have executed several experiments. Fig. 4 shows the time processing of the verification phase. The average time for the verification phase of our scheme is equal to 0.0045 seconds for each user who denotes the excelling solution of our proposed. This average time has been obtained from 200 runs of our proposed scheme, with each run consisting of 10,000 users.

Furthermore, with regard to system efficiency, we study the accuracy of our work. Fig.5 shows that we get 100% accurate results from 10,000 users in our experiment. For greater visibility, we use 5,000 users in Fig. 4 and 5.

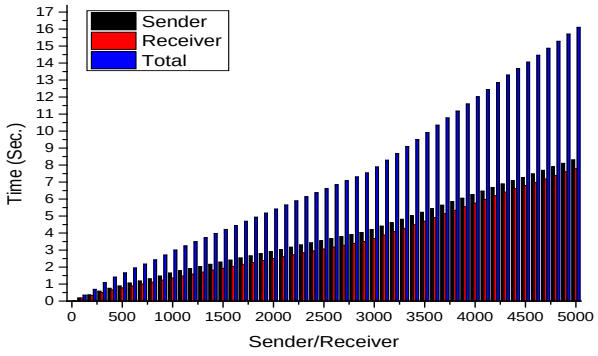


Fig. 4. Performance of our proposed scheme

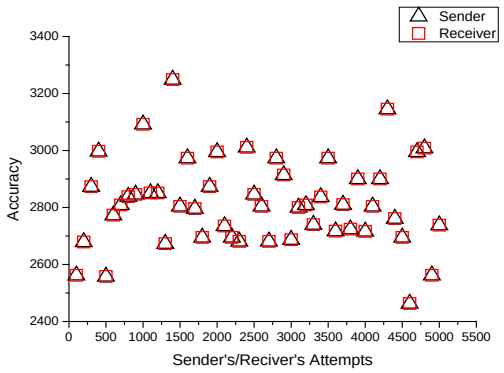


Fig. 5. Accuracy result of our proposed scheme

V. Related Works

We compare security properties of our proposed scheme with other five authentication schemes in Table I.

TABLE I. ComparisonofAuthentication Schemes

Feature	Our Scheme	Z. Liu et. al. [3]	N. Rabadi and S. Mahmud [4]	Z. Zhao et. al.[5]	A. Castiglione et. al. [6]	H. Al-Assam et. al. [7]
C1	Yes	Yes	No	No	Yes	Yes
C2	Yes	Yes	No	No	No	Yes
C3	Yes	Yes	Yes	No	No	No
C4	Yes	Yes	Yes	Yes	Yes	Yes
C5	Yes	No	No	No	No	Yes
C6	Yes	Yes	No	No	No	No

C1: One time key; C2: Bio-key; C3: one-time message anonymity; C4: Session key agreement; C5: Biometrics key management; C6: Cloud environment

VI. Conclusion

Our paper presents a new and efficient message authentication code between users in the cloud computing environment. The excelling method is emerged from the iris-biometrics features extraction to generate symmetric bio-key. The aim behind this scheme is to provide more roles and prevent known attacks.,The substantial aspects and advantages are that, first, an adversary may fail to get the keys because this depends on iris feature extraction. Second, an adversary may not get the bio-shared image because it depends on the intersection of sender and receivers' irises. Third, it provides a one-time bio-key that leads to one-time message anonymity. Fourth, valid users can freely submit a message. Fifth, it provides biometrics key management. Additionally, the proposed scheme has the ability to resist replay attacks and forgery attacks, as shown in the security analysis section. Finally, the performance of our presented scheme has been evidenced to achieve robust security with minimal time processing and the cost comparedwith predecessors' schemes.We can conclude that the integration between shared iris biometric features of two endpoints and the cryptography one-way hash function is secure enough to prevent the message from being modified by transferring between users.

References

[1] T. Rethika, I. Prathap, R. Anitha, and S.V. Raghavan, "A novel approach to watermark text documents based on Eigen values,"*in Proceedings of the Ninth International Conference on Network and Service Security (N2S'09)*,Paris, France, IEEE, pp. 1-5, June 2009.

[2] H. T. Dinh, C. Lee, D.Niyato, and P. Wang, "A survey of mobile cloud computing: architecture, applications, and approaches,"*Wireless Communications and Mobile Computing*, John Wiley, vol.13, no.19, pp.1587-1611, Dec. 2012.

[3] Z.Liu ,H.S.Lallie, L. Liu, Y. Zhan, and K. Wu, "A hash-based secure interface on plain connection,"*inProceedings ofthe sixth International Conference on Communications and Networking in China (ChinaCom'11)*, Harbin, China, IEEE, pp.1236-1239, Aug. 2011.

[4] N. Rabadi and S. Mahmud, "Drivers' anonymity with a short message length for vehicle-to-vehicle communications network,"*in Proceedingsof the fifth IEEE Consumer Communications and Networking Conference (CCNC'08)*, Las Vegas, NV, USA, IEEE, pp. 132–133, Jan. 2008.

[5] Z. Zhao, Y.F. Liu, H. Li, and Y.X. Yang, "An efficient user-to-user authentication scheme in peer-to-peer system,"*in Proceedingsof the First International Conference on Intelligent Networks and Intelligent Systems (ICINIS'08)*, Wuhan, China, pp. 263-266, Nov. 2008.

[6] A. Castiglione, D. Santis, andF. Palmieri, "An efficient and transparent one-time authentication protocol with non-interactive key scheduling and update,"*in Proceedingsofthe 28th International Conference on Advanced Information Networking and Applications (AINA'14)*, Victoria, BC, Canada, IEEE, pp.351-358, May 2014.

[7] H. Al-Assam,R.Rashid,and S. Jassim, "Combining steganography and biometric cryptosystems for secure mutual authentication and key exchange,"*in Proceedingsofthe8thInternational Conference for Internet Technology and Secured Transactions (ICITST'13)*, London, UK, pp.369-374, Dec. 2013.

[8] W. Stallings, *Cryptography and Network Security: Principles and Practice*, Prentice Hall, 6th Edition, 2013.

[9] C.Paar, J.Pelzl, and B.Preneel, *Understanding Cryptography*, Springer, 1st Edition, 2010.

[10] H. Handschuh, *Encyclopedia of Cryptography and Security*,Springer, 2nd Edition, 2011.

[11] S. Hariprasath, and V. Mohan, "Biometric personal identification based on iris recognition using complex wavelet transformations,"*in Proceedings ofthe International Conference on Computing, Communication and Networking (ICCCN'08)*,VI, USA, pp. 1-5, Dec.2008.

[12] J. G. Daugman, "High confidence visual recognition of persons by a test of statistical independence,"*IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no.11, pp. 1148-1161, Nov. 1993.

[13] L. Yu, D. Zhang, and K. Wang, "The relative distance of key point based iris recognition", *Pattern Recognition Pattern Recognition*, vol. 40, no. 2, pp.423-430, Feb. 2007.

Publicly Verifiable Delegation of Set Intersection

Tingting Wang
Soochow University
Suzhou, China
20124227042@suda.edu.cn

Yanqin Zhu*
Soochow University
Suzhou, China
yqzhu@suda.edu.cn

Xizhao Luo
Soochow University
Suzhou, China
xzluo@suda.edu.cn

***Abstract*—We study the verifiable delegation of set intersection in the model of authenticated data structures, where a weak client in storage and computation stores its dataset composed of some sets into a powerful server. The client can issue update and set intersection query over outsourced data. The server returns the result and a proof of its correctness. Our scheme allows any entity to publicly verify the correctness of set intersection query, and requires no secret key. Based on characteristic polynomials of sets, our authenticated data structure firstly provides a practical and specific method for set intersection through polynomial evaluations. It achieves optimal verification and proof complexity, as well as optimal update complexity without bearing any extra space overhead. The overhead for computing actual answer and proofs is reduced significantly in contrast with previous methods based on bilinear-map accumulators. The intersection query comes up in keyword search, database queries and other applications. Indeed, our scheme can be effectively used in such scenarios. The security of our constructions is based on co-Computational Diffie-Hellman assumption.**

Keywords—Verifiable computation; Authenticated data structures; Set intersection; Data outsourcing

I. INTRODUCTION

The rise of Cloud Computing raises the outsourcing of storage and computation to the cloud for both enterprises and individuals. The typical setting is that a client with bounded computational and storage capabilities wishes to outsource his database and issue queries that are answered by powerful servers. In such settings, verifying the correctness of computations performed by servers becomes a crucial property for the trustworthiness of cloud services. Thus the server must provide the result of the computation together with a proof of its correctness. Crucially, the verification of such correctness proof must incur minimal overhead to the client, otherwise the benefits of computation outsourcing are dismissed. A question is whether the verification can be made public: i.e. can any third part verify it? This is important, for example, in contexts where the computation has to be checked by several clients who cannot necessarily share a secret key, or the proof of correctness must be transferable(Such as, a digital signature on a message).

In [1,2,3] works on outsourced verifiable computation, they achieve verification of general functionalities. Although they cover set intersection as a special case, and meet our goal with respect to optimal verifiability, they inadequate to meet our goals with respect to public verifiability and dynamic updates, both properties are important in data querying. To address the problem of inefficiency, the recent work [4] considers a practical secure database delegation scheme supporting restricted class of queries. They consider functions expressed by arithmetic circuits of degree up to 2. Their construction is based on homomorphic MAC's and their protocol appears practical, however their solution is only privately verifiable and it does not support deletions from the dataset. The work in [5,6] proposed a scheme for publicly verifiable secure delegation of set intersection. It does not involve translating the problem to an arithmetic or boolean circuit. It's based on the notion of a bilinear accumulator[7]. They hash the accumulation values of sets in the dataset using an accumulation tree, introduced in [8]. They make use of the algebraic structure of the accumulators to gain in efficiency.

In this paper, we also propose a scheme for publicly verifiable secure delegation of set intersection. However, it's based on characteristic polynomials rather than bilinear accumulators. Characteristic polynomials for set representation have been used before in cryptography literature. We can get the result of set intersection through polynomial evaluation. Several different metrics of efficiency have been considered. Firstly, we would like that the time it takes for the client to verify a proof is short, and the time is independent of the size of server's computation cost. Secondly, we would like the server's computational overhead in computing proofs to be minimal. Additional efficiency considerations include the proof size, and the efficiency of update queries. The cost of communication should also be minimized. The main advantage of our scheme over previous approach is that we simplify the work at server without increasing the overhead of client. In addition, compared with [5,6], we provide a specific method for set intersection, rather than simple verification method. This model captures a variety of real-world applications such as outsourced SQL queries, authenticated keyword search with elaborate queries, similarity measurement and outsourced file systems, hence a practical protocol would be of great importance.

II. PRELIMINARIES

In the following, we denote with λ the security parameter and with ϵ a negligible function of λ.

Bilinear pairings. Let $\mathbb{G}_1, \mathbb{G}_2$ be cyclic multiplicative groups of prime order p, generated by g_1, g_2. Let also $\mathbb{G}_T$ be a cyclic multiplicative group with the same order p and $e: \mathbb{G}_1 \times \mathbb{G}_2 \to \mathbb{G}_T$ be a bilinear pairing with the following properties:(1) Bilinearity: $e(u^a, v^b) = e(u,v)^{ab}$ for all $u \in \mathbb{G}_1, v \in \mathbb{G}_2$ and $a,b \in \mathbb{Z}p$;(2) Non-degeneracy: $e(g_1, g_2) \neq 1$; (3) Computability: There is an efficient algorithm to compute $e(u,v)$ for all $u \in \mathbb{G}_1, v \in \mathbb{G}_2$. We denote with $(p, \mathbb{G}_1, \mathbb{G}_2, \mathbb{G}_T, e, g_1, g_2)$ the bilinear pairings parameters, output by a PPT algorithm on input 1^λ.

Assumption 1(co-Computational Diffie-Hellman). Let $\mathbb{G}_1, \mathbb{G}_2, \mathbb{G}_T$ be groups of prime order p, so that $e: \mathbb{G}_1 \times \mathbb{G}_2 \to \mathbb{G}_T$ is a bilinear map. Let $g_1 \in \mathbb{G}_1, g_2 \in \mathbb{G}_2$ be generators, and

*Corresponding author: Yanqin Zhu (yqzhu@suda.edu.cn)

$a,b \xleftarrow{\$} \mathbb{Z}p$ be chosen at random. We define the advantage of an adversary $\mathcal{A}$ in solving the co-Computational Diffie-Hellman problem as

$$\mathrm{Adv}_{\mathcal{A}}^{cdh}(\lambda)=\Pr[\mathcal{A}(p, g_1, g_2, g_1^a, g_2^b) = g_1^{ab}].$$

We say that co-CDH Assumption ϵ-holds in $\mathbb{G}_1, \mathbb{G}_2$ if for every PPT algorithm $\mathcal{A}$ we have that $\mathrm{Adv}_{\mathcal{A}}^{cdh}(\lambda) \leq \epsilon$.

Characteristic polynomial of set. As introduced in [5,6], given a set $X=\{x_1, x_2, \ldots, x_m\}$, the polynomial $Poly_X(s)=\prod_{i=1}^{m}(x_i - s)$ is called the characteristic polynomial of X. Characteristic polynomials constitute representation of sets by polynomials that have the additive inverses of their set elements as roots. The efficiency of computing the polynomial of a set is described as the following lemma.

Lemma 1(Polynomial interpolation with FFT) Let $Poly(s) = \prod_{i=1}^{n}(x_i - s) = \sum_{i=0}^{n} a_i s^i$ be a degree-n polynomial. The coefficients $a_n, a_{n-1}, \ldots, a_0$ can be computed with O($n$log$n$) complexity, given $x_1, x_2, \ldots, x_n$[9].

Closed Form Efficient PRF. It was introduced in [10]. A closed form efficient PRF consists of algorithms (PRF.KG,F). The key generation PRF.KG takes the security parameter 1^λ as input, and outputs a secret key K and some public parameters pp that specify domain X and range Y of the function F. On input $x \in X$, $F_K(x)$ uses the secret key K to compute a value $y \in Y$. It must satisfy the pseudorandom property. Namely, (PRF.KG,F) is ϵ-secure if for every PPT adversary $\mathcal{A}$ it holds:

$$\left|\Pr\left[\mathcal{A}^{F_K(\cdot)}(1^\lambda, \mathrm{pp}) = 1\right] - \Pr[\mathcal{A}^{R(\cdot)}(1^\lambda, \mathrm{pp}) = 1]\right| \leq \epsilon$$

where $(K,\mathrm{pp}) \xleftarrow{\$} \mathrm{PRF.KG}(1^\lambda)$, and $R(\cdot)$ is a random function from X to Y. In addition, it is required to satisfy closed-form efficiency property. For example, an arbitrary computation Comp that takes input l random values $R_1, \ldots, R_l \in Y$ and an arbitrary x, and assume that the best algorithm to compute Comp($R_1, \ldots, R_l, x$) takes time T. $z=(z_1, \ldots, z_l)$ is a l-tuple of arbitrary values in domain X of F. We say that a PRF(PRF.KG,F) is closed-form efficient for (Comp,z) if there exists an algorithm $\mathrm{PRF.CFEval}_{\mathrm{Comp},z}(K,x)=\mathrm{Comp}(F_K(z_1), \ldots, F_K(z_l), x)$ and its running time is o(T).

Authenticated data structures. It is originally defined in [11], which is a model of computation where untrusted responders answer queries on a data structure on behalf of a trusted source and provide a proof of the validity of the answer to the user. Our scheme is based on authenticated data structures, which comprises a collection of six polynomial-time algorithms {GenKey, Setup, Update, Refresh, Query, Verify} such that (a) GenKey() produces the secret and public key of the system; (b) Setup() initializes the authenticated data structure Auth(D), on input a plain data structure D; (c) Having access to the secret key, Update() updates the authenticated data structure(e.g., the polynomial coefficients and its digest), so that it could be used later for query verification; (d) Without having access to the secret key, Refresh() updates the polynomial coefficients and its digest as a whole so that it could be used later for query execution; (e) Query() computes cryptographic proofs $\prod(q)$ for answers $\alpha(q)$ to intersection query q; (f) Verify() processes the proof $\prod(q)$ and the answer $\alpha(q)$ and either accepts or rejects the answer. Note that Verify() is required to have no access to the secret key.

III. Our Construction

Firstly, we show the construction of pseudorandom function that enjoy closed-form efficiency for polynomial evaluation in our authenticated data structure. We just need to consider the situation that polynomials in one variable and degree at most d. As we defined in section 2, a closed form efficient PRF consists of algorithms (PRF.KG,F).

PRF.KG(1^λ ,s). Generate a group description $(p, g, \mathbb{G}) \xleftarrow{\$} \mathcal{G}(1^\lambda)$. Choose $4s+2$ values $y_0, z_0, \{y_j, z_j, w_j, v_j\}_{1 \leq j \leq s} \xleftarrow{\$} Z_p$. Output $K=y_0, z_0, \{y_j, z_j, w_j, v_j\}_j$.

$\mathbf{F_K(i)}$. The domain of the function is $i \in [0,\ldots,d]$, but we interpret each $i=(i_1, \ldots, i_s)$ as a binary string of $s=\lceil \log d \rceil$ bits. The function is computed by the following algorithm:

Initialize $a \leftarrow y_0, b \leftarrow z_0$

For $j=1$ to s:

If $i_j=0$, then $a \leftarrow a, b \leftarrow b$.

Else, $a \leftarrow a \cdot y_j + b \cdot z_j, b \leftarrow a \cdot w_j + b \cdot v_j$

Output g^a

In what follows we show that our PRF admits closed form efficiency for polynomials.

Consider any polynomial $p(x)$ in one variable of degree at most d. This polynomial has up to $l=d+1$ terms which can index with i (i=0,1…,d). $\mathrm{Poly}(\{R_i\}_{0 \leq i \leq d}, x)=\prod_{0 \leq i \leq d} R_i^x = g^{p(x)}$, where $p(\cdot)$ is the polynomial whose coefficients are the discrete logs of the R values. We now show that if we set $R_i=F_K(i)$, then there exists an algorithm $\mathrm{PRF.CFEval}_{Poly}(K,x)$ that compute $g^{p(x)}=\prod_{0 \leq i \leq d} F_K(i)^x$ in time O(log$d$), instead of the regular computation running time O($d$log$d$).

For ease of exposition, we first describe an alternative equivalent algorithm for computing $F_K(i)$. Denote $f_K(i) = (f_K^1(i), f_K^2(i))$ as the following recursive function:

If i=0, then $f_K^1(0)=y_0$ and $f_K^2(0)=z_0$.

Else:

Let $i=(i_1, i_2, \ldots, i_s)$, j be such that $i_{j+1} = \cdots = i_s = 0$ and $i_j \neq 0$.

$$f_K^1(i)=f_K^1(i_1, \ldots, i_{j-1}, 0, \ldots 0)y_j + f_K^2(i_1, \ldots, i_{j-1}, 0, \ldots 0)z_j$$

$$f_K^2(i)=f_K^1(i_1, \ldots, i_{j-1}, 0, \ldots 0)w_j + f_K^2(i_1, \ldots, i_{j-1}, 0, \ldots 0)v_j$$

Finally, the value of the function $F_K(i)=g^{f_K^1(i)}$.

$\mathbf{PRF.CFEval_{Poly}(K,x)}$. Set $s=\lceil \log d \rceil$, $p_s(x) = (p_s^1(x), p_s^2(x))$ be the following recursive function:

If s=1, then $p_s^1(x)=y_0, p_s^2(x)=z_0$.

Else

$$p_s^1(x)=p_{s-1}^1(x)+x^{2^{s-1}}(p_{s-1}^1(x) \cdot y_s + p_{s-1}^2(x) \cdot z_s)$$

$$p_s^2(x)=p_{s-1}^1(x)+x^{2^{s-1}}(p_{s-1}^1(x)\cdot w_s+p_{s-1}^2(x)\cdot v_s)$$

The algorithm output $g^{p(x)}=g^{p_s^1(x)}$.

The algorithm makes only one recursive call at each step, thus it runs in time O(log$d$).

The description of our authenticated data structure scheme $\mathcal{ACD}$ for a sets collection data structure is as follows.

{sk,pk}←Genkey(1^λ): Generate the description of bilinear groups $(p, g_1, g_2, \mathbb{G}_1, \mathbb{G}_2, \mathbb{G}_T, e)$ and keys of PRFs, $K_i \xleftarrow{\$}$ PRF.KG(1^λ,⌈log$i$⌉): i=1,…,max{|S|}, where max{|S|} is the maximum size of set that system specified. Subsequently, a function h: $Z_p \to Z_p^*$ which expands the input x to the vector $(h_1(x), h_2(x), \dots)$, where $h_i(x)=x^i$. Choose a random $\beta \xleftarrow{\$} Z_p$, and compute $e(g_1,g_2)^\beta$. Finally, the algorithm outputs sk = $\mathbf{K}=\left[K_1,\dots,K_{max\{|S|\}}\right]$, pk = $\{h(\cdot),(p,g_1,g_2,\mathbb{G}_1,\mathbb{G}_2,\mathbb{G}_T,e), e(g_1,g_2)^\beta\}$. This algorithm has O(1) complexity.

{D_0,Auth(D_0)}←Setup(D_0,sk,pk): Let D_0 be our initial data structure, i.e. the sets $S_1,S_2,\dots,S_m$. The authenticated data structure Auth(D_0) is built as follows: For each set S_i, the characteristic polynomial $\text{Poly}_{S_i}(s)=\prod_{x\in S_i}(x-s)=\sum_{j=0}^{|S_i|}a_{i,j}s^j$ coefficients $\boldsymbol{a_i}=\left[a_{i,0},\dots,a_{i,|S_i|}\right]$ are computed with complexity O(|S_i|log|S_i|). Subsequently, the algorithm computes the corresponding digest $\mathbf{W_i}=\left[W_{i,0},\dots,W_{i,|S_i|}\right]$, where $W_{i,j}=g_1^{\beta\cdot a_{i,j}}\cdot F_{K_{|S_i|}}(j), \forall j=1,\dots,|S_i|$ for later polynomial evaluation with complexity O(log|S_i|). The algorithm outputs all sets S_i as data structure D_0, all polynomial coefficients $\boldsymbol{a_i}$ and its digest $\mathbf{W_i}$ for $1\le i\le m$ as Auth(D_0).

{D_{h+1},Auth(D_{h+1}),upd}←Update(u, D_h,Auth(D_h),sk,pk): Suppose the update is "insert element $x\in\mathcal{U}$ into set S_i". The algorithm initially sets $\text{Poly}'_{S_i}(s)=\text{Poly}_{S_i}(s)\cdot(x-s)$. Note that in the case of deleting x from S_i, the algorithm sets $\text{Poly}'_{S_i}(s)=\text{Poly}_{S_i}(s)/(x-s)$. Then the updated polynomial coefficients $\boldsymbol{a'_i}$ and their digest $\mathbf{W'_i}$ are computed with complexity O(|S_i|log|S_i|+ log|S_i|). The algorithm output $\boldsymbol{a'_i}$ and $\mathbf{W'_i}$ as information upd. Information upd also include x. Finally, the new authenticated data structure Auth(D_{h+1}) is updated by replacing $\boldsymbol{a_i}$ with $\boldsymbol{a'_i}$ and $\mathbf{W_i}$ with $\mathbf{W'_i}$.

{D_{h+1},Auth(D_{h+1})} ←Refresh (u, D_h,Auth(D_h),upd,pk): Suppose the update is "insert element $x\in\mathcal{U}$ into set S_i". The algorithm updates the polynomial coefficients and its digest of S_i by using information upd. Finally, it output the updated sets collection as D_{h+1} and the updated polynomial coefficients and digests as Auth(D_{h+1}).

{α(q), $\prod(q)$}←Query(q, D_h,Auth(D_h),pk): The query q is set of indices {1,2,…,t}, requiring the intersection of $S_1,\dots,S_t$. Let $S_r=\{x_1,x_2,\dots,x_n\}$ represent the set which has minimum elements among sets that involved in this query, i.e. $|S_r|=\text{Min}_{i=1,\dots,n}\{|S_i|\}$. For each element x_j(j=1,…,n) in S_r, compute the valuation of polynomial $\text{Poly}_{S_i,j}(x_j)=\sum_{k=0}^{|S_i|}a_k\cdot h_k(x_j)$ and the witness $\text{V}_{S_i,j}=\prod_{k=0}^{|S_i|}W_{i,k}^{h_k(x_j)}$ with complexity O(|S_i|), where $S_i\subset\{S_1,S_2,\dots,S_t\}/S_r$ and $h_k(x_j)=(x_j)^k$. The intersection result α (q) will be consisted of the element $x_k(1\le k\le n)$ in S_r that satisfies the conditions: $\text{Poly}_{S_i,k}(x_k)=0, \forall S_i\subset\{S_1,S_2,\dots,S_t\}/S_r$. The proof $\prod(q)$ consists of the following parts: (1) S_r which owned minimum elements among involved sets; (2)The witness of polynomial evaluation $\text{V}_{S_i,j}$ for all $S_i\subset\{S_1,S_2,\dots,S_t\}/S_r$, j=1,…,n; (3)The public verification key $\text{VK}_{S_i,j}=\text{PRF.CFEval}_{Poly_{S_i,j}}\left(K_{|S_i|},h(x_j)\right)$ which needs to be computed by client with complexity O(log|S_i|), because it requires secret key $K_{|S_i|}$ for computation.

{Accept, Reject}←Verify(q, α(q), $\prod(q)$,pk): If the answer I=α(q) is correct for the intersection query q. It must satisfy the following two conditions. (1) Subset condition: $\text{I}\subseteq S_1\wedge\text{I}\subseteq S_2\wedge\dots\wedge\text{I}\subseteq S_t$. It can be checked by checking the equations

$$e(\text{V}_{S_i,j},g_2)=\text{VK}_{S_i,j} \quad (1)$$

for all $S_i\subset\{S_1,S_2,\dots,S_t\}/S_r$, $x_j\in$I. This step has O($\delta+t$) complexity, where δ is the size of α(q). If any of the above checks fails, the algorithms outputs reject; (2)Completeness condition: $(S_1-\text{I})\cap(S_2-\text{I})\cap\dots\cap(S_t-\text{I})=\emptyset$. It can be checked by checking that for all $x_j\in S_r-$I, $\exists S_i\subset\{S_1,S_2,\dots,S_t\}/S_r$ satisfies the inequation

$$e(\text{V}_{S_i,k},g_2)\neq\text{VK}_{S_i,j}. \quad (2)$$

This step has O(|S_r|-δ+t). If those relations hold, the algorithm accepts α(q) as the correct intersection.

IV. SECURITY AND COMPLEXITY ANALYSIS

In this section, we give a correctness specification of our authenticated data structure scheme and a proof for the security of one intersection query. We also compare our protocol's complexity with various schemes.

Definition 1. (Correctness of authenticated data structure scheme) Let $\mathcal{ACD}$ be an authenticated data structure scheme {GenKey, Setup, Update, Refresh, Query, Verify}. We say that the authenticated data structure scheme $\mathcal{ACD}$ is correct if, for all $\lambda\in\mathbb{N}$, for all {sk,pk} output by algorithm Genkey(), for all D_h,Auth(D_h) output by one invocation of Setup() followed by polynomial-many invocations of refresh(), where $h\ge0$, for all intersection queries and for all α(q), $\prod(q)$ output by Query(q, D_h, Auth(D_h),pk), with all but negligible probability that Verify(q, α(q), $\prod(q)$,pk) rejects.

We node that for relation 1, it indeed holds $e(\text{V}_{S_i,j},g_2)=e\left(\prod_{k=0}^{|S_i|}W_{i,k}^{h_k(x_j)},g_2\right)=e\left(\prod_{k=0}^{|S_i|}\left(g_1^{\beta\cdot a_{i,j}}\cdot F_{K_{|S_i|}}(j)\right)^{x_j^k},g_2\right)=e\left(g_1^{\beta\cdot Poly_{S_i}(x_j)+R_{K_{|S_i|}}(x_j)},g_2\right)=e\left(g_1^{R_{K_{|S_i|}}(x_j)},g_2\right)=$ PRF.CFEval$_{Poly_{S_i,j}}\left(K_{|S_i|},h(x_j)\right)=\text{VK}_{S_i,j}$,when all the witnesses $\text{V}_{S_i,j}$ and verification keys $\text{VK}_{S_i,j}$ have been computed honestly for all element in the result set I which satisfies $\text{Poly}_{S_i}(x_j)=0$. However, for the completeness verification, we didn't verify the equation $e(\text{V}_{S_i,k},g_2)=\left(e(g_1,g_2)^\beta\right)^y\text{VK}_{S_i,j}$, where y= $\text{Poly}_{S_i}(x_j)$, $x_j\in S_r-$ I, but the inequation $e(\text{V}_{S_i,k},g_2)\neq\text{VK}_{S_i,j}$. Because we don't care the correctness of

polynomial evaluation when $Poly_{S_i}(x_j) \neq 0$ and verifying the inequation is much easier than the equation due to the computation of exponents.

Definition 2. (Security of authenticated data structure scheme) Let $\mathcal{ACD}$ be an authenticated data structure scheme {GenKey, Setup, Update, Refresh, Query, Verify}, {sk,pk} ← Genkey(1^{λ}), and $\mathcal{A}$ be a polynomial-bounded adversary that is only given pk. The adversary has unlimited access to all algorithms of $\mathcal{ACD}$, except for algorithms Setup() and Update() to which has only oracle access. If $\mathcal{G}$ is such that the co-CDH assumption ϵ_{cdh} -holds, and characteristic polynomial $Poly_i$ corresponding to set S_i which participates in the intersection query Q is ϵ_i-secure, then for any PPT adversary $\mathcal{A}$ making at most q=poly(λ) queries it is

$$\Pr\left[\begin{matrix}\{\mathrm{Q}, \prod(\mathrm{Q}), \alpha(\mathrm{Q}), q\} \leftarrow \mathrm{Adv}_{\mathcal{A}}^{\mathcal{ACD}}(1^{\lambda}, \mathrm{pk}); \\ \mathrm{Accept} \leftarrow \mathrm{Verify}(\mathrm{Q}, \alpha(\mathrm{Q}), \prod(\mathrm{Q}), \mathrm{pk}); \\ \mathrm{Reject} \leftarrow \mathrm{Check}(\mathrm{Q}, \alpha(\mathrm{Q}), D_h)\end{matrix}\right] \leq \epsilon_{cdh} + \sum_{i=\{1,\ldots,t\}} \epsilon_i,$$

where {Accept,Reject}←Check(Q, α(Q), D_h) is a method that decides whether α(Q) is a correct answer for query Q on data structure D_h and t is the number of sets involved in this intersection query.

To prove the security of our authenticated data structure scheme, we define the following games, where $\mathcal{R}_i(\mathcal{A})$ denotes the output of Game i run with adversary $\mathcal{A}$:

Game 0: For our verifiable computation scheme $\mathcal{ACD}$, we define the following experiment:

Experiment $\mathrm{Exp}_{\mathcal{A}}[\mathcal{ACD}, D_0, \lambda, \mathrm{Q}]$

$\{sk, pk\}$←Genkey(1^{λ})

$\{D_0, \mathrm{Auth}(D_0)\}$←Setup($D_0, sk, pk$)

For i=1 to q:

$(D_i, \mathrm{Auth}(D_i)) \leftarrow \mathcal{A}(pk, D_1, \mathrm{Auth}(D_1), upd_i, \ldots, D_{i-1}, \mathrm{Auth}(D_{i-1}), upd_{i-1})$

$(D_{i+1}, \mathrm{Auth}(D_{i+1}), upd_{i+1}) \leftarrow \mathrm{Update}(u_i, D_i, \mathrm{Auth}(D_i), sk, pk)$

$(D^*, \mathrm{Auth}(D^*)) \leftarrow \mathcal{A}(pk, D_1, \mathrm{Auth}(D_1), upd_1, \ldots, D_{q+1}, \mathrm{Auth}(D_{q+1}), upd_{q+1})$

$(\widehat{D}^*, \mathrm{Auth}(\widehat{D}^*)) \leftarrow \mathrm{Refresh}(u, D^*, \mathrm{Auth}(D^*), upd, pk)$

$(\hat{\alpha}(\mathrm{Q}), \hat{\Pi}(\mathrm{Q})) \leftarrow \mathcal{A}(\mathrm{Q}, D_1, \mathrm{Auth}(D_1), upd_1, \ldots, D_{q+1}, \mathrm{Auth}(D_{q+1}), upd_{q+1}, \widehat{D}^*, \mathrm{Auth}(\widehat{D}^*), pk)$

$\hat{\mathcal{R}} \leftarrow \mathrm{Verify}(\mathrm{Q}, \hat{\alpha}(\mathrm{Q}), \hat{\Pi}(\mathrm{Q}), pk)$

Game1: this game is similar to Game 0, except for the following change in the evaluation of the pseudorandom function algorithm. For the query Q asked by the adversary during the game, instead of computing verification key VK_Q using PRF.CFEval algorithm, the inefficient evaluation $\mathrm{VK}_Q = \prod_{i=1}^{|S|} F_K(i)^{h_i(x)}$ is used.

Game2: this game is the same as Game1, expect that each $F_K(i)$ is replaced by an element $R_i \leftarrow \mathbb{G}_1$ chosen uniformly at random.

The proof of the security proceed by a standard hybrid argument, and is obtained by combining the proof of the following claims.

Claim. $\Pr[\mathcal{R}_0(\mathcal{A})=\mathrm{Accept}] = \Pr[\mathcal{R}_1(\mathcal{A})=\mathrm{Accept}]$.

Proof. The only difference between the two games is in the computation of the pseudorandom function algorithm. However, by correctness of PRF.CFEval, such difference does not change the distribution of the values verification key returned to the adversary. Thus, the probability of the adversary winning the Game 1 doesn't change.

Claim. $\Pr[\mathcal{R}_1(\mathcal{A})=\mathrm{Accept}] - \Pr[\mathcal{R}_2(\mathcal{A})=\mathrm{Accept}] \leq \sum_{i=\{1,\ldots,t\}} \epsilon_i$

Proof. The difference between Game 2 and Game 1 is that we replace the pseudorandom function with uniformly random group elements. It is easy to see that any adversary $\mathcal{A}$ for which such difference is greater than $\sum_{i=\{1,\ldots,t\}} \epsilon_i$ can be reduced to an attacker that has same advantage against the security of the PRF.

Claim. $\Pr[\mathcal{R}_2(\mathcal{A})=\mathrm{Accept}] \leq \epsilon_{cdh}$.

Proof. Assume by contradict that there exists a PPT adversary $\mathcal{A}$ that has advantage greater than ϵ_{cdh} of winning in Game 2, then we show that we can build an efficient algorithms $\mathcal{B}$ which uses $\mathcal{A}$ to solve the co-CDH problem for $\mathcal{G}$ with the same probability ϵ_{cdh}.

$\mathcal{B}$ takes input as a group description $(p, g_1, g_2, \mathbb{G}_1, \mathbb{G}_2, \mathbb{G}_T, e)$ and random elements g_1^a, g_2^b, and it proceeds as following. It computes $e(g_1^a, g_2^b)$ as public key. For each set $S_i, 1 \leq i \leq m$, the characteristic polynomial is $\mathrm{Poly}_{S_i}(s) = \sum_{j=0}^{|S_i|} a_{i,j} s^j$, it chooses $\mathbf{W_i} = [W_{i,0}, \ldots, W_{i,|S_i|}]$, where $W_{i,j} \xleftarrow{\$} \mathbb{G}_1$. It is easy to check that the public key and data structure digest $\mathbf{W_i}$ are perfectly distributed as in Game 2. Next, for each $\mathbf{W_i}$, it computes $\boldsymbol{Z_i} = e(\mathbf{W_i}, g_2)/\mathrm{pk} \in \mathbb{G}_T$. For a set intersection query Q, $\mathcal{B}$ runs $\mathcal{A}(q, pk)$ and answer this query as follows. $\mathcal{B}$ computes $\mathrm{VK}_{S_i,j} = \prod_{k=1}^{|S_i|} Z_{i,k}^{\,h_k(x_j)}$, where $x_j \in S_r$, $S_i \subset \{S_1, S_2, \ldots, S_t\}/S_r$, $h_k(x_j) = (x_j)^k$ and returns to $\mathcal{A}$. By the bilinear property of $e(\cdot, \cdot)$, this computation of $\mathrm{VK}_{S_i,j}$ is equivalent to the one in Game 2.

Finally, let $\hat{\alpha}(\mathrm{Q}) = \{\hat{x}_1, \ldots \hat{x}_\delta\}$, $\hat{\Pi}(\mathrm{Q}) = \{\hat{V}_{S_i,j}\}_{S_i \subset \{S_1, S_2, \ldots, S_t\}, 1 \leq j \leq |S_r|}$ be the output of $\mathcal{A}$ at the end of the game, such that for some $\widehat{D}^*, \mathrm{Auth}(\widehat{D}^*)$ chosen by $\mathcal{A}$ it holds Verify(Q, $\hat{\alpha}$(Q), $\hat{\Pi}$(Q),pk)= Accept, but $\hat{\alpha}(\mathrm{Q})$ is not the correctness result of query Q. By verification, this means that for each element $\hat{x}_j$ in $\hat{\alpha}(\mathrm{Q})$, it satisfies the equation $e(\hat{\mathrm{V}}_{S_i,j}, g_2) = \mathrm{VK}_{S_i,k}$. Let $y_j \neq 0$ be the correct output of the polynomial $\mathrm{Poly}_{S_i}(\hat{x}_j)$. Then, by correctness it also holds: $e(\mathrm{V}_{S_i,j}, g_2) = e(g_1^a, g_2^b)^{y_j} \cdot \mathrm{VK}_{S_i,k}$ where $\mathrm{V}_{S_i,j} = \prod_{k=0}^{|S_i|} W_{i,k}^{h_k(\hat{x}_j)}$. So, dividing the two verification equations, we obtain that $e(\hat{\mathrm{V}}_{S_i,j}/\mathrm{V}_{S_i,j}, g_2) = e(g_1^a, g_2^b)^{-y_j}$. $\mathcal{B}$ can thus compute $g_1^{ab} = (\hat{\mathrm{V}}_{S_i,j}/\mathrm{V}_{S_i,j})^{-1/y_j}$. Therefore, if $\mathcal{A}$ wins in Game 2 with probability ϵ_{cdh}, then $\mathcal{B}$ solves co-CDH with the same probability.

Efficiency analysis. To explicitly compare our protocol complexity with other schemes, we adopt the complexity model

used in memory checking[5]. It considers the number of primitive cryptographic operations. The access complexity of an algorithm is defined as the number of memory accesses this algorithm performs on the authenticated data structure stored in memory. The group complexity of a data collection is defined as the number of elementary data objects(e.g., group elements or elements in Z_p) contained in that object. Our protocol complexity is described as following. Firstly, the access complexity for running Setup is O(M)(the computation of polynomial coefficients and its digest), where M=$\sum_{i=1}^{m}|S_i|$, and the group complexity of Auth(D_0) is also O(M) since the algorithm stores polynomial coefficients and its digest for each set S_i as Auth(D_0), and the number of coefficients corresponding to S_i is $|S_i|$+1. For the Update and Refresh algorithms, they need to overwrite the polynomial coefficients and their digest values. Therefore the access complexity of these algorithms is O(N), where N is the size of updated set. For an intersection query on t sets, the access complexity of offline polynomial evaluation and its witness computation is O(N),where N is the sum of sizes of sets that are involved in this query and of total group complexity is O(t). Whereas the client's online access complexity for verification key computation is O(logN) and has total group complexity O(t).The Verify algorithm involves subset verification and completeness verification, the total group complexity is O(t+n), where n is the minimum size of sets involved in this query. A comparison of our work with existing schemes appears in the following table.

TABLE I. COMPLEXITY COMPARATION

	Setup	**Update& Refresh**	**Query**	**Verify**	**upd**
DGMS00[12] YPPK09[13]	m+M	logN+log$m$	N+log$m$	N+log$m$	1
MBKK04[14]	m+M	m+M	N	N	N
PT04[15]	m^t+M	m^t	1	δ	m^t
PTT11[5] CPPT14[6]	m+M	log$m$	tm^{ϵ}log$m$ +Nlog^{3}N	t+δ	log$m$
this	M	N	N+logN	t+n	N

This table describes the asymptotic access and group complexity of various schemes. For a set collection data structure of m sets: The sum of sizes of all the sets is M. For an intersection query on t sets, it outputs δ elements. And N is the sum of sizes of these t sets, n is the minimum set size among these t sets, $0< \epsilon <1$ is a constant.

V. CONCLUSIONS

In this paper, we present an optimal delegation of set intersection over dynamic dataset. Our solution provide two important properties, namely public verifiability and efficiency at the same time. According to the definition of optimality proposed in [14], our construction is nearly optimal. And we have solved the problem "Whether an optimal authenticated sets collection data structure is possible?" proposed in [4]. We reduced the query time from O(Nlog^{3}N) to O(N). In addition, on the problem that whether outsourced verifiable computation with secrecy, public verifiability and efficiency exists. We consider homomorphicly encrypting the coefficients of characteristic polynomial and the server will execute homomorphic addition and multiplication for polynomial evaluation, any party can verify the correctness of the encrypted result, the client just need to decrypt it and get the result However, the homomorphic computation will increase the overhead at server. It will result in a decline in the efficiency of the system. Thus, a better solution for secret, public verifiable and efficient outsourced computation is our future work.

ACKNOWLEDGEMENT

This work was supported by National Natural Science Foundation of China under grant No. 61373164, Research Project of Jiangsu Province under grant No.BY2013030-06, Suzhou Application Foundation Research Project under Grant No. SYG201238.

REFERENCES

[1] B. Applebaum, Y. Ishai, and E. Kushilevitz. From secrecy to soundness: Efficient verification via secure computation. In International Colloquium on Automata, Languages and Programming (ICALP), pages 152–163. Springer, 2010.

[2] K.-M. Chung, Y. Kalai, and S. Vadhan. Improved delegation of computation using fully homomorphic encryption. In CRYPTO, 2010.

[3] R. Gennaro, C. Gentry, and B. Parno. Non-interactive verifiable computing: Outsourcing computation to untrusted workers. In CRYPTO, 2010.

[4] M. Backes, D. Fiore, and R. M. Reischuk. Verifiable delegation of computation on outsourced data. Cryptology ePrint Archive, Report 2013/469, 2013.

[5] C. Papamanthou, R. Tamassia, and N. Triandopoulos. Optimal verification of operations on dynamic sets. In CRYPTO, pages 91–110, 2011.

[6] R.Canetti, O.Paneth, D.Papadopoulos, N.Triandopoulos. Verifiable set operation over outsourced database. In Public-Key Cryptograpgy, pages 113-130, 2014.

[7] L. Nguyen. Accumulators from bilinear pairings and applications. In CT-RSA, pages 255–292, 2005.

[8] C. Papamanthou, R. Tamassia, and N. Triandopoulos. Authenticated hash tables. In ACM Conference on Computer and Communications Security, pages 437–448, 2008.

[9] F. Preparata, D. Sarwate, and I. U. A. U.-C. C. S. LAB. Computational Complexity of Fourier Transforms Over Finite Fields. Defense Technical Information Center, 1976.

[10] S. Benabbas, R. Gennaro, and Y. Vahlis. Verifiable delegation of computation over large datasets. In P. Rogaway,editor, Advances in Cryptology - CRYPTO 2011, volume 6841 of Lecture Notes in Computer Science, pages 111-131, Santa Barbara, CA, USA, Aug. 14-18, 2011. Springer, Berlin, Germany.

[11] R. Tamassia. Authenticated data structures. In ESA, pages 2–5, 2003.

[12] P. Devanbu, M. Gertz, C. Martel, and S. G. Stubblebine. Authentic third-party data publication. In Fourteenth IFIP 11.3 Conference on Database Security, 2000.

[13] Y. Yang, D. Papadias, S. Papadopoulos, and P. Kalnis. Authenticated join processing in outsourced databases. In ACM International Conference on Management of Data (SIGMOD), pages 5–18, 2009.

[14] R. Morselli, S. Bhattacharjee, J. Katz, and P. J. Keleher. Trust-preserving set operations. In INFOCOM,2004.

[15] H. Pang and K.-L. Tan. Authenticating query results in edge computing. In Proc. Int. Conf. on Data Engineering, pages 560–571, 2004.

Workload Forecasting Framework for Applications in Cloud

Shuang Jiang, Haopeng Chen, Fei Hu
School of Software Engineering,
Shanghai Jiaotong University,
Shanghai, China
Email: {daxianji008,chen-hp,hufei}@sjtu.edu.cn

***Abstract*—With the developing of cloud computing technics, an increasing number of applications prefer to be deployed in cloud. Load balancing becomes the key technicfor cloud provider to control the resources and cost.But using load balancing with real time data can`t react in time towards workload peak or valley. Thus, workload forecasting is presented to let the cloud provider to get ready for a possible workload change. There are already many kinds of predicting methods. In this article, we study the workload of applications in cloud and propose a workload forecasting framework. This framework monitors workloads of applications in real time, processes the data, and provides feedback of the predicted workload value according to historical data,guiding the cloudprovider to allocate resources.**

Keywords—cloud; application; workload; forecast; framework;

I. INTRODUCTION

Today we experience an increasingly need for collaboration, data sharing, and other modes of interaction that involve multiple heterogeneous and geographically distributed resources, including supercomputers, PC's, PDA's, workstations, storage systems, databases, and special purpose applications with various requirements (CPU, I/O, or memory) [1]. For this reason, new abstractions and concepts should be introduced at network architecture and middleware level to allow applications to access and share resources or services in an efficient manner.[1,2].

Many technics and concepts are presented to solve part of those problems, such as virtualization, utility computing, distributed computing, computational grid.

Based on those technics, cloud computing is presented. Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. [6] It inherits many advantages of the above technics.

Load balancing and resources allocation are two of the key technics of cloud computing. Load balancing distributes workloads across multiple computing resources, such as computers, a computer cluster, network links, central processing units or disk drives [7]. Allocation mechanisms should know the status of each element/resource in the distributed cloud environment and, based on them, intelligently apply algorithms to better allocate physical or virtual resources to applications according to their pre-established requirements [8]. The cloud provider should monitor the workload at run time and allocate resources accordingly to avoid overload of the application as well as resources waste.

However, a problem with such a resource allocation scheme is that allocating at run time has more or less delay while the users of the application may not patient enough to wait for that long. Another problem is the chance of thrashing where due to frequent variation of workload, machines can be added and released on every sample – a process that involves significant overhead. A desirable solution would require an ability to predict the incoming workload on the system and allocate resources a priori. This capability in turn will enable the application to be ready to handle the load increase when it actually occurs [9].

Thus, it is really important to apply workload forecasting method in cloud. This article presents a 'Workload Forecasting Framework for Applicationsin Cloud'. The framework monitors workloads of applications in real time, processes the data, and provides feedback of the predicted workload according to historical data, which guides the cloudprovider to allocate resources in advance.

II. BACKGROUND

A. Why We Need Workload ForecastingAlgorithms?

For the cloud provider, only when he gets precise predicted workload can he make suitable load balancing. Good predict algorithm means good resources allocation and economic benefit. We can say it is also a key technic in cloud. But we can`t simply guess the future workload as a predicted value. Finding a suitable workload forecasting method is not an easy and straightforward task.Overcoming these challenges will require algorithms which take into account the following: (i) overheads related to state transition when number of resources are changed, (ii) ability to accurately predict future workload, and (iii) compute the right number of resources required for the expected increase or decrease in workload [9].

So it is important to have a suitable forecasting algorithm.Fortunately, forecastingalgorithm is not first presented for cloud computing and has a long history. Many

people devote their effort to the algorithms. We have many algorithms based on a variety of models. We can easily leveragethem and have many articles to consult.

B. Whynot using a 'Silver Bullet'?

We have large numbers of workload forecasting methods nowadays. They have many differences: They are based on different models, and make different improvements. Some focus on the seasonality of data and some focus on the noise reduction. Different models have different innate advantages and disadvantages. They also have different result on different benchmarks. That means the specific forecasting method`s accuracy is related to the specific benchmark.

To be detailed, suppose we have 3 applications in cloud: an online game, a world cup live website, a road monitoring system. For the workload issue, online game should have more workload on day than night, more workload on holidays than workdays. World cup live website should have a workload peak on match days while little workload on other days. Road monitoring system should have a constant amount of workload everyday.For the workload types, online game may have high CPU utilization and many upload and download operations. World cup live website`s bottleneck must be the bandwidth, and almost all of them are downloading. Road monitoring system may consider the Disk I/O as an important indicator, and almost all of them are writing.

Therefore, the applications have different features, different workload characteristics and different main workload types. We can`t simply use one 'silver bullet' forecasting method to solve the problem. We need a unified forecasting framework, guiding the cloudprovider to apply the most suitable method to the specific applications.

III. Related Work

A. Cloud Computing

Cloud computing is firstly presented by Google CEO Eric Schmidt in 2006. And its first cloud computing project comes from project 'Google 101' [10]. Cloud computing technics started fast developing since then. Thanks to the efforts those people devote in cloud technics, today we can already see cloud applications everywhere just after 8 years.

B. Basic Technics

Load balancing is presented on many areas, including software and hardware. In 1995, it is used in parallel and distributed systems which are the base technics of cloud [11].It becomes one of the key technics of cloud computing later.

Workload prediction is also not first presented for cloud computing. It was even used in many areas besides computer science. It is basically a math method.

It would be better to predict workload if we know some statistical properties of the workload. In 1997 and 1998, Peter A. Dinda collected week-long, 1 Hz resolution Unix load average traces on 38 different machines including production and research cluster machines, compute servers, and desktop workstations. Dinda concluded several statistical properties which are important reference of later researches. The workload have mainly 7 kinds of statistical properties:

- The change of workload is a random process.
- The amount of workloads is mainly at a low level while it has strong volatility.
- Heavier workload has more absolute volatility but less relative volatility. And it is more valuable predicting heavy workload.
- The workload distribution is complex especially for heavy workload. Heavy workload has multiple characteristics. It is better using pattern-driven predict method than just analyzing the workload distribution.
- Workload is strongly related to time series. The historical workload does impact the future workload. Thus, workload prediction is a possible task.
- The change of workload has self-similarity but it is complex. Thus, workload modelling and prediction is a tough task.
- Workload may have sudden change. This indicates that the prediction method should have adaptation mechanism.

C. Workload Prediction

Prediction algorithm has a long history. We main focus on those related to cloud computing.

(linear)Autoregression model,autoregression method utilizes the self-correlation of workload to build a regression formula, then predicts workload by the formula. It is the most classic prediction method presented by H. Akaike in 1969. Large numbers of later prediction algorithms are based on autoregression.

Urgaonkaret. al. [12] have used virtual machines(VM) to implement dynamic provisioning of multi-tieredapplications based on an underlying queuing model. Foreach physical host, however, only a single VM can berun. Wood et. al. [13] use a similar infrastructure as in [12].They concentrate primarily on dynamic migration of VMsto support dynamic provisioning. They define a uniquemetric based on the consumption data of the three resources:CPU, network and memory to make the migration decision.Cunha et. al. [14] Padalaet.al. [15] provide a control-theoretic solution where each tierof the application is executed on each virtual machine.Authors carry out black box profiling of the applicationsand build an approximated model which relates performanceattributes such as response time to the fraction of processorallocated to the virtual machine running the application.Wang et. al. [16] describe a two-level control architecture for avirtualized environment. A load balancing controller ensuresthat the virtual machines are all load-balanced and theresponse time of the applications in all the virtual machinesare the same. Moreno et. al. [17]Schopf and Berman presented a confidence window method [18]. This algorithm outputs a confidence interval of workload instead of a single value. This algorithm can achieve a high accuracy.

Dinda and O'Hallaron evaluated several mainstream prediction models [19]. The conducted detailed experiments and came to a conclusion that it is already acceptable using autoregressive method on single CPU utilization. This

explained the situation that many prediction algorithms choose to use this simple model.

IV. SYSTEM ARCHITECTURE

A typical application in cloud may be deployed as Fig. 1. In one cloud there may be several applications. One application has a corresponding dispatcher and servers. When user sends a request to the application, the request should be first sent to the dispatcher, and be dispatched to the corresponding server according to 'VM-app' mapping rules. After request processed, the response will be sent back to the dispatcher then sent to user. There exists a module 'Resources Allocator' to dynamically allocate the virtual machines to the application and maintain the 'VM-app' mapping rules.

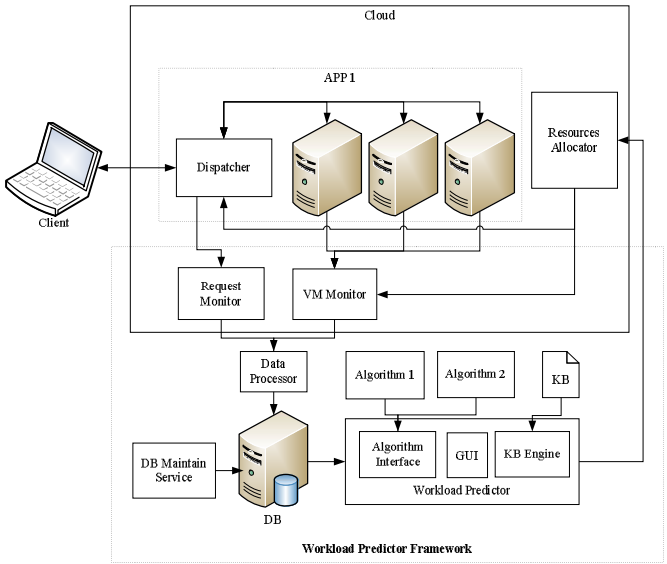


Fig. 1. Cloud&Framework deployment

The forecasting framework consists of 5 modules, they are 'monitor module', 'pre-process module', 'database module', 'maintain data module', 'workload forecasting module'. They should be deployed into cloud as plug-ins(that is to say the framework is application-independent). Fig. 1 shows the main architecture.

A. Monitor Module

Monitor module consists of 2 sub module:

Request monitor, the implementation of this sub module is related to the dispatcher in cloud, but not related to the application. It should be a plug-in of the dispatcher, monitoring the requests received in real time. It would record the amount of requests every T_{mr} seconds. While the requests are recorded as a vector (T,I,H,P_o,D,P_u,O), meaning the record time, the record time interval(the same as T_{mr}), the 'request intensity'(it will be discussed later) of GET, HEAD, POST, DELETE, PUT during that time interval.

VM monitor, this is an application-independent sub module. But it would access the 'VM-app' mapping rules. The module reads the system resources utilization of all the VMs which are running the specific application every T_{mv} seconds. It records the resources utilization every T_{mw} seconds. While the resources utilization are recorded as a vector (T,I,C,M,D), meaning the record time, the record time interval(the same as T_{mw}), CPU utilization, the memory utilization, the disk utilization at that time.

The result of the 2 sub modules will be merged and sent to the next 'data processor module' every T_{ms} seconds;

B. Data Processor Module

Data processor module consists of 4 tasks. They are all triggered at constant time intervals.

The module triggers 'data merge' task every Tdm seconds. It gets the two recorded data of the monitor module, merges them and put them into a circular queue.

The module triggers'data filter' task every Tdf seconds. It traverses the circular queue. Filter the data. The filter is custom. Whether checking the input format or Gaussian Filter is OK.

The module triggers'data accelerator' task every Tda seconds. It pre-processes the data, accelerating the speed of the later calculating.

The module triggers'data record' task every T_{dr} seconds. It will write the latest processed data into the database.

C. Database Module

We record the data in database for long-time storage and interaction with other module. The data is merged from the HTTP requests and the VM resources. It is recorded as a vector (T,I,G,Po,D,Pu,O,C,M,D), meaning the record time, the record time interval, the 'request intensity' of GET request, HEAD request, POST request, DELETE request, PUT request, OPTIONS request, the CPU utilization, the Memory utilization, the Disk I/O utilization separately. The amount of data should be large. It is recommended to use non-relational database.

D. Maintain Data Module

The amount of the data should be large. This framework has a 'maintain data module' to keep the amount of the recorded data in a certain range. It accesses the database every T_m seconds,and reduces the amount of data to N_{min} when it is larger than N_{max} according to some rules(merging, feature extraction or simply deleting is OK).

E. Workload Predictor Module

The workload predictor module consists of 3 sub modules:

Display module, this is the only module that the GUI is required. It read the database every T_{pd} seconds. Firstly, the module merges the GET,HEAD,OPTIONS into 'read request' while merging the PUT,DELETE,POST into 'write request'. Because actually the amount of HTTP request besides GET and POST is little (In practice, the HTTP requests of the later simulation`s 98 world cup workload are almost GET) and those requests have similar features. Secondly, the module smoothes and interpolates the data, generating a series of data with a custom time interval T_{pi} (seconds).Because the recorded data may not have strictly same recording time interval although the tasks are triggered every same time interval and the forecasting step T_{pi} should be custom. Finally display this series of data in a diagram on the GUI.

Knowledge Base Engine. Firstly we need to add a label for every application in cloud to represent its feature. There is a file (Knowledge Base) storing the label and the relevant accuracy of all the forecasting algorithm. The Knowledge Base Engine will read the Knowledge Base when the workload predictor starts. It provides the most accurate algorithm directly to the cloud-provider and modify the Knowledge Base content according to the workload predictor at run time.

The forecasting algorithm interface. This module provides a unified forecasting algorithm interface. If two algorithms are basically the same method with different coefficients, they will be treated as different algorithms. The input of the interface is a series of workloads data(vector (G,P,C,M,D) , representing the 'read requests', the 'write requests', the CPU utilization, the Memory utilization, the Disk utilization separately.). The output of the interface is onepredicted workload(also a vector (G,P,C,M,D)) of next step. The real algorithms are dynamically loaded. They can be easily added or removed.

On the whole, this module do the following things every Tpd seconds: reads the latest workload in the database,merges, filters and smoothes the data,predicts the future workload according to the algorithm that the knowledge base provides,displays those historical data and predicted data on the GUI,sends the predicted workload to the 'Resources Allocator' module, which will decide if the resources of the application need to be re-allocated.

F. Resources Allocator

This module does not belong to the forecasting framework. It is provided by the cloud provider. The module can dynamically allocate resources to the applications in cloud. In this framework, the VM monitor needs to access the 'VM-app mapping rule' which is generated by Resources Allocator.

To specify the whole framework, Fig. 2 displays the workflows of request and workload data.

The left 'Request workflow' rectangle covers the workflow of user`s requests which is the normal situation of cloud. When user sends a request, the request will be first dispatched to server according to the cloud`s VM-app mapping rules. Then the server processes the request and sends response. The response will be dispatched back to the original user. Dispatcher logs the request during the processes. Then the request`s workflow ends.

The right 'Workload data workflow' rectangle covers the workload data workflow which is included in the forecasting framework. The flow is triggered at constant time. We first refer to the monitor module to get the requests and resources workload. Then merge and pre-process the two workloads and write them into database. When we want to use the data, the data will be sent to 'workload predictor' module. We process the data and predict based on the data. The prediction algorithm is related to knowledge base`s recommendation. At last we display the processed workload and predicted workload on the GUI. Then the workload data workflow ends.

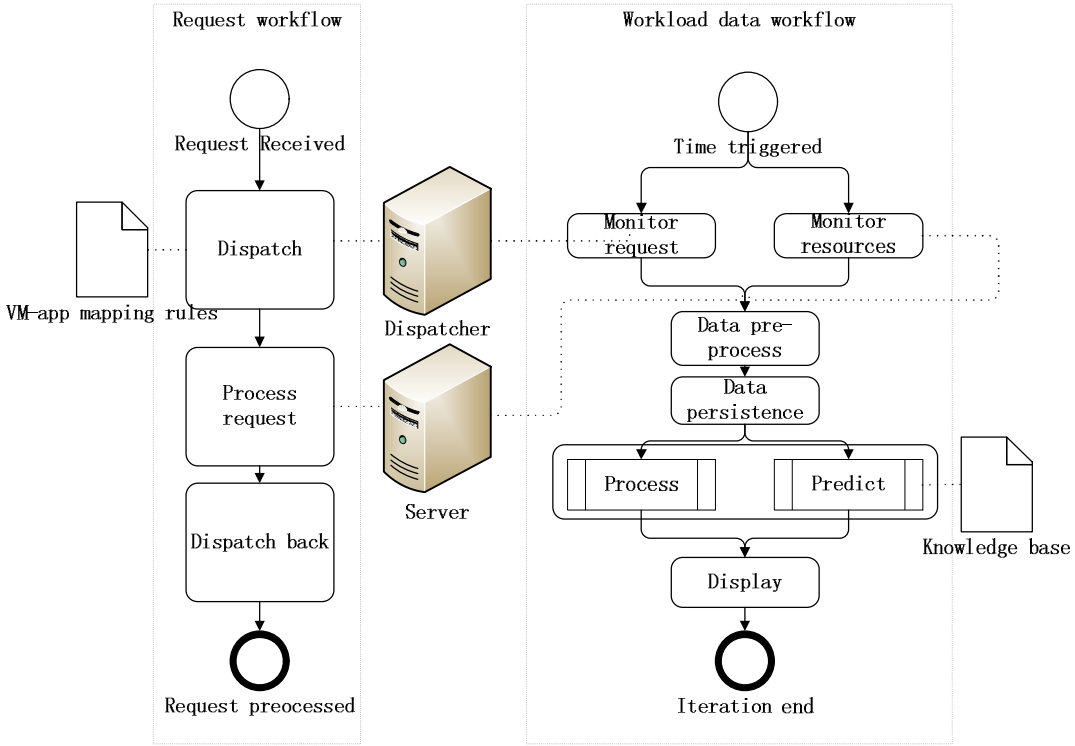


Fig. 2. Workflow of the forecasting framework

V. KEY TECHNICS

Besides the described framework, there are some other technics we use.

A. Big-data Oriented Database

In this framework, we need to handle big data. There are many kinds of workloads, and the amount of workload grows

very fastas time goes. For this reason, we need a database with high performance. Traditional relational database is not recommended. It would be better using distributed and non-relational database.Distributed database can store the big data if the cloud is quite large. In some researches, non-relational database has better efficiency than relational database under big data.

For example, if we choose MongoDBas our database. One record here means a document. Every document should have same content as following:

```
{
      id: 1363 ,
      read:{
                GET: 0.153682 ,
                HEAD: 0 ,
                OPTIONS: 0
      }
      write:{
                POST: 0 ,
                PUT: 0 ,
                DELETE:    0
      }
      CPU: 34.166667 ,
      MEMORY: 917485.333333 ,
      DISK: 47 ,
      time: 1413744115.633 ,
      interval: 3000
}
```

B. Request Intensity

During the research, we found that it is not suitable to just count the number of the HTTP requests as the workload. Because different HTTP request takes different time to deal with. See the following figure.

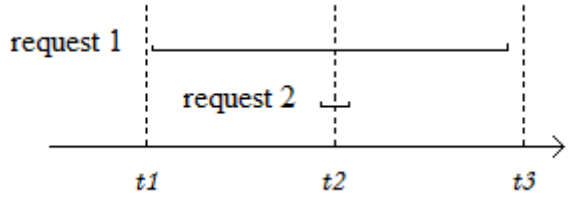


Fig. 3. Requestexample

Assume we have 2 requests during t1 and t3. And it takes different time to handle them. If we just count the number of requests that we are processing at the given time. We will get Req(t_1) = 0, Req(t_2) = 2, Req(t_3) = 0. The two requests have the same statistical effect. But the two requests are obviously different.

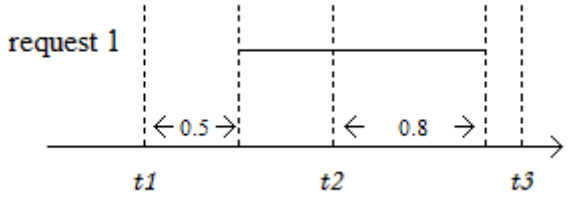


Fig. 4. Request intensity example

Therefore, 'request intensity' is introduced. The 'request intensity' of one request is defined as 'the percentage of time this request takes during the time interval'. For example, in Fig.4, the time interval is 1s. The request starts 0.5 seconds after t_1 and ends 0.8 seconds after t_2. We have the 'request intensity' as Reqi(t_1) = 0, Reqi(t_2) = 0.5, Reqi($t3$) = 0.8.

In the following simulation, we use 'request intensity' to represent the HTTP requests.

C. Maintain Data Module

In practice, in the 'Maintain data module' the N_{max} would better be larger than N_{min}. If they are equal, this module will be triggered every time. While reducing only a little data which is produced during the latest time interval. This is quite low-efficiency.

The reducing method depends on the business. You can either merge the older data into smaller and representative data or just removing the older data. In our simulation, the 98 world cup workload has little seasonality. Thus, the older data may have little effect on the predicted value. We choose to remove them in this module.

For the reason that we may frequently put records in database while deleting the redundant ones, it is better using higher Isolation Levels. We can use a lock for every record instead of one lock for the whole table.

D. Algorithm Interfaces

The forecasting framework does not contain specific algorithm. But we have somealgorithm interfaces (they are discussed later). When we start the workload predictor module, it will read a configuration file. This file will assign the specific algorithms we will use as dynamic link library. (Also, the configuration file can assign all the arguments in the framework.)

In the simulation, we implementedalgorithm interfaces using C++. We would like to introduce them in C++.

1. Workload data, we use a struct workload to provide the capsulation of all kinds of workload. The member ‘data_length’, ‘read’, ‘write’, ‘CPU’, ‘MEMORY’, ‘DISK’, ‘time_interval’represent the amount of workload, the data position of read requests, write requests, CPU utilization, MEMORY utilization, DISK utilization, the time interval between two adjacent data separately. Here is the struct:

```
structWorkload{
          intdata_length;
          double *read;
          double *write;
          double *CPU;
          double *MEMORY;
          double *DISK;
          inttime_interval;
}
```

2. Initialize interface, many algorithm should first use a series of data to build its model, and this operation always takes a lot of time. The input should be a series of historical workload, and output shows if the operation succeeds. Here is the interface:

```
int initialize(const Workload &workload);
```

3. Predict interface, obviously we should have a predict interface to get the each kind of predicted workload. The input is a series of the latest workload. The output is the value of predicted workload. Here are the interfaces:

```
doubleget_predicted_read(const Workload &workload);
doubleget_predicted_write(const Workload &workload);
doubleget_predicted_CPU(const Workload &workload);
doubleget_predicted_MEMORY(const Workload &workload);
doubleget_predicted_DISK(const Workload &workload);
```

VI. SIMULATION

A. Simulate the Request

Although we can use the web browser to send requests directly. It can`t simulate large amount of requests and can`t represent the real users` requests. Here we use the '98 world cup web servers` access log' [20] to simulate real requests. The amount of the access log is huge. We resolved the data format and generate a new simpler file named '98 World Cup Workload'. The file records the request in vector (*T,M,S*) representing the request time, the request method, the requested data length. We developed a tool 'Request Simulator' to send requests with just the same HTTP methods at the same time,while the server will read a file and response a random string in the file with exactly the same lengths.

B. Setup

The figure shows the simulation deployment of this framework. According to the real situation, some modules will be deployed in one physic machine.

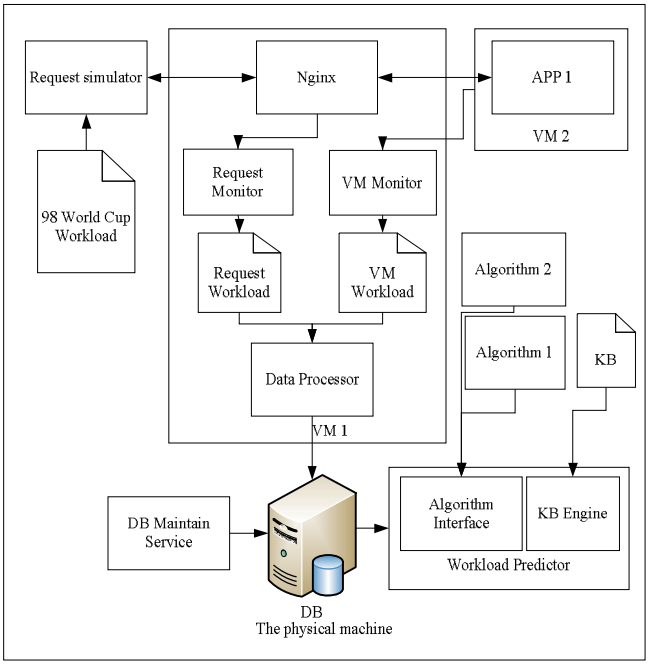


Fig. 5. Simulation deployment

We used one physic machine and two VMs in total. The following is the details classified by machines/VM.

On the physic machine, we run:

a) Database service.

b) Maintain data module, the T_m should neither be too big or too small. If big, the amount of data grows fast and we maintain it slowly. If small, we maintain it frequently, which is low-efficiency. Thus, we set T_m=60s. And we set N_{max}=4000 ,N_{min}=3500. Thus, we can keep about 3 hours` data (we record data every 3s), which is a suitable amount.

c) Request simulator, the simulator reads a file named '98 World Cup Workload' every 1 second. And sends the same requests as those logged in the file during this 1 second (because the original access log records data every 1s). According to the real situation, we just send half amount of requests due to the machine performance.

d) Workload predictor, set T_{pd}=1s, thus we can observe predicted workload quickly. We simply use two forecasting algorithms. Algorithm 1 just predicts that the next workload will be the same as last workload. Algorithm 2 is a simple auto-regressive method.

On VM1, we run:

a) Nginx, a light-weighted efficient reverse proxy server. It plays the role of 'dispatcher'. According to the VM-app mapping rule, it can send the requests of application to the real servers, and handle the response.

b) Request monitor. In the simulation, we implement it as an Nginx plug-in. Using the log system, it records the request time, request method, the time request takes into file 'Request Workload'.

c) VM monitor，set T_{mv}=1s, T_{mw}=1s (because we want to observe the predicted workload every 1s, it is better monitoring data every 1s), and record the result into file 'VM Workload'.

d) Data processor,

- In 'data merge' task, set T_{dm}=1s. The queue has a length of 64(because we display 180 points in simulation, we need at least 60 points of original data).
- In 'data filter' task, setT_{df}=3s(filter normally takes more points to output one point, thus we set a little more time here). The filter simply checks if data is available(eg. CPU utilization can`t be higher than 100%).
- In 'data accelerator' task, set T_{da}=3s for the same reason as above. The accelerator generates the mean value of 3 sequential workloads.
- In 'data record' task, setT_{dr}=3s also for the same reason as above.

On VM2, we run:

This VM runs only the application server. The application responds by opening a file and sending a random string in the file. The length of string is decided by the parameter in the request.

C. Result

For the simulation, we counted the whole requests of all the 92 days of 98 world cup. The requests diagram is shown in Fig.6. We can see the requests reach a high level on some specific days (because those days have matches) and keep a low level on other days. Here we scaled the amount of requests in the scale [0, 160].

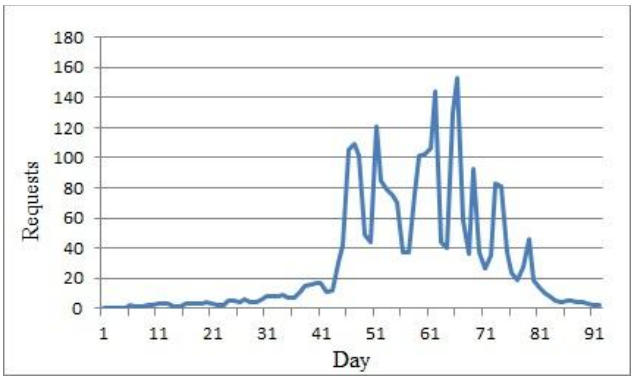


Fig. 6. Simulation-Original Requests

Then we started the forecasting framework. We can observe the workload changing as shown in the following 2 figures.

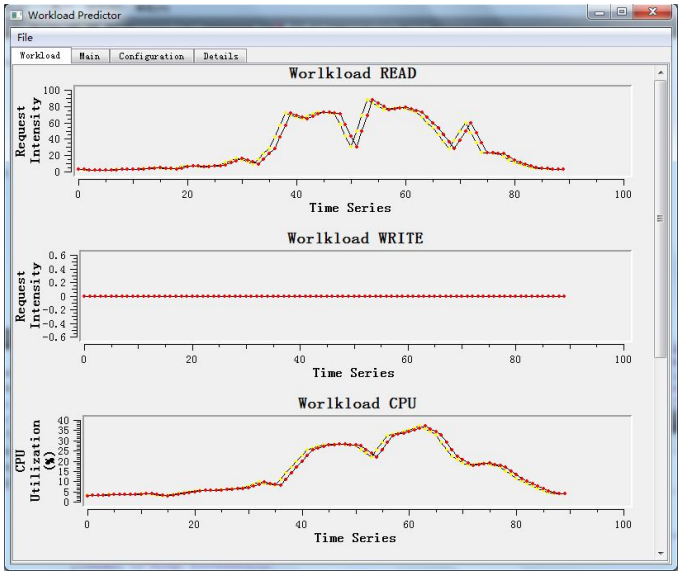


Fig. 7. Simulation-Predicted workload(Part A)

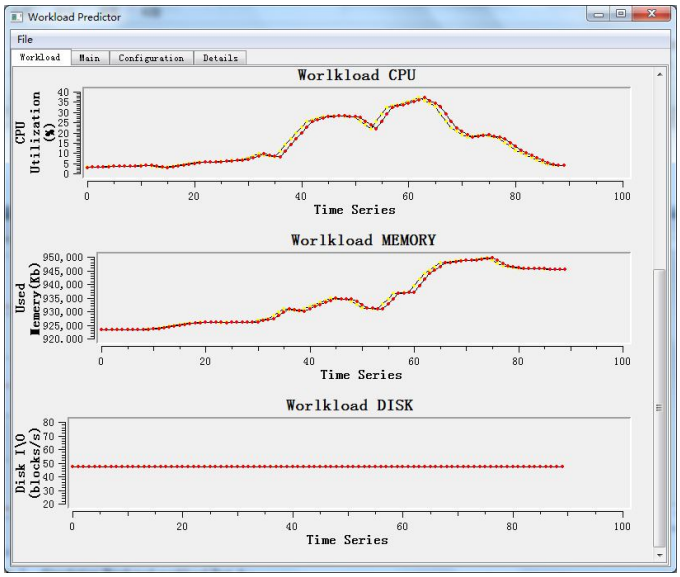


Fig. 8. Simulation-Predicted workload(Part B)

In the figures, the red lines show the real workload while the yellow lines show the predicted workload according to the algorithm that the KB recommended. We can see that the Workload READ and Workload CPU are high related to the requests. And workload MREMORY may have some relationship with the requests. Workload WRITE keeps 0 during the period of time (because we only sent the GET requests for this simulation). All the predicted workloads show some degree of accuracy towards the real workload.

We can also go to details page (Fig.9) to manually check the accuracy of every algorithm. We can see the algorithm 1 reaches accuracy of 89.7% , higher than the algorithm 2`s 45.3%, which is in accordance with our KB`s recommendation. From this result we can see there is no ‘silver bullet’ algorithm. Different algorithms may have different accuracies according to different applications.

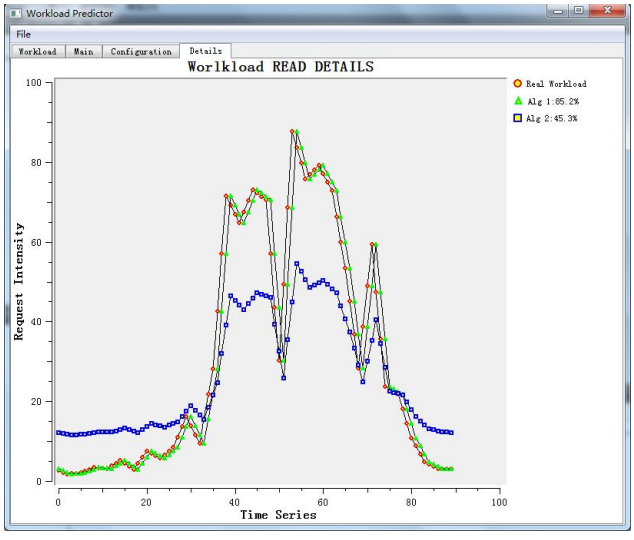


Fig. 9. Simulation-Details

We also check the CPU utilizations of our framework.

VMmonitor takes CPU: 0.3%.

Request monitor, this is a plug-in of Nginx, we can`t explore directly its CPU utilization. But we find that Nginx takes CPU: 15% both with and without the plug-in. This indicates that the request monitortakes little CPU utilization.

Workload predictor takes CPU: 6%.

Others take CPU: 2%.

And the modules besides monitor can be deployed outside the cloud. Therefore, we can say the framework is application-independent and takes little resources in cloud.

The CPU utilization and other value may have fluctuation during different simulation. But they indicated same results:

This 'Workload Forecasting Framework for Applications in Cloud':

1. Can monitor the real time amount of workload.
2. Can process and record the big workload data.
3. Can display the workload and the predicted workload from multiple algorithms while showing their accuracy.
4. Can automatically choose algorithm by knowledge base.
5. Is application-independent and light-weighted.

VII. CONCLUSION

We have proposed a workload forecasting framework for applications in cloud. After simulation, we can see the framework works the same way as we planned: it monitors workload in real time, processes and records the data, automatically predicts workload based on the algorithm recommended by knowledge base, while affecting little on the original application and cloud. However, in this article, we do not discuss the Knowledge Base`s details, and just implement a simple one. In the future, we can study how to build a full-featured Knowledge Base and set suitable label for application.

REFERENCES

[1] I. Foster, C. Kesselman, S. Tuecke, The anatomy of the grid: enabling scalable virtual organizations, International Journal Supercomputer Applications 15 (3) (2001).

[2] W. Leinberger, V. Kumar, Information power grid: the new frontier in parallel computing? IEEE Concurrency 7 (4) (1999) 75–84.

[3] Barham P, Dragovic B, Fraser K, et al. Xen and the art of virtualization[J]. ACM SIGOPS Operating Systems Review, 2003, 37(5): 164-177.

[4] Ross J W, Westerman G. Preparing for utility computing: The role of IT architecture and relationship management[J]. IBM systems journal, 2004, 43(1): 5-19.

[5] Thain D, Tannenbaum T, Livny M. Distributed computing in practice: The Condor experience[J]. Concurrency and Computation: Practice and Experience, 2005, 17(2‐4): 323-356.

[6] Mell P, Grance T. The NIST definition of cloud computing[J]. 2011.

[7] Randles M, Lamb D, Taleb-Bendiab A. A comparative study into distributed load balancing algorithms for cloud computing[C]//Advanced Information Networking and Applications Workshops (WAINA), 2010 IEEE 24th International Conference on. IEEE, 2010: 551-556.

[8] Endo P T, de Almeida Palhares A V, Pereira N N, et al. Resource allocation for distributed cloud: concepts and research challenges[J]. Network, IEEE, 2011, 25(4): 42-46.

[9] Roy N, Dubey A, Gokhale A. Efficient autoscaling in the cloud using predictive models for workload forecasting[C]//Cloud Computing (CLOUD), 2011 IEEE International Conference on. IEEE, 2011: 500-507.

[10] Schwartz B. Search memories: live from SES San Jose. SearchEngineWatch Retrieved 17 March, 2006[J]. 2004.

[11] Shirazi B A, Kavi K M, Hurson A R. Scheduling and load balancing in parallel and distributed systems[M]. IEEE Computer Society Press, 1995.

[12] B. Urgaonkar, P. Shenoy, A. Chandra, P. Goyal, and T. Wood,"Agile dynamic provisioning of multi-tier internet applications,"ACM Trans. Auton. Adapt. Syst., vol. 3, no. 1, pp.1–39, 2008.

[13] T. Wood, P. J. Shenoy, A. Venkataramani, and M. S. Yousif,"Black-box and gray-box strategies for virtual machine migration,"in NSDI, 2007.

[14] S. Cunha, J. M. Almeida, V. Almeida, and M. Santos,"Self-adaptive capacity management for multi-tier virtualizedenvironments," in Integrated Network Management, 2007, pp.129–138.

[15] P. Padala, K. Shin, X. Zhu, M. Uysal, Z. Wang, S. Singhal,A. Merchant, and K. Salem, "Adaptive control of virtualizedresources in utility computing environments," ACM SIGOPSOperating Systems Review, vol. 41, no. 3, p. 302, 2007.

[16] Y. Wang, X. Wang, M. Chen, and X. Zhu, "Power-efficientresponse time guarantees for virtualized enterprise servers,"in RTSS '08: Proceedings of the 2008 Real-Time SystemsSymposium. Washington, DC, USA: IEEE Computer Society,2008, pp. 303–312.

[17] R. Moreno-Vozmediano, R. S. Montero, and I. M. Llorente,"Elastic management of cluster-based services in the cloud,"in ACDC '09: Proceedings of the 1st workshop on Automatedcontrol for datacenters and clouds. New York, NY, USA:ACM, 2009, pp. 19–24.

[18] Wu Y, Hwang K, Yuan Y, et al. Adaptive workload prediction of grid performance in confidence windows[J]. Parallel and Distributed Systems, IEEE Transactions on, 2010, 21(7): 925-938.

[19] Dinda P A, O'Hallaron D R. An evaluation of linear models for host load prediction[C]//High Performance Distributed Computing, 1999. Proceedings. The Eighth International Symposium on. IEEE, 1999: 87-96.

[20] ftp://ita.ee.lbl.gov/html/contrib/WorldCup.html

[21] Y. Yuan, Y. Wu, G. Yang, and W. Zheng, "Adaptive Hydrid Model for Long-Term Load Prediction in Computational Grid," Proc. IEEE Eighth Int'l Conf. Cluster Computing and Grid (CCGrid '08), pp. 340-347, May 2008.

[22] Z. Xu and K. Hwang, "Early Prediction of MPP Performance: SP2, T3D, and Paragon Experiences," J. Parallel Computing, vol. 22, no. 7, pp. 917-942, Oct. 1996.

[23] L. Yang, I. Foster, and J.M. Schopf, "Homeostatic and Tendency-Based CPU Load Predictions," Proc. Int'l Parallel and Distributed Processing Symp. (IPDPS '03), pp. 42-50, 2003.

[24] L. Yang, X. Ma, and F. Muller, "Cross-Platform Performance Prediction of Parallel Applications Using Partial Execution," Proc. Supercomputing Conf., 2005.

[25] Ganapathi A, Chen Y, Fox A, et al.. Statistics-drivenworkload modeling for the cloud[C]. Proceedings ofWorkshop on Self-Managing Database Systems (SMDB2010), California, 2010: 87-92.

[26] Wang Meng, Meng Xiao-qiao, and Zhang Li. Consolidatingvirtual machines with dynamic bandwidth demand in datacenters[C]. Proceedings of the 30th IEEE InternationalConference on Computer Communications (INFOCOM2011), Shanghai, 2010: 71-75.

[27] Beloglazov A and Buyya R. Adaptive threshold-basedapproach for energy-efficient consolidation of virtualmachines in cloud data centers[C]. Proceedings of the 8thInternational Workshop on Middleware for Grids, Cloudsand e-Science (MGC10), New York, 2010: 4:1-4:6.

[28] Beloglazov A and Buyya R. Optimal online deterministicalgorithms and adaptive heuristics for energy andperformance efficient dynamic consolidation of virtualmachines in cloud data centers[J]. Concurrency andComputation: Practice and Experience (CCPE), 2012, 24(13):1397-1420.

Cloud government- a proposed solution to better serve the nation

Adeel Akbar Memon[1], Chengliang Wang[1]
1. School of Software Engineering
Chongqing University
Chongqing, P.R. China
adeelitsme@hotmail.com

Muhammad Rashid Naeem[1], Muhammad Aamir[1],
Muhammad Ayoob[2]
2. Sindh Medical College
Dow University of Health Sciences
Karachi, Pakistan

***Abstract*— The main responsibility and duty of government employees is to serve the nation. We have proposed a combination of cloud computing solution and data mining to the third world countries to better serve the nations. We have focused data/statistics gathered from Pakistan as exemplary data. There are four provincial governments of Pakistan headed by the Chief Ministers and ministries/departments are headed by the ministers/department heads. Having the up to date and accurate information on finger tips is headache for the stakeholders as the existing system is manual (Paper) work. The main objective of the proposed solution is to provide the up to date and accurate information (knowledge) to the stakeholders on their finger tips providing them a dashboard displaying the information with the drill down/up facility to get information of the own will. The proposed solution is intended to be used by the governments of third world countries to better serve the nations. However; the proposed solution is quite in general and can be easily adopted by the private organizations currently using individual information systems at different locations.**

Keywords*— Cloud computing; Data mining; Government; Information system

I. INTRODUCTION

The constant development in the field of information technology is making the computer infrastructure much more powerful and less expensive day by day. This development has increased the number of computer and internet users. Today, almost every person related with the field of information technology (IT) is familiar with the new trend named as Cloud Computing. Although the rest of the educated population doesn't know about the trend but indirectly using it for every day work like OneDrive, LinkedIn, Dropbox and Cloud call etc.

Most of the work is carried out manually (paper work) in third world countries. However; some portion of the work is carried out using computers. Data security [1] always being a big concern and due to the data security risks, independent information systems are developed.

A huge number of researchers is focusing on this new trend named as cloud computing [2]. In [3] the efficient and effective solution is given to solve the challenges of E-Government. Cloud computing is helping to shift from E-government to C-government [4].

Our research paper is organized as following: In Section II we explain/go through the problematic and major issues of existing information system and inspiration of our research work. Section III is about the background of Information System, Cloud Computing and Data Mining. A new combined solution of cloud computing and data mining is proposed in Section IV. In last, we have concluded our research work in Section V.

II. MOTIVATION

As described above, the independent ministries/departments have developed their standalone information systems. These systems are incapable to provide the up to date information to the stakeholders as there is lack of the data sharing. As a whole the stakeholder (Chief Minister) is unable to have over all view of the up to date information. Some problematic issues of the current system are defined below.

A. Data Sharing

Taking example of the ministry of health in consideration, the minister of health is unknown to the number of diseases found among the population. He is unable to have view of the useful information like number of deaths per day as the government hospitals in different districts or locations are not sharing the data among each other, or there isn't any single information system where this data can be gathered. Lackness of data sharing is illustrated in figure below.

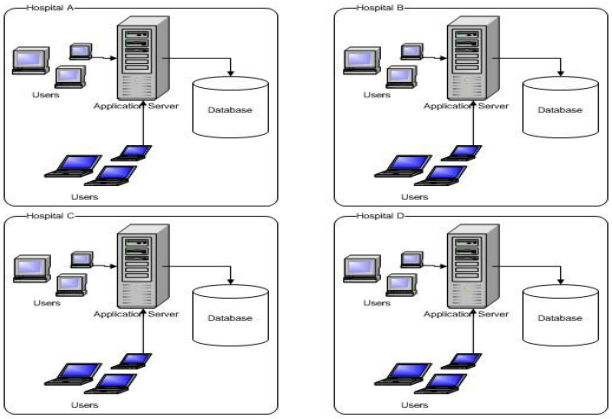


Fig. 1. Standalone HIS Systems

In [5] a cloud computing solution is proposed for the government hospitals to overcome data sharing problem currently in hand; hence patients will be better treated. Moreover cloud computing in medical aspects work is carried out in [6,7].

B. Management and Maintenance

It is well known that setting up and developing an information system has never been big concern, however managing and maintaining the information system is big issue. Standalone information systems are hard to manage and maintain as the process requires continuous investment. As the user get into the information system, user's needs/requirements change with time to time and it's hard to upgrade individual information systems. Different software maintenance types are defined in [8].

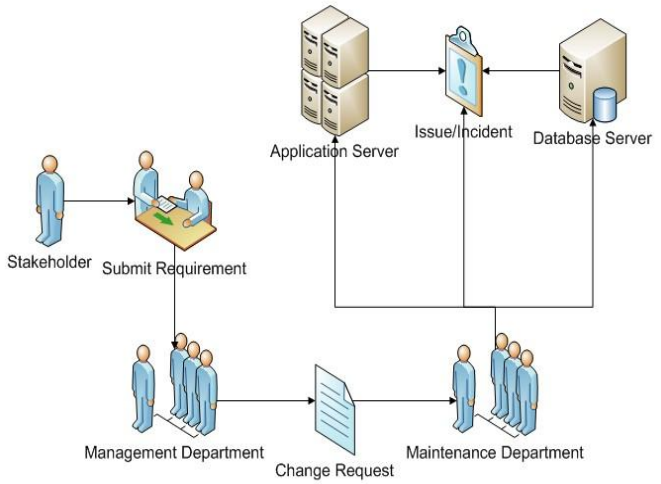


Fig. 2. Maintenance and Management

C. High Cost and Wastage of Resources

Setting up, managing and maintaining standalone information systems require high costs, including the infrastructure. Every organization is very much focused on spending the less amount of money. The individual Information system requires app/web server and Database server. In contrast to mention above individual information system requires separate network administrators to maintain the network, separate database administrators for database and server administrators for maintaining the app/web server.

III. BACKGROUND

A. Information System

There is variety of definitions for information system but the easy to understand and simple definition of information system will be a set of different components used to gather/collect, store and process data to deliver information or knowledge to the stakeholders. According to the type of information delivered to the different type of stakeholders, information systems are categorized as:

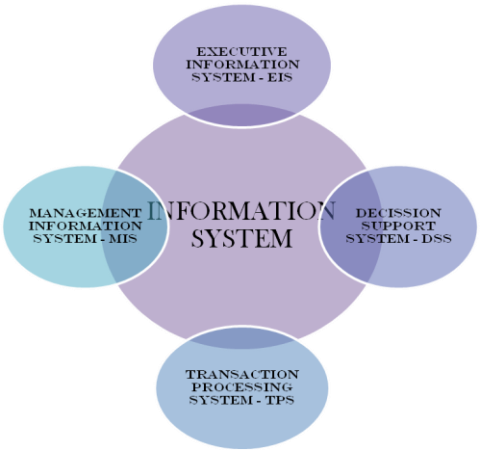


Fig. 3. Information Systems

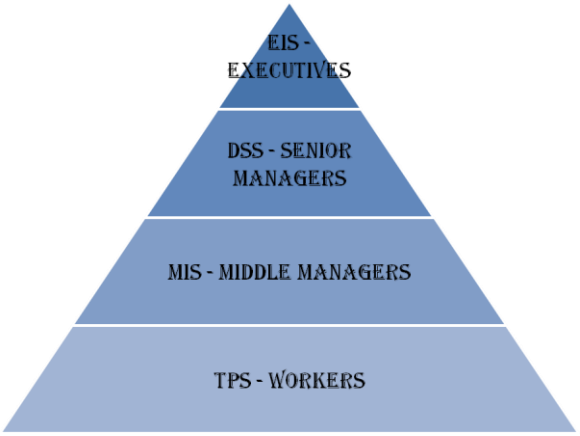


Fig. 4. User pyramid for Information Systems

B. Cloud Computing

Cloud Computing is the most focusing area for the information technology (IT) researchers since few years and will remain the favorite area ahead for few years until it get mature. Cloud computing is defined as a network (high speed internet) based model, in which platforms, infrastructure and software sold as a service. The main objective of cloud computing is to save organizations from the cost burden by reducing the infrastructure cost.

Depending on the definition described above, cloud computing has 3 types as illustrated below.

Fig. 5. Types of Cloud Computing

Cloud computing is being promoted by many of the giants of the information technology industry like Google, Microsoft, Apple, Amazon, Cisco and others. Cloud computing has come out like a high wave in the ocean of Information Technology; that's the main reason for most of software companies to move towards cloud computing to sail their boats in the ocean. Software engineers are working on developing open source cloud computing tools as in [9]. A little extended work is done in [10-12].

C. Data Mining

The word data mining itself represents the meaning as mining/extracting the data. The appropriate definition will be the process of discovering data patterns from a large dataset. The huge databases of big enterprises contain tons of data; however from these tons of data only some kilograms of data are useful for the upper level management. The goal of data mining is to dig out the useful information (called as knowledge) from dataset and transform (using different transformation techniques available) it into understandable structure to represent to the stakeholders.

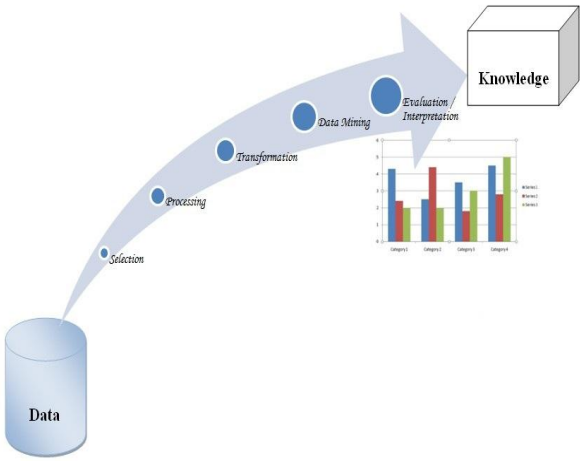


Fig. 6. Data Mining

The Data mining is wide area having educational data mining as sub area. Most of the work is carried out these days in EDM and a simple example is given in [13].

IV. Proposed Solution

A. Cloud Computing Solution

To overcome the above stated problems of the current system we proposed a hybrid cloud computing solution. Hybrid solution for each ministry/department is based on the private cloud of the ministry/department which facilitates the different offices located at different locations to manage the important information of department. In contrast public cloud of the ministry/department is capable to handle the normal management tasks and also facilitates the outside users. In order to focus the security risks the offices at different locations can access the private cloud via virtual private network and the communication link between private and public cloud is internet. The outside users can only access the public cloud through internet.

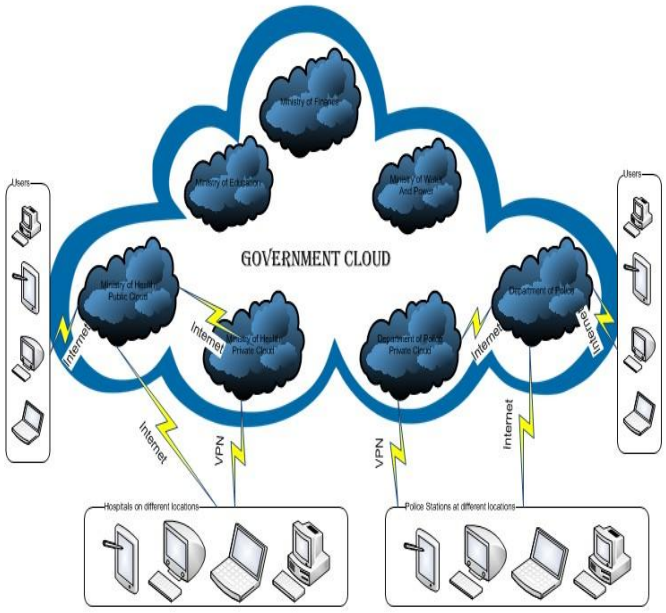


Fig. 7. Cloud Computing Solution

B. Data Mining Solution

After setting up the cloud government, the stakeholders (ministers/department heads) can be facilitated with the up to date information using the dashboards [14]. Using data mining technique the useful information can be extracted and can be represented in terms of bar charts, doughnuts, pie charts etc. There should be the facility to drill down and drill up the information so that the stakeholders can look up the data according to their own will.

1) Using BI Application

Displaying data in terms of charts has two main benefits, first is charts are very much attractive and second is they are easy to understand.

a) Ministry of health:

A number of various diseases are found in the population of Pakistan. The statistics gathered from the health department are given in table below:

TABLE I. Disease Found

Name of disease	Percentage Found
Acute Respiratory Infection	41
Coronary Heart Disease	27
Viral Hepatitis	7
Malaria	13
Diarrhea	12

From the statistics collected in Table I, a pie chart can be constructed using BI Application as shown in the following figure:

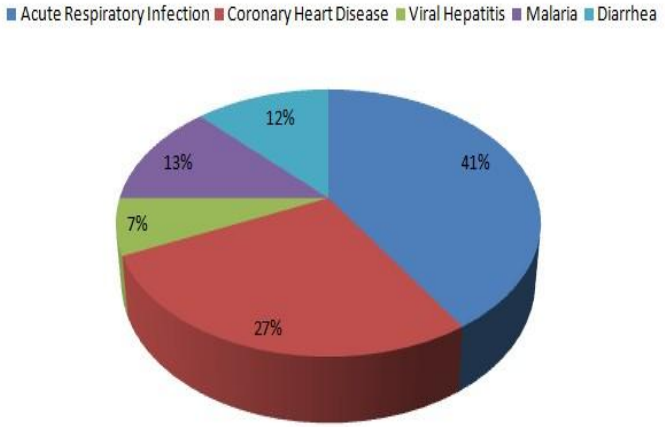


Fig. 8. Disease found in Pakistan

The pie chart given above represents the diseases found in Pakistan's population with percentage. This chart can be very useful to the minister to take in account actions against these diseases and take some actions to prevent these diseases.

The vaccination completed among children within period of time is given in table below:

TABLE II. VACCINATION COMPLETED

Time Period	Percentage Completed
1990-91 PDHS	35
2006-07 PDHS	47
2012-13 PDHS	54

The line chart drawn from Table II data using BI Application below depicts the percentage of vaccination completed among the children. The chart shows that every year there is increase in vaccination, however yet 100% is not achieved. The chart gives the meaningful information that the department must focus on the percentage of vaccination so that the innocent children will be away from diseases.

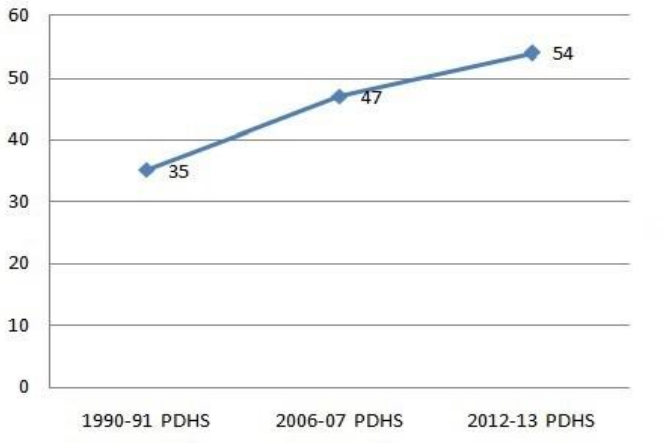


Fig. 9. Vaccination completed in Pakistan

b) Department of Police:

The crime statistics gathered for year 2011 and 2012 are given in the following table:

TABLE III. CRIME STATISTICS 2011-2012

Crime Type	Year	
	2011	*2012*
Attempted Murders	1,861	983
Murders	632	633
Sexual Assaults	779	557
Home & Bank Robberies	4,083	3,973
Non-violent Crimes	6257	6206
Vehicle Theft	7,066	6,941
Carjacking	815	923

Using BI Application bar chart constructed from the table III is displayed below.

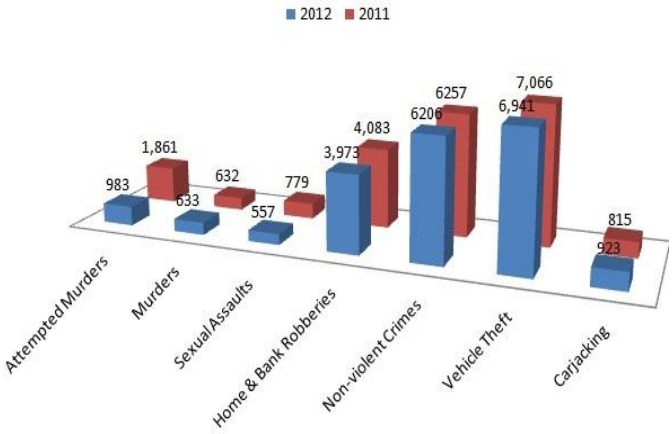


Fig. 10. Crime Statistics of Pakistan

Using the data in table III, another chart is given below:

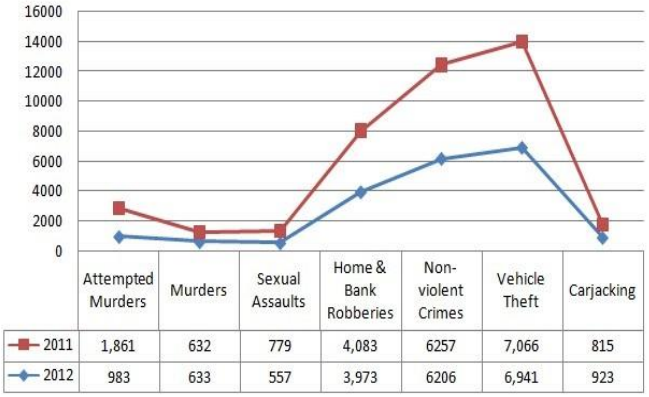


Fig. 11. Stacked line chart of Pakistan's crime statistics

The chart helps the head of department to understand that they are succeed in controlling crime in year 2012 than year 2011.

2) Using third party tools

Another way of doing data mining is using the third party software/tools. We used weka 3.7.10 as data mining tool for this paper. These tools can be directly linked with the database by passing the database URL or just by providing .csv file of the dataset. The process is illustrated in the figure below:

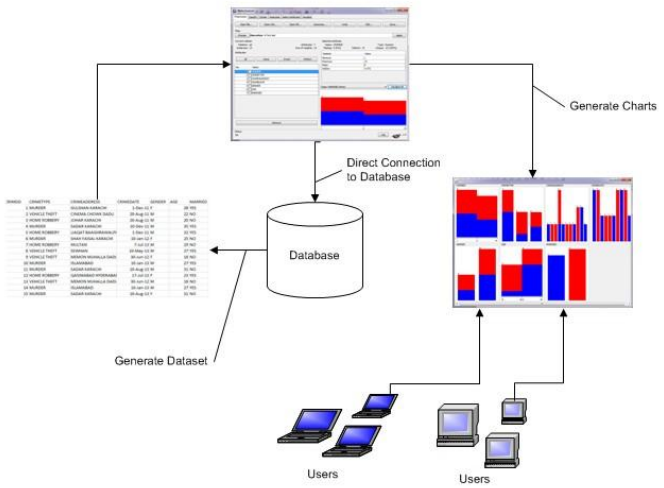


Fig. 12. Data mining using third party tool

To demonstrate the usage of third party data mining tool we took the sample data of police department containing some information about the crime and the criminal as given in following table:

TABLE IV. SAMPLE CRIME DATA

Crime ID	Crime Information			Criminal Information		
	Crime Type	*Crime Address*	*Crime Date*	*Sex*	*Age*	*Married*
1	Murder	Gulshan Karachi	1-Dec-11	F	28	YES
2	Vehicle Theft	Cinema Dadu	20-Aug-11	M	22	NO
3	Home Robbery	Johar Karachi	20-Aug-11	M	20	NO
4	Murder	Sadar Karachi	10-Dec-11	M	35	YES
5	Home Robbery	Rawalpindi	1-Dec-11	M	32	YES
6	Murder	Shah Faisal Karachi	10-Jan-12	F	25	NO
7	Home Robbery	Multan	7-Jul-13	M	29	NO
8	Vehicle Theft	Sewhan	10-May-13	M	27	YES
9	Vehicle Theft	Dadu City	30-Jun-12	F	18	NO
10	Murder	Islamabad	10-Jan-13	M	27	YES
11	Murder	Sadar Karachi	10-Aug-13	M	31	NO
12	Home Robbery	Hyderabad	17-Jul-13	F	23	YES
13	Vehicle Theft	Moro	30-Jun-12	M	18	NO
14	Murder	Islamabad	10-Jan-13	M	27	YES
15	Murder	Sadar Karachi	10-Aug-13	F	31	NO

By taking the sample dataset in table IV as .csv file, we get the following as in figure:

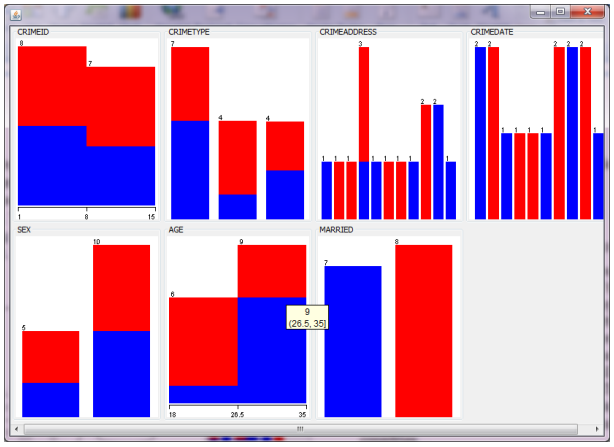


Fig. 13. Sample dataset chart

The chart contains the rich information, highlighting some of the following points.

- "Murder" is the top ranking crime type.
- Most of the crime occurred at location "Sadar Karachi".
- Criminals aging from "26.5 to 35" years are more involved in crime as marked in figure above.
- Most of the criminals are "Married".

V. CONCLUSION

We have proposed a combined cloud computing and data mining solution for the governments of third world countries to better serve their nations. The system facilitates the government stakeholders to have a better view of up to date and accurate data on finger tips. The proposed solution offers some major advantages like reduced development, management and maintenance cost, drill down/up data facility and interactive view of information. The security concern of the system can be removed by deploying hybrid cloud model. The proposed solution is for governments of third world countries.

REFERENCES

[1] M. Turle, Data security: Past, present and future. Computer Law and Security Review, 2009, pp.51-58.

[2] S. Zhang, S. Zhang; X. Chen; X. Huo, Cloud Computing Research and Development Trend, Second International Conference on Future Networks ICFN, 2010, pp. 93-97.

[3] M. A. Aziz, J. Abawajy, and M. Chowdhury, The Challenges of Cloud Technology Adoption in E-government, Advanced Computer Science Applications and Technologies (ACSAT), 2013, pp. 470-474.

[4] W. Zhang. and Q. Chen, From E-government to C-government via Cloud Computing, E-Business and E-Government (ICEE), 2010, pp. 679-682.

[5] A. A. Memon, M. R. Naeem, M. Tahir, M. Aamir, A. A. Wagan, A New Cloud Computing Solution for Government Hospitals to Better Access Patients' Medical Information, American Journal of System and Software, 2014, vol. 2, no. 3, pp. 56-59.

[6] A. Tejaswi, N. M. Kumar, G. Radhika, S. Velagapudi, Efficient use of cloud computing in medical science. American journal of computational mathematics, 2012, vol. 2, pp. 240-243.

[7] C. 0. Rolim, F. L. Koch, C. Westphall, L. Werner, A. Fracalossi, and G. S. Salvador, A cloud computing solution for patient's data collection in health care institutions, Proc. IEEE Symp. 2010 Second International Conference on Health, Telemedicine, and Social Medicine, 2010, pp. 95-99.

[8] N. Chapin, Software Maintenance Types – A Fresh View, International Conference on Software Maintenance (ICSM 2000), 2000, pp. 247–252.

[9] M. Rodriguez-Martinez, J. Seguel, M. Greer, Open Source Cloud Computing Tools: A Case Study with a Weather Application, International Conference on Cloud Computing (CLOUD), 2010, pp. 443-449.

[10] N. Sakamoto, Availablity of software services for a hospital information system, International Journal of Medical Informatics, 1998, vol. 49, pp. 89-96.

[11] A. Weiss, Computing in the Clouds. NetWorker, 2007 vol. II, pp. 16-25.

[12] W. kim, Cloud computing: Today and tomorrow, Journal of Object Technology, 2009, vol. 8, pp. 65-72.

[13] A. A. Memon, C. Wang, M. R. Naeem, M. Tahir, M. Aamir, A New Web Based Student Annual Review Information System (SARIS) With Student Success Prediction, International Journal of Computer Trends and Technology (IJCTT), 2014, vol. 10, no. 5, pp. 275-278.

[14] X. Zhang, K. Gallagher, S. Goh, BI application: Dashboards for healthcare, 17th Americas Conference on Information Systems (AMCIS), 2011, vol. 5, pp. 3898-3902.

Research on necessity of adjusting PLE configuration

Bindi Huang
School of Electronic Information and Electrical Engineering
Shanghai Jiao Tong University
Shanghai 200240, P.R. China
huangbindi@sjtu.edu.cn

Minjun Zhu
School of Electronic Information and Electrical Engineering
Shanghai Jiao Tong University
Shanghai 200240, P.R. China
zmj1989@sjtu.edu.cn

***Abstract*—Virtualization is important in server consolidation and essential for cloud computing. Using virtualization technology, a single physical machine can provide multiple isolated virtual machine to users, which will improve the overall system resources usage. However, virtualization will also introduce new challenges and mechanisms that work well in traditional operating system may lead to the performance degradation of virtual machines. LHP problem is one such problem that arises due to spin lock mechanism could not work pretty well in virtualization environment. To solve LHP problem, many solutions have been proposed to reduce its performance impact on virtual machines and improve the overall system performance. A hardware assisted technology called Pause Loop Exit(PLE) has been proposed to help detect spin lock waiter in virtual machines and then reduce the cpu time used in spinning. In this paper, we verify the necessity of adjusting PLE configuration in different scenarios. Our experiment results show that it is necessary to adjust PLE configuration when the virtual machine is running different applications and the number of virtual machines varies.**

Keywords—Virtualization;Lock Holder Preemption;Pause Loop Exit;

I. INTRODUCTION

Virtualization technology is widely used in cloud computing and server consolidation to improve overall system resource usage. Virtualization technology introduce a software layer called Virtual Machine Manager (VMM) or hypervisor [1] [2] between physical resource and guest operating system. VMM manages all the physical resources and also the upper virtual machines. In this way, virtualization technology can provide multiple isolated virtual machines to the users and improve the overall system performance and resource usage.

Although virtualization can improve system performance, new challenges have arisen due to the introduction of VMM. Spin lock is used to synchronize data between different threads and is heavily used in the kernel of modern operating system. When using mutex or semaphore to synchronize data between different threads, a thread will yield the CPU to other threads and goes to sleep when the lock is held by other thread and it can't get the lock. And the thread that holds the lock will wake up sleeping threads which are waiting for the lock when it releases the lock. Thus, the CPU resource can be fully used to do meaningful works. Unlike mutex and semaphore, the design principle of spin lock is quite different. When the lock is held by a thread in a pretty short time, the overhead of yielding will be unbearable since the context switch is too high compared with the time spent in spinning. Besides, the frequency of spin lock invoked in the kernel is pretty high and yielding is not suitable.

When the operating system runs on physical machine, the mechanism of spin lock works well since it runs on physical CPU and a CPU that holds the spin lock can release the lock in a pretty short time. But in virtualized environment, guest operating system runs on virtual CPU (vCPU) and all the vCPUs are scheduled by VMM and share the physical CPU (pCPU). When a vCPU in the VM gets the spin lock (Lock Holder) and then is scheduled out of pCPU by VMM before it can release the spin lock. Then other vCPUs in this VM will keep spinning when they attempt to get the lock (Lock Waiter) since the lock is held by the vCPU that is not running on pCPU. The problem discussed above is Lock Holder Preemption (LHP) problem and will severely degrade the performance of virtual machines.

There are already many researches focusing on reducing the impact of LHP. Traditional co-scheduling algorithm [3] [4] which schedules all the vCPUs of a VM into pCPU simultaneously was proposed to reduce the occurence of LHP. There are also several researches focusing on improving VM performance for para-virtual machine or parallel applications [5] [6] [7]. Hardware assisted technology called Pause Loop Exit (PLE) [8] [9] was also proposed to detect spin lock holder in the VM. Based on PLE technology, several lock waiter aware algorithms [10] [11] were proposed. Lock-visor [11] schedules others vCPUs of a VM into pCPU when a vCPU of this VM is detected as a lock waiter to allow the lock holder to release the lock and reduce the time spent in spinning.

The PLE technology relies on the preset parameters to approximately detect the lock waiter so the value of the preset parameters will affect the accuracy of lock waiter detection. The changes of applications running in the VMs and numbers of VMs in the VMM may require different parameters values to detect lock waiter. In this paper, we analysis PLE technology, discuss performance deviation of the VMs when PLE has different configuration. We then perform different experiments to verify the necessity of adjusting PLE configuration in different scenarios by checking the VM performance changes in different scenarios.

The contributions of this paper can be summarized as follows:

- We analyze the PLE hardware assisted technology and discuss performance deviation when PLE have different configuration.
- We performance several experiments to verify the performance deviation in different PLE configuration and different system workload.
- We make a conclusion on the necessity of adjusting PLE configuration when system workload changes.

The rest of the paper is organized as follows: Section Ⅱ presents a detailed analysis of PLE technology and the problem. Section Ⅲ presents the design of experiments and the result and analysis for each experiment. Section Ⅳ conclude the result of the experiments.

I. BACKGROUND AND PROBLEM

In this section, we first introduce the background knowledge of PLE technology and the problem. We then analysis the effect of PLE configuration when application changes in the VM and workload changes in the VMM system.

A. PLE Technology

Pause Loop Exit (PLE) is the hardware assisted technology that detects spin lock waiter in the VMs. As shown in Figure 1, the hardware in each physical CPU will keep track of the execution of PAUSE instruction and when the execution condition of the PAUSE instruction satisfy the preset condition, a Pause VM-Exit will be produced and sent to the VMM and VMM will send this kind of VM-Exit to Pause VM-Exit handler.

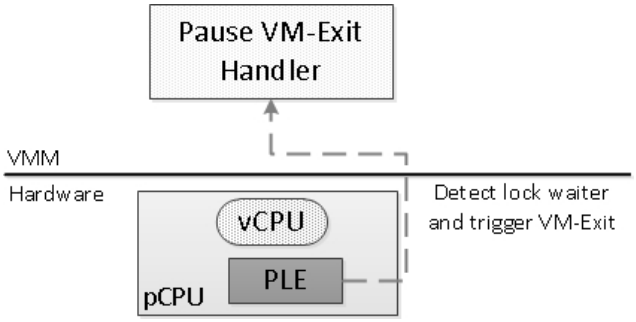


Fig. 1. PLE Technology

There are two parameters in PLE, namely PLE_GAP and PLE_WINDOW. These two parameters are used in the internal workflow of PLE. As shown in Figure 2, there is a hardware in the CPU which keep track of consecutive execution of PAUSE instruction. In the operating system, when the CPU is spinning, it will execute PAUSE instruction so consecutive execution of this instruction gives us the hint that the CPU is spinning for a spin lock and it is a spin lock waiter. It can be used to detect PAUSE instruction execution in virtual machine and detect the vCPUs that a waiting for the spin lock. However, the time between two PAUSE instructions should not be too long, otherwise these two instructions may belong to two different spin lock loop. Hence, the PLE_GAP parameter is used to define two PAUSE instructions as consecutive and the time between two PAUSE instructions should not be larger than PLE_GAP. Another parameter PLE_WINDOW is used to define the executing vCPU as lock waiter. When a vCPU spends too much time executing consecutive PAUSE instruction, namely larger than PLE_WINDOW, we can define it as lock waiter and the lock holder may not be running on pCPU since this vCPU has spent too much time spinning.

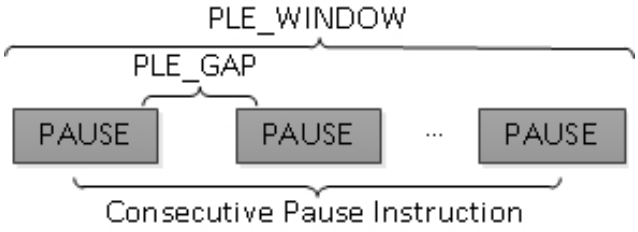


Fig. 2. PLE Internal Workflow

B. Problem Discussion

As described above, the hardware in CPU uses preset parameters to detect vCPU as spin lock waiter. The value of these two preset parameters are very important for detecting lock waiter accurately. If the value of PLE_GAP is too small, then the actual consecutive PAUSE instruction may be defined as non-consecutive and two PAUSE instruction belong to two different spin lock loop may be defined as consecutive when the value is too large. For the PLE_WINDOW parameter, the overhead of handling Pause VM-Exit will be too high when the value is too small and frequently trigger VM-Exit and the system will waste too much time spinning when the value is too large.

Even if we have choose the proper value for these two parameters, the average CPU cycles used in a spin lock varies with the workload changes in each VM and the whole system. So keeping the parameter values unchanged during the whole life cycle of a VM and applying the same values for all the VMs are not suitable and will impact the effect of detecting lock waiter in the VM.

From previous discussion, we can know the problem is that whether we should adjust the PLE configuration, namely PLE parameters, during the life cycle of a VM and apply different PLE configuration for different VM and we should also consider the performance deviation when applying different configuration. When the performance deviation is pretty small, we don't have to adjust the configuration otherwise the overhead of adjusting the configuration may be larger than the performance improvement introduced by adjusting the configuration. So it is necessary to perform experiments to check the performance deviation when applying different configuration for different applications and workload.

II. EVALUATION

In this section we perform several experiments to demonstrate the performance deviation of the VMs in different scenario using different PLE configuration.

C. Experimental Environment

In our experiments, we use a sever with 24 core Intel Xeon CPU and 64GB of RAM. The operating system is Ubuntu 13.04 and we use KVM as the VMM in our experiments. We add a module in KVM kernel module to change the PLE parameter values and the kernel version is 3.8.13. The benchmarks we use in our experiments are hackbench [12] and webbench [13].

D. Performance with different parameter value and application

To evaluate the performance of the benchmarks with PLE disabled and different PLE parameter values. We modify the linux kernel to disable PLE and also set different values for PLE parameters. We disable PLE and set PLE parameter PLE_WINDOW as 4096, 2048 and 1024. We run 4 VM with 8 vCPU and 4 GB RAM in our server and all the VM run the same benchmark at the same time. The performance of hackbench and webbench are shown in Figure 3.

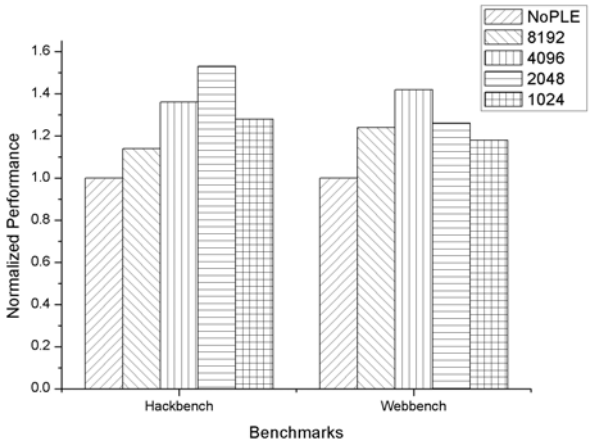


Fig. 3. Performance of Benchmarks

As shown in Figure 3, we normalize the performance of each benchmark. The performance of each benchmark with PLE disabled is used as the base and set as 1. As we can see, with PLE enabled, the performance of all the benchmarks can get at least 20% of performance improvement. For hackbench, the performance improvement is the highest when PLE_WINDOW is set as 2048 and can be up to 55%. For webbench, it gets the highest performance when PLE_WINDOW is set to 4096. From the experiment above, we can see that with different parameter value, the performance of each benchmark are different and the performance deviation can be about 30%.

From Figure 3 we can also see that the PLE_WINDOW parameter value is different in order to allow different kind of benchmarks to achieve their highest performance. The parameter value for hackbench is 2048 while for webbench it is 4096.

E. Performance with different number of VMs

To evaluate the performance deviation of the VMs when the total number of VMs in the system are different, we perform experiments using also hackbench and webbench and running different number of VMs in the server. We still use different values for PLE_WINDOW parameter and run 4 VMs and 8 VMs. The performance of different kind of benchmarks are shown in Figure 4 and Figure 5.

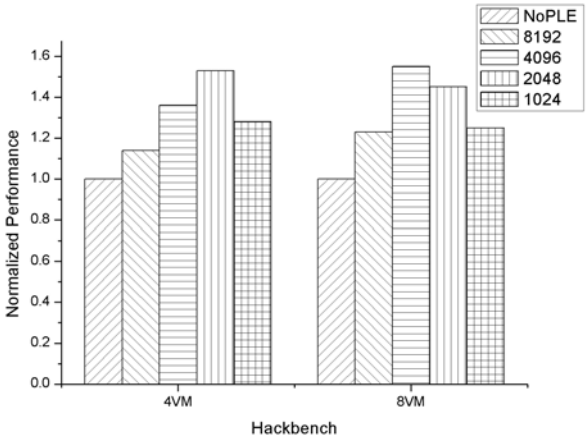


Fig. 4. Performance of Hackbench

As shown in Figure 4, it shows the normalized performance of hackbench when running 4 VM and 8 VM in different parameter values including 8192, 4096, 2048 and 1024. The overall performance improvement when running 8 VM is higher than running 4 VM. However, the best parameter value for hackbench is different when the number of VMs changes.

As shown in Figure 5, it is similar to the experiments in Figure 4. The difference is that we run webbench this time. The overall performance improvement for 8 VM is better than 4 VM. This is easy to understand since spin lock contention increase with the increase of the number of the VMs. But this time, the best parameter value for webbench is the same for both 4 VM and 8 VM.

From the two figures above, we can see that the performance deviation between different number of VMs is not as much as the performance deviation between different value of the parameters. And the best parameter value for each kind of benchmarks not only depends on the benchmark itself but also the overall system workload, such as the number of VMs in the system.

F. Summary

In this section, we perform several experiments to check the performance of the VMs in different scenarios. The factors we consider include the value of parameters, the benchmarks we run and the number of VMs in the system. By comparing the experiment results, we can see that the best parameter for a VM changes with the workload changes both in the VM and in the system.

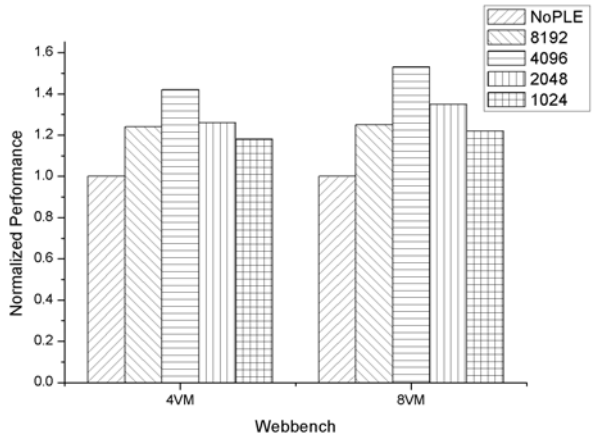


Fig. 5. Performance of Webbench

II. Conclusion

In this paper, we demonstrate the necessity of adjusting PLE configuration when workload changes in both VMs and the whole system. We first introduce and analyze the LHP problem in virtualize environment. Then we introduce and analyze the PLE hardware assisted technology which is used to detect spin lock waiter in VMs.

We discuss and analyze the importance of the PLE parameter values when detecting lock waiter in VMs and make a conclusion that adjusting values is necessary in theory. However, the performance deviation tells whether it is really necessary. So we perform several experiments to verify the performance deviation in different scenario using different PLE configuration.

From the experimental results, we conclude that the performance deviation is pretty large in different scenarios using different PLE configuration and it is necessary to adjust PLE configuration dynamically to fully improve the overall system performance.

References

[1] A. Kivity, “kvm: the Linux virtual machine monitor,” in OLS ’07: The 2007 Ottawa Linux Symposium, Jul. 2007, pp. 225–230.

[2] P. Barham, B. Dragovic, K. Fraser, S. Hand, T. Harris, A. Ho, R. Neugebauer, I. Pratt, and A. Warfield, “Xen and the art of virtualization,” in Proceedings of the nineteenth ACM symposium on Operating systems principles, ser. SOSP ’03. New York, NY, USA: ACM, 2003, pp. 164–177.

[3] J. K. Ousterhout, “Scheduling techniques for concurrebt systems.” In ICDCS, vol. 82, 1982, pp. 22–30.

[4] W. Chuliang, L. Qian, Y. Lei, and L. Minglu, “Dynamic adaptive scheduling for virtual machines,” in Proceedings of the 20th international symposium on High-Performance Parallel and Distributed Computing, ser. HPDC’11. ACM, 2011, pp. 239–250.

[5] R. Jia and Z. Xiaobo, “Towards fair and efficient smp virtual machine scheduling,” in Proceedings of the 19th ACM SIGPLAN symposium on Principles and practice of parallel programming, ser. PPoPP ’14. Orlando, FL, USA: ACM, 2014, pp. 273–286.

[6] B. Yuebin, X. Cong, and L. Zhi, “Task-aware based co-scheduling for virtual machine system,” in Proceedings of the 2010 ACM Symposium on Applied Computing, ser. SAC ’10. Sierre, Switzerland: ACM, 2010, pp. 181–188.

[7] “Drummonds. vmware, inc. co-scheduling smp vms in vmware esx server,” http://communities.vmware.com/docs/DOC-4960.

[8] “Intel,”http://www.intel.com/content/www/us/en/processors/architectur essoftware-developer-manuals.html.

[9] “Amd,”http://developer.amd.com/documentation/guides/pages/default.as px.

[10] Y. Dong, X. Zheng, X. Zhang, J. Dai, J. Li, X. Li, G. Zhai, and H. Guan, “Improving virtualization performance and scalability with advanced hardware accelerations,” in Workload Characterization (IISWC), 2010 IEEE International Symposium on. IEEE, 2010, pp. 1–10.

[11] L. Zhang, Y. Chen, Y. Dong, and C. Liu, “Lock-visor: An efcient transitory co-scheduling for mp guest,” in IEEE International Conference on Parallel Processing, ser. ICPP ’12. Pittsburgh, PA, RUS: IEEE, 2012, pp. 88–97.

[12] “ Hackbench,”http://people.redhat.com/mingo/cfsscheduler/tools/hackb ench.c.

[13] “Webbench,” http://cs.uccs.edu/~cs526/webbench/webbench.htm.

I/O-intensive Scheduling in Multiprocessor Virtualized System

Haoxiang Mao
Department of Computer Science and Engineering
Shanghai Jiao Tong University
Shanghai 200240, P.R. China
maohaoxiang123@sjtu.edu.cn

Bindi Huang
Department of Computer Science and Engineering
Shanghai Jiao Tong University
Shanghai 200240, P.R. China
huangbindi@sjtu.edu.cn

***Abstract*—Virtualization is becoming an important part of cloud computing and high performance calculating. In the virtualized system, the scheduler plays an important role in effecting the performance of whole system. Traditional scheduler that focus on the fairness of VMs would cause problems in I/O performance and latency. In order to eliminate the I/O latency issue, we propose a virtual machine scheduling model based on multiprocessor system. We also implement a prototype in Xen 4.3.0 and evaluate it with several benchmarks. Experiment results demonstrate that our scheduling model can improve the I/O performance effectively. The bandwidth of disk increase by 53.6% and that of network increase by 39.4%. Meanwhile, our method does not change the scheduling algorithm of CPU-intensive VM so that the scheduling fairness is ensured.**

Keywords—Virtualization; I/O; scheduling; Xen;

I. Introduction

High performance calculating and cloud computing are becoming more and more popular, which cause virtualization becoming an important role in modern life [1]. For example, commercial cloud providers such as Amazon EC2 [2] use virtualization to allocate different types of virtual machines on the same hardware platform for customers to meet their different demands. Virtualization technology provides the improvement of resource utilization, application portability and efficiency of system management [3].

However, there are several challenges remaining to be addressed when using virtualization in high performance computing (HPC) platform. One of them is the I/O latency issue. Generally, the I/O operations such as disk and network requests are very important in the HPC platform. In non-virtualized environment, the OS has information about all the tasks so the I/O tasks can preempt other CPU tasks to achieve a better performance [4][5][6]. While in virtualized environment, traditional virtual machine scheduler focus on the fairness of CPU time between domains, which makes I/O resources less important. So that the latency of I/O-bound applications will become longer than in non-virtualized environment.

When it comes to the cloud environment, the latency problem becomes even worse. In the cloud environment, different users often apply for virtual machines to run different types of applications. Some of them are CPU-bound and others are I/O-bound. The customers may desire a fast response for I/O-bound VM, but they may desire a well calculation ability for CPU-bound VM. Since there will many VMs in many physical cores, when more than one VMs run in one same physical core, a VM with I/O requests has to wait for its turn to access CPU to process the requests. The latency is supposed to be a multiple of default scheduling time slice (e.g. 30ms in Xen). This is harmful for I/O-bound applications.

To avoid these problems, one could pin only one VM to a physical core and this can eliminate competition forever. However, that would increase the cost which is not desired by the customers. There are also many other researches on how to improve the virtualized I/O performance. Hwanju Kim et al. proposed a task-aware virtual machine scheduling mechanism using gray-box knowledge [13]. And Govindan et al. proposed a communication-aware VM scheduling mechanism in consolidated hosting environment [14].

In this paper, we propose our solution named iSche to reduce I/O latency in multiprocessor virtualized system. We analysis and address the problem of I/O latency in Xen credit scheduler. We design and implement a prototype of virtual machine scheduler based on multiprocessor system in Xen 4.3.0. And our evaluation shows that the performance of I/O-bound VMs has been improved. We improve the bandwidth of disk 54.6% and the throughput of network 39.4%.

The rest of this paper is organized as follows. We discuss the background in Section II. Then Section III presents the design and implementation of iSche. Section IV gives the details of evaluation. And we conclude our work in the last section.

II. Background

In this section, we first introduce the Xen credit scheduler. Then we discuss the problems by a simple experiment.

A. The Xen Credit Scheduler

Xen [7] is an open source virtual machine monitor (VMM). In Xen, domain0 (the host) runs the control software and the backend drivers while domainU (the guest) runs users applications. Xen has three schedulers, including Borrowed Virtual Time (BVT) scheduler, Simple Earliest Deadline First (SEDF) scheduler and the Credit scheduler. BVT is usually

used for some delay-sensitive tasks [8]. SEDF is suitable for real-time tasks [9]. The default scheduler in Xen is the Credit scheduler [10].

The credit scheduler is a fair share CPU scheduler as the default scheduler of Xen on SMP hosts. Each domain (including the host OS) is assigned a weight and a cap. A domain with more weight will get more CPU time on a contended host. The cap limits the maximum amount of CPU time that a domain will be able to consume. When cap is set, a domain must wait for next schedule period if he consumes up all his credit even if the host system has no other domains to run.

Each physical CPU manages a run queue of runnable VCPUs. This queue is sorted by VCPUs' priority. A VCPU's priority can be **over** or **under**, which represent whether this VCPU has or hasn't exceeded its fair share of CPU resource. When the scheduler insert a VCPU into a run queue, it put the VCPU after all other VCPUs which have the equal priority. When a VCPU runs, it consumes **credits**. Negative credits imply a priority of over. A VCPU's priority becomes under if it consumes all its allocated credits. At every scheduling decision, the next VCPU to run is picked off the head of the run queue by the scheduler. When a CPU doesn't find a VCPU whose priority is under on its local run queue, it will come to other CPUs to pick one. This is the load balance which guarantees each VM receives its fair share of CPU resources system-widely [10].

B. Problems of I/O Processing Latency

In Xen (as well as other VMM such as KVM [11], VMware [12]), I/O requests are first delivered to domain0 and then the domain0 forwards them to the target VM. The target VM could process the request and send a response at the next time it is scheduled. Since the scheduling policy of credit is round robin, here we comes some problems. For instance, suppose we have 4 VMs in one physical core, as shown in Fig. 1.

When VM1 has consumed all its credit, the scheduler selects VM2 to run. When an I/O request comes for VM1, it has to be buffered until VM1 is rescheduled. Suppose we use 30ms as the default time slice, VM1 must wait for 90ms (e.g. 3*30ms) to process its requests. So the I/O latency can be reached as high as 90ms in the worst case. In fact, the worst waiting time could be even longer.

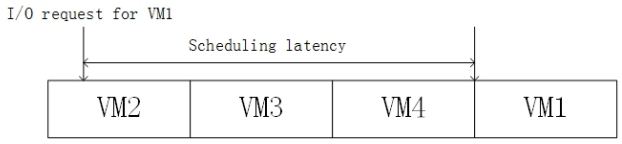


Fig. 1. Scheduling latency in Xen

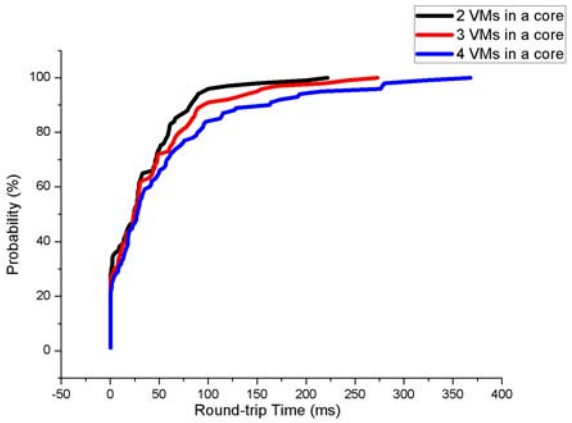


Fig. 2. CDF of ping round-trip time

We perform a simple experiment to demonstrate this problem. We vary the number of VMs sharing one physical CPU from 2 to 4, each VM running CPU loop, and calculate the CDF of the round-trip time (RTT) of VM1. The result is shown in Fig. 2. Our results show that the ping RTT increases with the number of VMs in one core. And the worst RTT in 4VMs is as much as two times of that in 2VMs.

III. Design and Implementation

In this section, we present our design and implementation of iSche. Subsection A presents the architecture of our system and subsection B and C discuss the details of implementation.

A. Design of the Scheduling Model

In virtualized environment, the requirements for CPU resource of domainUs are different when they run different applications. CPU-intensive VMs generally run continuously and consume up all its credit. On the contrary, I/O-intensive VMs need less CPU time. What they need are high-rate scheduling switch, so that they can process I/O requests in time. On the one-processor system, all of VMs are running in the same core to share CPU time. I/O-intensive VMs will be treated the same with the CPU-intensive ones. So the I/O-intensive VMs have to wait for their turn to handle the I/O requests. And it's hard to apply different scheduling strategy to only one processor. When it comes to multiprocessor system, we can apply different strategies into different processors.

In our system, we define two types of cores, General Core and I/O core. The general core is used to run domain0 and CPU-intensive VMs. While the I/O core is used for I/O-intensive VMs to improve the performance. We use different strategies on different types of cores. The general cores use the credit scheduling algorithm of Xen to fairly schedule all the CPU-intensive VMs. The I/O cores switch VMs more frequently to provide more opportunities for I/O-intensive VMs to process I/O requests. The architecture of our system is shown in Fig. 3.

As shown in Fig. 3, one model is added into VMM. The scheduling decision model is mainly used to collect all the domainUs' information and schedule them in a suitable core. At first, this model separate general cores and I/O cores according to the admin's strategy. Then we pin domain0 to a general core because it's the host and we can always treat it as a CPU-intensive VM. Thirdly, it monitor the type of all the domainUs and pin I/O-intensive VMs to I/O cores and CPU-intensive VMs to general cores. The algorithm in the I/O cores is different from that in general cores so we can treat I/O-intensive VMs differently.

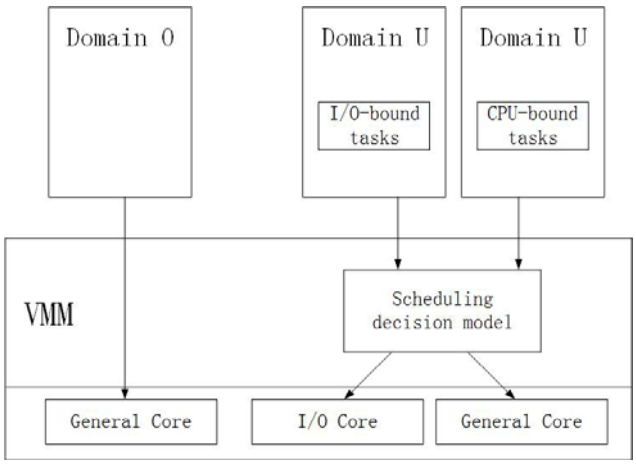


Fig. 3. Architecture of iSche

B. *Implementation of Scheduling Decision Model*

As mentioned above, the first thing the model should do is to separate general cores and I/O cores. We provide an interface to the administrative tools (such as xl in Xen) to configure the number of each core. Then this model will randomly pick several cores and label them as general cores, and label others as I/O cores. Whenever the administrator change the configure file, this model will relabel all the cores.

The scheduling decision model needs to distinguish I/O-intensive VMs from CPU-intensive VMs. We left the decision on whether a particular VM is I/O-intensive or CPU-intensive to the administrator. And the xl tools is provided. The decision model maintains a list of all the I/O-intensive VMs and CPU-intensive VMs. When the type of a VM is changed, it can easily change the list the VM belongs to. The algorithm of I/O cores is changed so that when picking next VM to run, I/O cores will pick VMs from the I/O list instead of all the VMs.

C. *Time Slice of I/O Core*

We use 10ms as the default time slice in I/O cores (the default time slice in Xen is 30ms). So the switch frequency of I/O-intensive can be three times more than before.

IV. Evaluation

We perform several experiments to evaluate our system. We evaluate both disk I/O performance and network I/O performance of iSche. And we discuss the improvement made by our system.

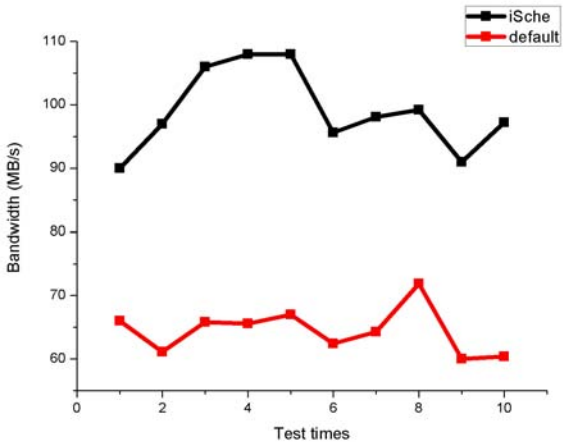


Fig. 4. Test result of disk bandwidth

A. *Experiment Setup*

The hardware configuration of our experiment is as follows: Intel Core i5-3450 quad-core CPU, 16GB memory, 100Mbps Ethernet NIC, 1TB SATA disk. Our system runs Xen 4.3.0 as virtual machine monitor. We use Ubuntu 13.04 as the operating system of the domain0 and all domainUs. The version of the Linux distribution is Linux-3.8.0 kernel.

B. *Evaluation with Disk I/O*

To evaluate the performance of disk I/O, we run 4 VMs concurrently. Three of them are CPU-intensive and one is I/O-intensive which is the one to be evaluated. Here we denote it as domain1. We firstly run CPU-bound application (here is the CPU loop) in all four VMs, and then use the dd command in domain1 to test the bandwidth of the disk. We use dd to read and write 2GB data with 64kB block size. The bandwidth test result is shown in Fig. 4.

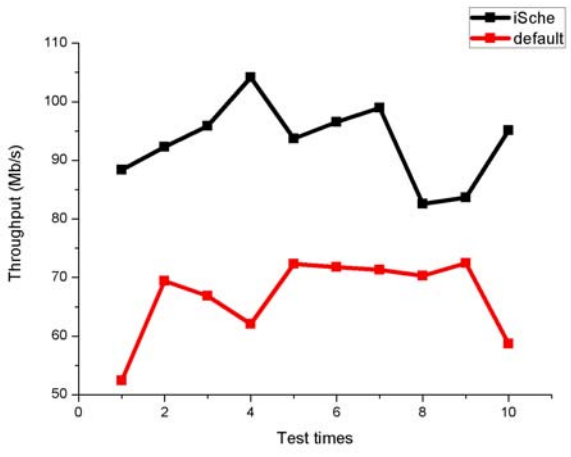


Fig. 5. Fig 5. Test result of network throughput

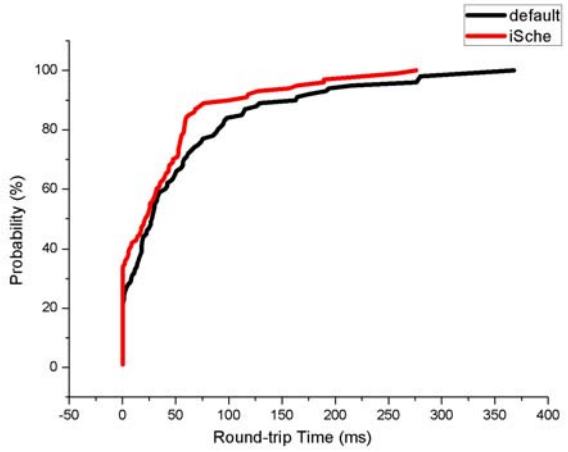


Fig. 6. CDF of ping RTT

As shown in Fig. 4, the average disk bandwidth is about 50MB/s in default case. The iSche improves the average bandwidth from 64.45MB/s to 99.01MB/s. The increasing rate is nearly 53.6%.

C. *Evaluation with Network I/O*

To evaluate the performance of network I/O, we run 4 VMs concurrently. The case situation is the same as evaluation with disk I/O. We first run the netperf [15] to evaluate the throughput of network I/O. We evaluate the TCP_STREAM between iSche and default Xen. As shown in Fig. 5, the iSche improves the average throughput from 66.77Mb/s to 93.11Mb/s with the increasing rate of 39.4%.

We then test the ping latency comparing with default Xen credit scheduler. As shown in Fig. 6, we can find that the RTT of iSche is less than default one. This is because the I/O-intensive VM is scheduled more frequently. So when network I/O events come, VM can handle these very quickly and thus reduces the latency caused by the scheduling delay.

CONCLUSION

We have presented iSche to pin different types of VM to different types of cores. Our system can reduce the latency of I/O-intensive VMs without much effect on the performance of CPU-intensive VMs. We increase the switch frequency of I/O-intensive VM to process more I/O requests. Our evaluation of a Xen-based prototype demonstrates improvement at both disk I/O and network I/O.

REFERENCES

[1] L. M. Vaquero, L. Rodero-Merino, J. Caceres, and M. Lindner, “A break in the clouds: towards a cloud definition,” SIGCOMM Comput. Commun. Rev., vol. 39, no. 1, pp. 50–55, Dec. 2008.

[2] http://aws.amazon.com/ec2.

[3] Y. Hu, X. Long, J. Zhang, J. He, and L. Xia, “I/o scheduling model of virtual machine based on multi-core dynamic partitioning,” in Proceedings of the 19th ACM International Symposium on High Performance Distributed Computing. ACM, 2010, pp. 142–154.

[4] D. Bovet and M. Cesati, Understanding The Linux Kernel. Oreilly & Associates Inc, 2005.

[5] M. K. McKusick and G. V. Neville-Neil, “Thread scheduling in freebsd 5.2,” Queue, vol. 2, no. 7, pp. 58–64, Oct. 2004. [Online]. Available: http://doi.acm.org/10.1145/1035594.1035622

[6] M. E. Russinovich and D. A. Solomon, Microsoft Windows Internals, Fourth Edition: Microsoft Windows Server(TM) 2003, Windows XP, and Windows 2000 (Pro-Developer). Redmond,WA, USA: Microsoft Press, 2004.

[7] P. Barham, B. Dragovic, K. Fraser, S. Hand, T. Harris, A. Ho, R. Neugebauer, I. Pratt, and A. Warfield, “Xen and the art of virtualization,” in Proceedings of the nineteenth ACM symposium on Operating systems principles, ser. SOSP ’03. New York, NY, USA: ACM, 2003, pp. 164–177.

[8] K. J. Duda and D. R. Cheriton. Borrowed-virtual-time (BVT) Scheduling: Supporting Latency-sensitive Threads in a General-purpose Scheduler. SIGOPS Oper. Syst. Rev., 33(5):261–276, 1999.

[9] I. M. Leslie, D. Mcauley, R. Black, T. Roscoe, P. T. Barham, D. Evers, R. Fairbairns, and E. Hyden. The Design and Implementation of an Operating System to Support Distributed Multimedia Applications . IEEE Journal of Selected Areas in Communications, 14(7), 1996.

[10] http://wiki.xen.org/wiki/Credit Scheduler.

[11] A. Kivity, “kvm: the Linux virtual machine monitor,” in OLS ’07: The 2007 Ottawa Linux Symposium, Jul. 2007, pp. 225–230.

[12] C. A. Waldspurger, “Memory resource management in vmware esx server,” in Proceedings of the 5th symposium on Operating systems design and implementation, ser. OSDI ’02. USENIX, 2002, pp. 181–194.

[13] H. Kim, H. Lim, J. Jeong, H. Jo, and J. Lee. Task-aware Virtual Machine Scheduling for I/O Performance. In Proceedings of the 5th ACM SIGPLAN/SIGOPS International Conference on Virtual Execution Environments (VEE'09), pages 101–110, New York, NY, USA, Mar. 2009. ACM.

[14] S. Govindan, A. R. Nath, A. Das, B. Urgaonkar, and A. Sivasubramaniam. Xen and Co.: Communication-aware CPU Scheduling for Consolidated Xen-based Hosting Platforms. In Proceedings of the 3th ACM SIGPLAN/SIGOPS International Conference on Virtual Execution Environments (VEE'07), pages 126–136, New York, NY, USA, June 2007. ACM.

[15] http://www.netperf.org/netperf/.

Elastic Time Slice Scheduler in Virtualized System

Minjun Zhu
Department of Computer Science and Engineering
Shanghai Jiao Tong University
Shanghai 200240, P.R. China
zmj1989@sjtu.edu.cn

Bindi Huang
Department of Computer Science and Engineering
Shanghai Jiao Tong University
Shanghai 200240, P.R. China
huangbindi@sjtu.edu.cn

Xiaolong Jia
Department of Computer Science and Engineering
Shanghai Jiao Tong University
Shanghai 200240, P.R. China
jiaxiaolong@sjtu.edu.cn

***Abstract*—Virtualization is a key technology in the modern data centers as well as in the cloud computing industry. Modern virtual machine monitors (VMM) like KVM and Xen play a key role in the management and maintenance of virtual machines (VM). However, as a software middle layer, VMM brings in a negative impact in terms of VM's performance, among which the I/O events latency problem is one of the biggest issue.**

As the virtual CPUs inside the VMs share each other with the same set of physical CPUs, the scheduling latency will actually contribute to the I/O events processing delay, which will result in a long processing waiting time and low I/O throughput. To mitigate this negative impact while still take advantage of the sharing mechanism, we first divide all VMs into two basic classes, namely the I/O-sensitive VMs and the CPU-sensitive VMs. We give out more scheduling opportunities but a shorter time slice each scheduling to the I/O-sensitive VMs, while keep the scheduling mechanism unchanged for the non-I/O-sensitive VMs to improve the I/O events responding latency of I/O-sensitive VMs. At the same time, we also take into account the computing workload in our design, as a shorter time slice will holds an impact on the CPU-sensitive tasks as well.

We also conduct detailed experiments to evaluate tour design and implement under a physical network environment. The result shows that our design achieves a significant improvement in terms of the I/O latency, while only introduces a small overhead in the original system.

Keywords—Virtualization;VMM;Scheduling;I/O Latency;VM; Xen;

I. INTRODUCTION

Cloud computing [1] has brought in a deep impact on the traditional computing industry. With virtualization as the basic and fundamental technology, modern clouds and data centers allow users to create and maintain their virtual machines (VM) by multiplexing the underlying physical resources, for example CPU, hard disk, memory and other common devices [2]. In the virtualized world, the virtual machine monitor (VMM) plays a key role in managing and monitoring the physical resources in the physical machines. Common VMMs like Xen [3], KVM [4], VMWare [5], Hyper-v [6] and Virtual-box [14] are widely used in the cloud computing and virtualization technology. With the help of VMM, a VM can be easily created with specific resources and virtual devices configured to achieve the goals of flexible and scalable resources management, less energy wasting and a better server consolidation [7] [8]. However, virtualization and VMM might bring in problems that not exist in the physical world.

As many other researches have pointed out, I/O virtualization is one of the most important functionality of the virtualization technology, while it still remains a big challenge to researchers. With the fast development of cloud computing, nowadays many popular services and applications, for example online games, video sharing websites, live meetings and phone meetings are deployed inside the cloud. More and more users and publishers hold their applications and services in the virtual machine which are in the cloud instead of common physical ones. These applications or services in the cloud could also be very sensitive to the I/O processing latency, where a good I/O latency will result in a better user experience and higher service quality. But in modern virtualization environment, the physical computing resources are shared by all virtual CPUs of the VMs in that physical machine, so each virtual CPU has to line up and wait for its turn to be scheduled. The pending I/O events to a specified virtual CPU will not be handled actually until the corresponding virtual CPU is no longer waiting but scheduled through a content switch in the VMM. As a result, the waiting time, or we called scheduling delay of the virtual CPUs will actually become a significant part of the I/O events' processing delay, which a step further brings in a terrible user experience in terms of the application aspect.

Many researchers have focused on the I/O latency problem. Research [2] mainly aims at the I/O latency problem in the Symmetric Multi-Processing (SMP) VMs. As for the SMP VMs, since there is more than one virtual CPU in a VM, the mapping relationship between the specified virtual interrupt and virtual CPUs is not fixed. In research [2], they are always trying to find an active or online virtual CPU, i.e. a virtual CPU that is not waiting but actually running in a physical CPU to handle the virtual interrupts. In order to achieve this goal, they have modified the virtual interrupt mapping mechanism inside the guest Operating Systems based on the scheduling status information of each virtual CPU. Their solution mitigated the problem to some degree, but since it requires significant modification in the guest, their solution is not flexible and adaptable enough to most commercial OSes which are close-sourced. What's more, their solution is only for SMP VMs, ignoring the fact that single processor VMs also suffer from I/O latency problem.

Research [9] comes up with a credit sharing algorithm for Xen's credit scheduler [10] to solve this I/O latency problem. In their design, each virtual CPU gives out a certain part of its credits as the shared credits. When a specified virtual CPU runs out of credits and cannot be scheduled to handle the pending I/O events, it borrows some shared credits from other virtual CPUs so that the I/O events can be handled in time. Their solution has also mitigated the I/O problem to some degree, yet it relies on Xen and the credit scheduler closely and is not adaptable to other popular VMMs and their schedulers, since other popular schedulers like the SEDF [11] scheduler and the BVT [12] scheduler do not have the concept of credit.

In this paper, we proposal Elastic timE slicE controllEr (E4) software solution to solve the I/O latency problem. We use an elastic time slice to achieve a shorter I/O processing latency in the virtual machine. In E4, the time slice is not uniform among all the VMs. For VMs that have a strong I/O request, we use a shorter time slice, while for others the length of their time slices remain the same.

The rest of this paper is organized as follows. We discuss the background in Section II. Then Section III presents the design and implementation of E4. Section IV gives the details of evaluation. And we conclude our work in the last section.

II. DESIGN AND IMPLEMENTATION

In this section, we introduce the detailed design and implement. As the same set of physical CPUs in the physical machine are shared by all the virtual CPUs of the virtual machines, the scheduling latency, or the waiting time in the scheduling queues will actually contribute a lot to the virtual I/O events' processing delay time, which finally result in a low I/O throughput and a long processing waiting time. To mitigate this negative impact while still take the advantage of this resources sharing mechanism, we use non-uniform time slices for different kinds of virtual VMs instead of the uniform designed time slice in the credit scheduler.

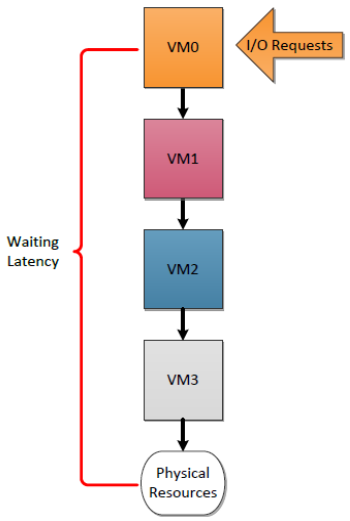


Fig. 1. Original Scheduling Latency

To achieve this goal, in this paper we proposal Elastic timE slicE controllEr (E4) software solution. We first divide all the virtual machines into two classes, which we called the I/O-sensitive VMs and the CPU-sensitive VMs. If a specified virtual machine belongs to the I/O-sensitive VMs, it means that this VM is sensitive to the I/O latency and requires timely responses for I/O events. On the other hand, CPU-sensitive VMs are not so sensitive to the I/O latency but requires a large computing time. For example, the VM holds a live meeting service or online game service might probably belongs to the class of I/O sensitive VMs, while a VM running some graphic algorithms are not. We then arrange more scheduling opportunities with a shorter time slice to the I/O-sensitive VMs, while keep the scheduling mechanism untouched for CPU-sensitive VMs. As a result, the I/O sensitive VMs will get more opportunities and more frequently to be scheduled, which will shorten the waiting time between their two scheduling round and contribute to a quick response for the I/O events. At the same time, we also control the whole scheduling time unchanged to ensure that the fairness among VMs is not disturbed by our design.

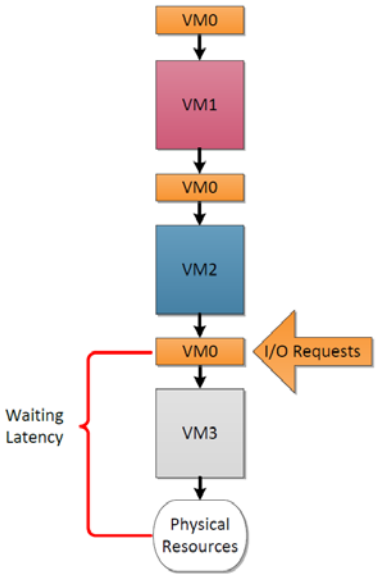


Fig. 2. Modified Scheduling Latency

Our design is based on the analysis of Figure 1 and Figure 2. As show in Figure 1, we assume that VM0 is an I/O sensitive virtual machine. When a virtual I/O event shows up, it has to wait for the virtual CPU of VM0 to be scheduled, which results in a long waiting time. On the other hand, if we use shorter time slice but more scheduling opportunities for VM0, the scheduling delay would become shorter, as shown in Figure 2.

III. EVALUATION

In this section, we carry out the detailed experiments for our design. We use the light-weight command Ping and a common test tool Netperf [13] to measure the performance improvement achieved by our design. In this section, we first have a clear introduction of our experiment environment. We

then go through the experiments and have a detail discuss based on the analysis of the results.

A. Experiment Environment Setup

In the experiment we use two physical machines as our test machines. One of the physical machine acts as the host machine, inside which we install Xen hypervisor and create virtual machines for testing, while the other is responsible for sending test requests as a client to the virtual machines inside the first physical machine. For each physical machine, a 3.10GHz Intel Core i5-3450 CPU is equipped with 1T SATA hard disk and 16GB physical memory. As for the network, both of the two machines are working and connected in a same real world network environment, where other 50+ machines are connected as well. As for the software, we use Xen 4.3.0 as the hypervisor and Ubuntu 13.04 as the operating system in the physical machine. The detailed machine configuration is shown in Table I.

TABLE I. EXPERIMENT ENVIRONMENT

Item Name	SPEC
Memory	16GB
CPU	Intel Core i5-3450
Hard Disk	1T SATA Disk
Network Adapter	1Gb network card
VMM	Xen 4.3.0
OS	Ubuntu 12.04

B. Evaluation with Ping

Ping is a very light weight command to test the network I/O latency. We first create four identical virtual machines in the host machine, with each virtual machine equipped with 1Gb memory, one virtual CPU and 10Gb hard disk. Each virtual machine uses Ubuntu as its operating system. Three of these virtual machines are classified as CPU-sensitive VMs, while one is regarded as I/O sensitive. We run a background CPU burning tasks in each CPU-sensitive virtual machine to keep up an appropriate CPU utilization. In this experiment, we keep Pinging the I/O sensitive virtual machine from the other physical machine in the same LAN for about ten minutes and record the responding latency of each ping command. We carry out the same testing on both E4 and the original Xen hypervisor to see how much improvement E4 could achieve compared to the original system under the same testing environment. The CDF of the result is shown in Figure 3. The green line in the CDF stands for the experimental results of E4, while the red line in the figure stands for the results of the original Xen 4.3.0 hypervisor.

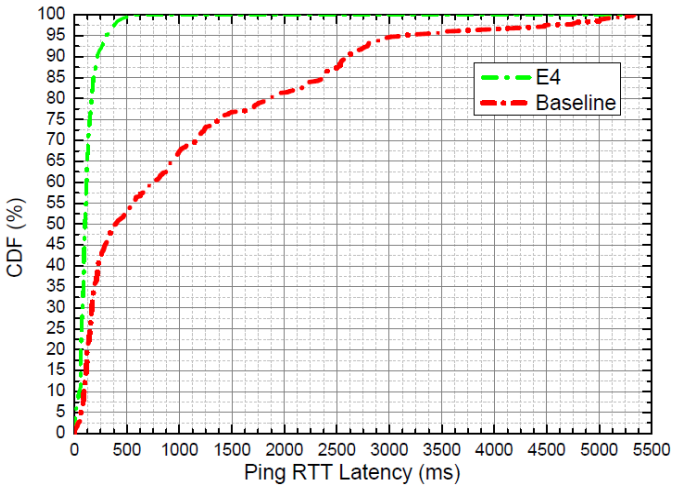


Fig. 3. Test result of Ping RTT

From Figure 3 we can see that, for the original Xen hypervisor, the Ping RTT delay is large and unpredictable. Up to 50\% results are larger than 500ms, while in the worst case the ping latency reaches 5,000ms. On the other side, the Ping results of E4 are much shorter. Up to 90\% results are shorter than 250ms, while in the worst case the largest Ping RTT latency is only around 500ms. These results show that under the same environment, our design E4 improves the network I/O latency drastically.

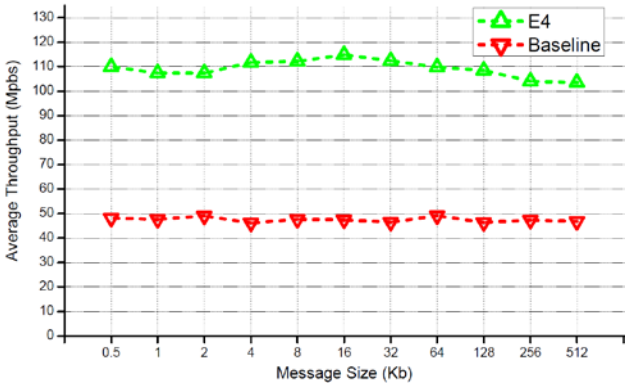


Fig. 4. Test result of TCP_STREAM

C. Netperf Experiment

Netperf is a widely used testing tool to measure the network throughput and I/O latency [13]. In this section, we also create the I/O sensitive virtual machine as the test machine, and other three CPU sensitive virtual machines as the background. We deployed the Netperf server in the I/O sensitive virtual machine, and use the other physical machine in the same LAN as the client. The client machine then sends out requests to the I/O sensitive virtual machine and records the time spent in total. The results of this experiment are in the Figure 4 and Figure 5.

Figure 4 shows the results of TCP_STREAM experiment with Netperf. We send out Netperf testing commands from

another physical machine with the message size ranging from 0.5Kb to 512Kb, and record the average throughput. As the previous experiment, the green line stands for the result of E4, while the red line in the Figure stands for the result of the baseline. From Figure 4 we see that E4 successfully achieves an average throughput over 100Mbps, while for the baseline, i.e. the original Xen 4.3.0 hypervisor, the average throughput is smaller than 50 Mbps.

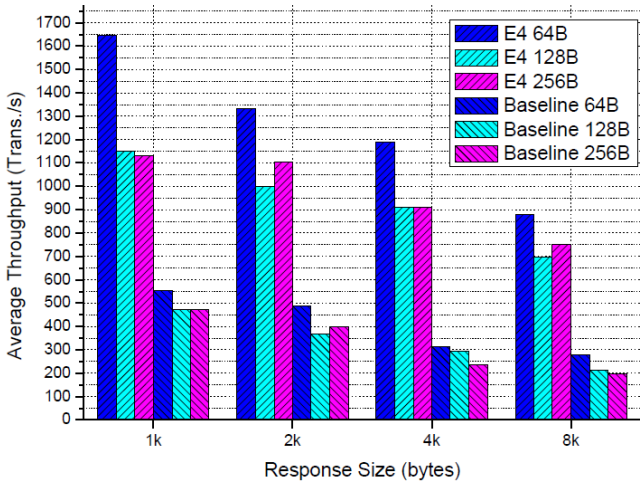


Fig. 5. Test result of UDP_RR

Figure 5 shows the results of UDP_RR experiment with Netperf. We also use different request and response message sizes to test E4 and the original baseline. From the figure we can also see that for different sizes, E4 shows a higher average RR throughput than the baseline. For example, when the request size is 64B and the response size is 1k, E4 achieves 1650 trans per second, while for the baseline the result is only 550 trans per second. From the experiment results we can draw the conclusion that E4 brings in a significant improvement on reducing the network I/O latency.

IV. CONCLUSION

In this paper, we propose Elastic timE slicE controllEr (E4) software solution to mitigate the I/O latency problem for virtual machines. We first introduce and analyze the I/O latency problem for virtual machines, and point out that the scheduling latency actually makes a lot of contributions to the I/O events' processing latency. We further more point out that a non-uniform time slice in the scheduler will shorten the scheduling delay for I/O sensitive tasks. Based on our analysis of the problem, we divide the VMs into two I/O sensitive class and CPU sensitive class, and propose a new scheduling algorithm with a non-uniform time slice. We use a shorter time slice for I/O sensitive VMs to get a quicker response ability, while a longer time slice is suitable for the CPU sensitive tasks. The detailed experiment results show that E4 is able to achieve a significant improvement in terms of network I/O delay and throughput.

REFERENCES

[1] L. M. Vaquero, L. Rodero-Merino, J. Caceres, and M. Lindner, “A break in the clouds: towards a cloud definition,” SIGCOMM Comput. Commun. Rev., vol. 39, no. 1, pp. 50–55, Dec. 2008.

[2] L. Cheng and C.-L. Wang, “vbalance: Using interrupt load balance to improve i/o performance for smp virtual machines,” in Proceedings of the Third ACM Symposium on Cloud Computing, ser. SoCC’12. San Jose, CA, USA: ACM, 2012, pp. 3–14.

[3] P. Barham, B. Dragovic, K. Fraser, S. Hand, T. Harris, A. Ho, R. Neugebauer, I. Pratt, and A. Warfield, “Xen and the art of virtualization,” in Proceedings of the nineteenth ACM symposium on Operating system sprinciples, ser. SOSP ’03. New York, NY, USA: ACM, 2003, pp. 164–177.

[4] A. Kivity, “kvm: the Linux virtual machine monitor,” in OLS ’07: The 2007 Ottawa Linux Symposium, Jul. 2007, pp. 225–230.

[5] C. A. Waldspurger, “Memory resource management in vmware esx server,” in Proceedings of the 5th symposium on Operating systems design and implementation, ser. OSDI ’02. USENIX, 2002, pp. 181–194.

[6] A. Velte and T. Velte, Microsoft Virtualization with Hyper-V, 1st ed. New York, NY, USA: McGraw-Hill, Inc., 2010.

[7] G. von Laszewski, L. Wang, A. Younge, and X. He, “Power-aware scheduling of virtual machines in dvfs-enabled clusters,” in Cluster Computing and Workshops, 2009. CLUSTER ’09. IEEE International Conference on, 31 2009-sept. 4 2009, pp. 1 –10.

[8] D. Gupta, L. Cherkasova, R. Gardner, and A. Vahdat, “Enforcing performance isolation across virtual machines in xen,” in Proceedings of the ACM/IFIP/USENIX 2006 International Conference on Middleware, ser. Middleware ’06. New York, NY, USA: Springer-Verlag New York, Inc., 2006, pp. 342–362.

[9] Z. Chang, J. Li, R. Ma, Z. Huang, and H. Guan, “Adjustable credit scheduling for high performance network virtualization,” in 2012 IEEE International Conference on Cluster Computing, ser. CLUSTER ’12. Beijing, BJ, CHN: IEEE, 2012, pp. 337–345.

[10] “Credit scheduler,” http://wiki.xensource.com/xenwiki/CreditScheduler

[11] I. Leslie, D. McAuley, R. Black, T. Roscoe, P. Barham, D. Evers, R. Fairbairns, and E. Hyden, “The design and implementation of an operating system to support distributed multimedia applications,” 1997

[12] K. J. Duda and D. R. Cheriton, “Borrowed-virtual-time (bvt) scheduling: supporting latency-sensitive threads in a general-purpose scheduler,” in Proceedings of the seventeenth ACM symposium on Operating systems principles, ser. SOSP ’99. New York, NY, USA: ACM, 1999, pp. 261–276.

[13] “Netperf,” http://www.netperf.org/netperf/

[14] “Virtualbox,” https://www.virtualbox.org/

Research on Significance of VCPU Scheduling for SR-IOV on NUMA platform

Xiaolong Jia
School of Electronic Information and Electrical Engineering
Shanghai Jiao Tong University
Shanghai 200240, P.R. China
jxlsjtu@sjtu.edu.cn

Minjun Zhu
School of Electronic Information and Electrical Engineering
Shanghai Jiao Tong University
Shanghai 200240, P.R. China
zmj1989@sjtu.edu.cn

***Abstract*—Virtualization is vital important in server consolidation. Using virtualization technology, a single physical machine can provide multiple isolated virtual machine for users, which will improve the overall system resources usage. However, In virtualization environment, Non-Uniform Memory Access Architecture (NUMA) problem is also a very important parameter that affects server's performance. To solve NUMA problem, many solutions have been proposed to reduce its performance impact on virtual machines and improve the overall system performance. Hardware Single Root I/O Virtualization (SR-IOV) is a solution for high performance network processing, but when SR-IOV technology is running on a NUMA platform, there will also generate the traditional problem about NUMA.**

We take into vCPU allocation strategy of VM to check the performance influence of a VM.

We also conduct detailed experiments to evaluate our design and implement under a physical network environment. The result shows that our design achieves a significant improvement in terms of network performance, while introducing a small overhead to the original system.

Keywords—Virtualization;NUMA;SR-IOV;High performance network;Xen;

I. INTRODUCTION

Recent years, network virtualization has gained great performance improvement. Virtualization technology introduce a software layer called Virtual Machine Manager (VMM) between physical resource and guest operating system. VMM manages all the physical resources like CPU and memory for the upper virtual machines. In this way, virtualization technology can provide isolated virtual machines for users and improve the overall system resource usage.

Although virtualization can improve system performance, new challenges have arisen due to the introduction of VMM. A very common scene is that when a VMM is running under NUMA platform, while NUMA is used to solve memory bottleneck in SMP and is widely used in high performance servers. But there is a traditional problem in NUMA[4], namely it's quite quicker to access data in local node than access data in remote nodes. So in virtualization environment, it's also a key factor that affect the server's whole performance.

SR-IOV[1][2] is used in high performance network performance, device like Intel 82599 achieved about 10 Gb/s network performance. It tries to use Virtual Function (VF) which is directly assigned to a VM, then the VM can access it's memory directly without the interface of VMM. So a VM can achieve a very high network performance and this technology is widely used in high performance network area.

So for a NUMA VM which is assigned with SR-IOV's Virtual Function (VF), when the VM starts its network interface, the VM allocate a buffer for the VF, but the hypervisor is not aware of that. Although through Hardware-Assisted Virtualization tools like Input/Output Memory Management Unit (IOMMU) has the address information translation of VF (PCIe device), the hypervisor has done nothing to utilize that information. In general, in this strategy hypervisor has its method to choose destination CPU to run the vCPU of a VM, which may lead to the situation that vCPU can't obtain the optimum affinity with its corresponding network buffer. Then while network packets arrives, vCPUs process the packets inefficiently. This is a classical problem for a task running under NUMA platform. And nowadays, network performance is a key indicator of servers, like in network station, cloud computing center, hadoop map-reduce and so on.

Previous Researches that is related with NUMA[5] are focusing on non-virtualized environment[6][7]. Someone tries to get the better affinity between the task's computations and data[8], and achieved good acceleration in performance. Someone tries to analysis NUMA resource contention (like cache). And there are also some researchers considering in the memory traffic when allocating memory for applications[9][10]. They all haven't considered about the virtualized environment. There are few researchers has done some work about VM memory allocation strategies, which show a big influence for VM's performance. One of the reason is that under virtualization environment, the problem get complex for that hypervisor is in charge of memory. So it becomes difficult when try to analyze a task's memory distribution in a VM, and there is little tools like NUMA API under linux to control the resources for a VM's task. So there is little researches has been done in the virtualized environment.

And researches that is related with SR-IOV is mostly about how to alleviate interface of VMM[3][11][13]. They doesn't consider in VF buffer location.

The contributions of this paper can be summarized as follows:

- We analyze the SR-IOV hardware assisted technology and discuss performance deviation when SR-IOV have different vCPU configuration.
- We performance several experiments to verify the performance deviation in different vCPU configuration.
- We make a conclusion on the necessity of adjusting VM vCPU configuration.

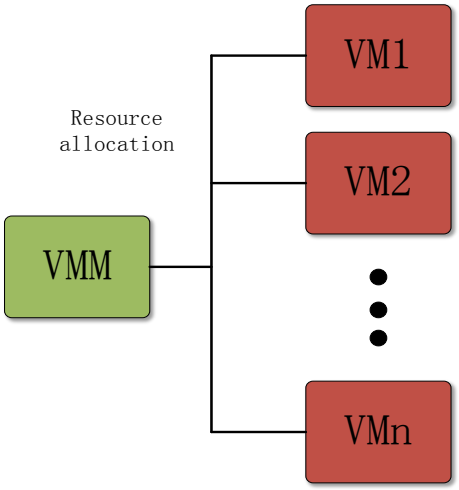


Fig 1.Original Resource allocation strategy

The rest of the paper is organized as follows: Section II presents a detailed analysis of NUMA technology, SR-IOV and the problem. Section Ⅲ presents the design of experiments and the result and analysis for each experiment. Section IV conclude the result of the experiments.

II. Background and Problem

In this section, we first introduce the background knowledge of NUMA technology, SR-IOV and the problem. We then analysis the effect of vCPU configuration changes in the VM.

A. *NUMA Technology*

Modern CPUs operate faster than the main memory they use, in early days, the CPU generally ran slower than its own memory. But large servers has tens of CPU cores. CPUs increasingly have found they are starved for data, so the NUMA platform try to solve this problem while introducing separate memory for each processor, avoiding the performance hit when several processors attempt to address the same memory. But there also bring in some problem. The classical problem is that a processor can access its own local memory faster than non-local memory, and while the nodes is enough

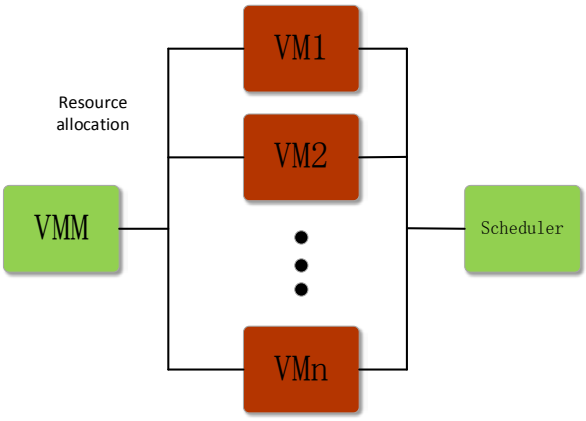


Fig 2.Original Resource allocation strategy

large, the time a processor access different non-local nodes is also different.

B. *SR-IOV Technology*

SR-IOV uses physical functions (PFs) and virtual functions (VFs) to manage global functions for the SR-IOV devices. PFs are full PCIe functions that include the SR-IOV Extended Capability which is used to configure and manage the SR-IOV functionality. It is possible to configure or control PCIe devices using PFs, and the PF has full ability to move data in and out of the device. VFs are lightweight PCIe functions that contain all the resources necessary for data movement but have a carefully minimized set of configuration resources. While using Hardware-assisted technology like Input/Output Memory Management Unit (IOMMU), a VM with SR-IOV VF can access its own device memory directly without the interface of VMM. So the network performance is higher than traditional network solution.

C. *Problem Discussion*

As described above, Fig 1 shows that in virtualization environment, the VMM will assigned a VM resources like vCPUs and memory for a VM. When the VM start it VF, the VM will create a buffer which is used to receive and send packets. This bring in a problem, the VM vCPU is not always running corresponding with the VM VF buffer. So Fig 2 shows our modify to the VMM vCPU scheduling strategy.

Although Xen 4.3 has done some work for NUMA aware scheduling, but it also can't promise the VM's vCPU will always running on the nodes that the the VM buffer allocated, so the VMM can't promise the most optimum performance for network processing.

From previous discussion, it is obvious that if we try to colocate the vCPU and VF buffer, the network performance will get a better performance.

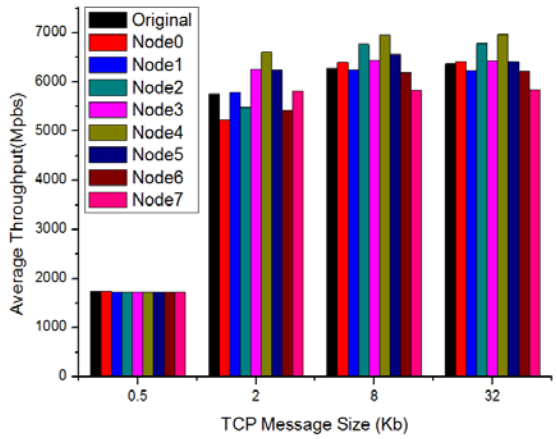


Fig 3. Problem

III. EVALUATION

In this section we perform several experiments to demonstrate the performance deviation of the VMs in different scenario using different vCPU configuration.

A. *Experimental Environment*

In our experiments, We conduct our experiments on two machines which are connected via a 10 Gigabit Ethernet Intel 82599 switch in LAN, with each equipped with a 64-core 2.3GHz, AMD Core 6376 CPU, 256GB physical memory, one 10 GbE network card and two 2TB SATA disk. One of the physical machines runs Ubuntu 13.04 as the server, with 64-bit Xen 4.3.0 installed.The benchmarks we use in our experiments are Netperf[12].

B. *Evaluation with netperf for test*

Netperf is a widely used as a testing tool to measure the network throughput[12]. In this section, we also create the I/O sensitive virtual machine as our test machine, and other four virtual machines as the background. We deployed the Netperf server in the I/O sensitive virtual machine, and use the other physical machine in the same LAN as the clients.

From Fig 3 we can see a big difference while allocating VM's vCPU to different node, so if we can find out the best allocation for the VM's vCPU, there will show a big improvement.

C. *Netperf Experiment*

Fig 4 shows the results of TCP_STREAM experiment with Netperf. We run Netperf testing commands from another physical machine with the message size ranging from 0.5Kb to 512Kb, and record the average throughput. From Fig 4 we see that our modified strategy successfully brings up the throughput from 5.5Gbps to over 6.5Gbps.

Fig 5 shows the results of UDP experiment with Netperf.

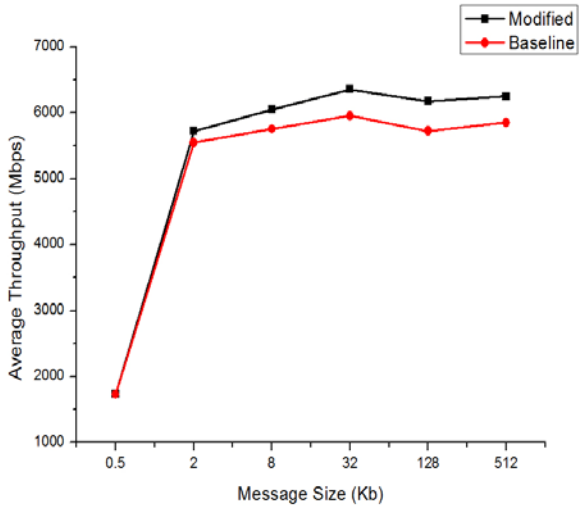


Fig 4. Test result of TCP_STREAM

We also use different request message sizes to test our modified strategy and the original baseline. From the figure we can see that for different sizes, our strategy shows a higher average UDP throughput than the baseline. For example, when the request size is 25.6Kb, we achieves 4.9 Gbps, while the baseline result is only 4.2Gbps. From the experiment results we can draw the conclusion that our strategy brings in a significant improvement on improving the network I/O throughput.

IV. SUMMARY

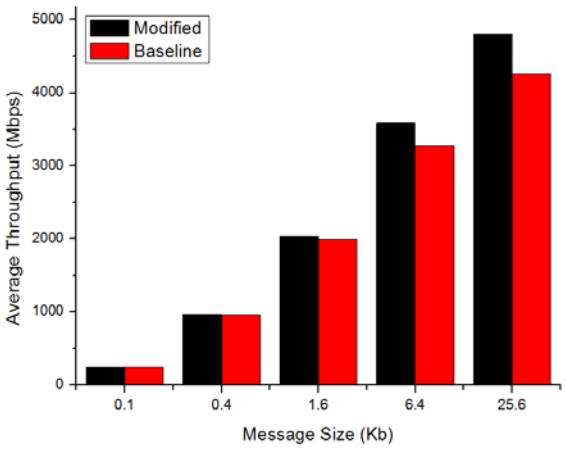


Fig 5. Test result of UDP_STREAM

In this section, we perform several experiments to check the performance of the VMs in different scenarios. The factors we consider in is the vCPU location for theVM. By comparing the experiment results, we can see that the best performance for a VM changes with the network buffer location of the VM.

V. CONCLUSION

In this paper, we demonstrate the necessity of adjusting NUMA vCPU configuration when A VM is using VF module of SR-IOV. We first introduce and analyze the SR-IOV problem under NUMA platform in virtualize environment. Then we introduce and analyze the SR-IOV network buffer and hardware assisted technology like IOMMU which is used to do address translation of VMs PCIe devices.

We discuss and analyze the importance of the vCPU parameter values when processing network packets in VMs and make a conclusion that adjusting vCPU location is necessary in theory. So we perform several experiments to verify the performance deviation in while using different vCPU configuration.

From the experimental results, we conclude that the performance deviation is pretty large in different scenarios using different vCPU configuration and it is necessary to adjust vCPU configuration dynamically to fully improve the overall system performance.

REFERENCES

[1] Y. Dong, Z. Yu, and G. Rose, "Sr-iov networking in xen: Architecture, design and implementation." in Workshop on I/O Virtualization, 2008.

[2] J. Liu, "Evaluating standard-based self-virtualizing devices: A performance study on 10 gbe nics with sr-iov support," in Parallel & Distributed Processing (IPDPS), 2010 IEEE International Symposium on. IEEE, 2010, pp. 1–12.

[3] H. Guan, Y. Dong, K. Tian, and J. Li, "Sr-iov based network interruptfree virtualization with event based polling," Selected Areas in Communications, IEEE Journal on, vol. 31, no. 12, pp. 2596–2609, 2013.

[4] "Non-uniform memory access(numa)," http://en.wikipedia.org/wiki/Non-uniform memory access.

[5] J. Ramanathan and L. M. Ni, "Critical factors in numa memory management," in Distributed Computing Systems, 1991., 11th International Conference on. IEEE, 1991, pp. 500–507.

[6] R. P. LaRowe Jr, C. S. Ellis, and M. A. Holliday, "Evaluation of numa memory management through modeling and measurements," Parallel and Distributed Systems, IEEE Transactions on, vol. 3, no. 6, pp. 686– 701, 1992.

[7] Z. Majo and T. R. Gross, "(mis) understanding the numa memory system performance of multithreaded workloads," in Workload Characterization (IISWC), 2013 IEEE International Symposium on. IEEE, 2013, pp. 11–22.

[8] Y. Ren, T. Li, D. Yu, S. Jin, and T. Robertazzi, "Design and performance evaluation of numa-aware rdma-based end-to-end data transfer systems," in Proceedings of SC13: International Conference for High Performance Computing, Networking, Storage and Analysis. ACM, 2013, p. 48.

[9] S. Blagodurov, S. Zhuravlev, A. Fedorova, and A. Kamali, "A case for numa-aware contention management on multicore systems," in Proceedings of the 19th international conference on Parallel architectures and compilation techniques. ACM, 2010, pp. 557–558.

[10] D. S. Rao and K. Schwan, "vnuma-mgr: Managing vm memory on numa platforms," in High Performance Computing (HiPC), 2010 International Conference on. IEEE, 2010, pp. 1–10.

[11] J. Liu, "Evaluating standard-based self-virtualizing devices: A performance study on 10 gbe nics with sr-iov support," in Parallel & Distributed Processing (IPDPS), 2010 IEEE International Symposium on. IEEE, 2010, pp. 1–12.

[12] "Netperf," http://www.netperf.org/netperf/.

[13] J. R. Santos, Y. Turner, G. Janakiraman, and I. Pratt, "Bridging the gap between software and hardware techniques for i/o virtualization," in USENIX 2008 Annual Technical Conference on Annual Technical Conference, ser. ATC'08. Boston, MA, USA: USENIX, 2008, pp. 29–42.

Survey of Structure from Motion

Gao Yi[1,2]
[1]College of Command Information System
PLA University of Science and Technology
[2]Department of Remote Sensing and Mapping
PLA Surveying and Mapping Navigation
Nanjing, China
iamninigao@sohu.com

Luo Jianxin, Qiu Hangping, Wu Bo
College of Command Information System
PLA University of Science and Technology
Nanjing, China

***Abstract*—As a useful technology of 3D reconstruction based on binocular stereo vision, structure from motion is widely used in many fields and highly valuable in applications. However, few review works have been focused on this technology. In this paper, the basic principles are overviewed. More specifically, the related works and main methods are discussed. Finally some future research directions are summarized.**

Keywords—structure from motion; 3D reconstruction; overview

I. INTRODUCTION

In 2009, Researches from Cornell University Of Washington enabled the reconstruction of datasets of 150,000 images with the keywords 'Rome' from Internet by Structure from Motion(SFM) algorithms. Exciting progress showing the world the strong ability of applying SFM algorithms on large image reconstruction.

In fact, SFM relates to estimate the locations of 3D points from a lot of images which only given a sparse set of correspondences between image features. During the main part of this process, both 3D geometry (structure) and camera pose(motion) are estimated simultaneously. It is commonly known as structure from motion[6].

One of the neatest application of SFM is in the reconstruction of 3D objects and scenes from collections of images. In this case, there are two questions: first, little is known about the cameras taking the photographs; second, the cost of the SFM is high for large scene components.

Many scholars do a lot of work to promote the development of this field, but now most previous records [1~4] describe the SFM in a general way, and there is no in-depth analysis, it is necessary to do a more comprehensive knowledge about the SFM.

In this paper, the basic principle of SFM is introduced, then the implementation technique and research status are reviewed and the main algorithms are analyzed and discussed, finally, a discussion of the problems of SFM and future direction of research are listed.

II. BASIC PRINCIPLE OF SFM

With the two-frame structure from motion problem as an example:

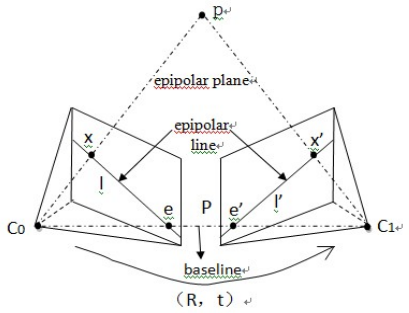


Fig. 1a. Two-frame structure from motion.

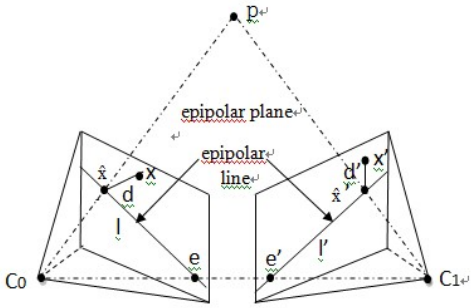


Fig. 1b. Minimization of the reprojection errors.

Consider Figure 1a, which shows a 3D point p being viewed from two cameras whose relative position can be encoded by a rotation R and a translation T. C0 and C1 are around the center of the two cameras. C0x and C1x are the 3D rays corresponding to the 2D matching feature locations x and x' .If arrive at the basic epipolar constraint[9] $x'Fx = 0$ (F is fundamental matrix), C0x and C1x are co-planar and meet in a 3D point p. However, due to image affected by noise, either x , x' or $\hat{x}$, $\hat{x}'$ (the estimation of x , x') has arrived at the basic epipolar constraint(Figure1b). As $\hat{x}$ and $\hat{x}'$ are known, point p can be determined by triangulation. Point p selection criteria is to minimize the projection error $d^2 + d'^2$.

III. RELATED WORK

Figure 2 summarizes the process of SFM implementation, containing data preprocessing, feature extraction, image matching, estimate fundamental matrix, camera calibration and bundle adjustment.

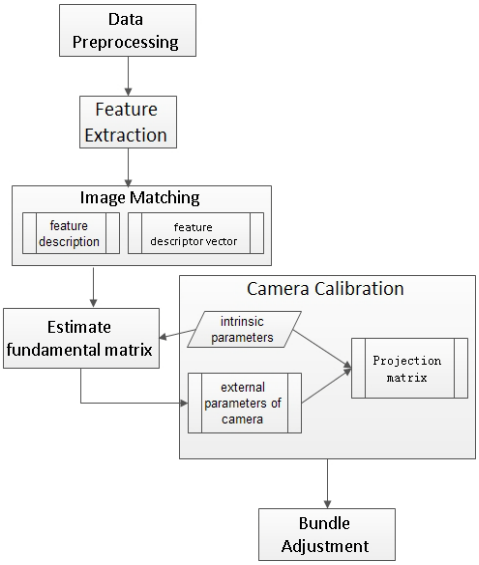


Fig.2. The process of SFM implementation

A. Data Preprocessing

Due to the data acquisition and the impact of computer storage and processing power, traditional SFM does not involve vital amounts of image data processing. With the rapid development of computer hardware and network, data acquisition has become simple and feasible, However, data inevitably appeared complications and redundancy. For example, two images present strong correlation when their distance are very close. Addition to denoise image, simplification has become a necessary operation in data preprocessing stage.

Snavely[8] et al develops a skeleton extraction algorithm. Firstly, estimates the position covariance between pairs of overlapping images. Then, computes a small skeletal subset of images. Finally, adds the remaining images using pose estimation. Kushal[4] et al plans the shortest path among with images by Dijkstra algorithms. Although removing the redundancy of the image and extracting the skeleton points have effectively reduce parameters, the number of remaining images is still huge.

B. Feature Extraction

Feature extraction and matching is the foundation of SFM method, its goal is to found and correct matching key points of the same object or scene in the two images. Harris[10] proposes one of the earliest corner detectors. The seminal work of Lowe et al presents SIFT[11] which filtered the image with differences of Gaussians. PCA-SIFT reduced the descriptor from 128 to 36 dimensions with Principal Component Analysis. Another widely used keypoints at the moment is SURF[12]. It has similar matching performances as SIFT but is much faster. However, the dimensionality of the feature vector is still too high for large-scale 3D reconstruction.

Leutenegger et al proposes a binary descriptor matching algorithm BRISK[13] based on Fast for keypoint detection. BRISK makes use of an easily configurable circular sampling pattern from which it computes brightness comparisons to form a binary descriptor string. Compared with the random sampling mode in BRISK, FREAK[14] based on fixed-point sampling mode is more robust than SIFT,SURF or BRISK. Gao Hong-bo et al[15] presents a binary descriptor matching algorithm based on hierarchical learning method which combines the advantages of BRISK and FREAK, the proposed algorithm outperforms the classical methods with lower computation time. Although the binary descriptor on the computing speed and memory capacity have improved significantly, the way may leads to lost a lot of information.

C. Image Matching

Time complexity of image matching is decided by two aspects: one is the time complexity of similarity comparison, the other is the time complexity of search.

In order to reduce the time complexity, we often creates a kd tree to organize the feature sets and using KNN(K-nearest neighbor) algorithm to speed up the matching process. Agarwal et al[2] puts forward a vocabulary tree-based approach, where K-Means algorithm is used to quantizing the image features. The clusterings of features extracted from a training set are defined as a word and changed into weight vector by TF-IDF model. This approach can calculate the similarity between word frequency and the target images; Irschara[17] is inspired by vocabulary tree-based approach and uses the GPS priors to filter the target images, Cao and Snavely[18]use SVM models to predict matching and non-matching images pairs better than TF-IDF for large-scale image matching.

D. Fundamental Matrix

The estimation of fundamental matrix is commonly use RANSAC. The realization of algorithm is in[9]. Faugeras[19] proved that as long as the fundamental matrix between two images is known, it can realize the projective reconstruction.

E. Camera Calibration

Camera calibration target is to determine the relationship between image coordinate and world coordinate. As can be seen from Figure 3.

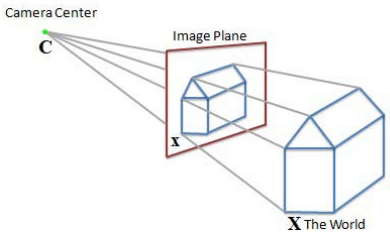


Fig.3. The realationship between world coordinate and camera coordinate[35]

The coordinates of the point X in the world coordinate are (X, Y, Z), and the corresponding image point x in the camera coordinate are (x,y):

$$\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} R & T \\ 0 & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \quad (1)$$

here f is the camera focal length, R is rotation matrix and T is translation matrix.

The accuracy of calibration directly affects the results of the subsequent reconstruction. Camera calibration is mainly divided into two categories:

Traditional camera calibration: such as DLT(Direct Linear Transformation)[20], two-stage camera calibration[21] and Zhang Zheng-you calibration[22] can achieve high accuracy, but usually algorithm is more complex and depends on the high precision calibration block, this paper introduces the SFM technology relies on the fundamental matrix F which plays a very important role in the process of recovering certain information about camera intrinsic.

Camera self-calibration: Faugeras[23] et al use Kruppa equation to calibrate camera parameters. Lourakis[24] simplifies Kruppa equation by using SVD decomposition. Most of these techniques assume that the scene is taken with the same camera and hence the images all have the same intrinsic, Pollefeys[25] enables the camera self-calibration with variable camera parameters .

F. Bundle Adjustment

Bundle Adjustment(BA)[5] is the key technology of SFM, the most accurate way to recover structure and motion is to perform robust non-linear minimization of the re-projection errors called bundle adjustment.

Sparse bundle adjustment (SBA)[26] is an available high-quality algorithm based on incremental standard equation. Irschara[16] proposes a fast feasible reconstruction algorithm, one-time to calculate rotation matrix of all cameras in the global coordinate system. However, this kind of seed and grow approach on dealing with huge amounts of data, the time complexity is still large. PCG (Preconditioned Conjugate Gradients)[27] is proved more effective method to solve large-scale bundle adjustment problem. Liu Xin[28] et al presents a novel distributed bundle adjustment algorithm for solving the massive-points BA problem, where the original BA problem is divided into sub-problems by partitioning the 3D reconstructed points.

IV. EXPERIMENT

Bundler is a SFM system for unordered image collections based on the Photo Tourism work of Noah Snavely et al. Due to running on single computer, we only randomly select seven images from Notre Dame dataset to test by bundler. Notre Dame dataset is a set of photos of the Notre Dame Cathedral in Paris. The computer was equipped with 4 GB of RAM and of local hard disk space and running the Microsoft Windows7 32-bit operating system. Then use CMVS-PMVS supported by Yasutaka Furukama to product dense mesh model. The reconstruction of seven images(Figure 4a) is shown in Figure 4c.

Fig.4a. Seven images of Notre Dame dataset

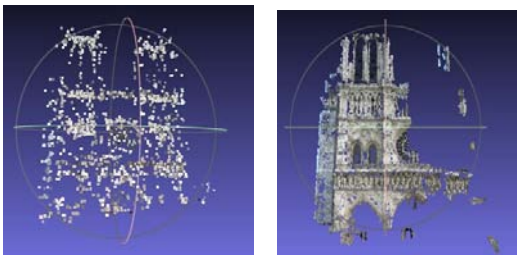


Fig.4b. The reconstruction of seven images

Fig.4c. The reconstruction after CMVS-PMVS.

Output files of Bundler which suffix is called '.out' and '.ply'. The former is used to store the current state of the images after analysis and registration(Figure 5), the latter is used to record the information of points and camera after reconstruction.

```
# Bundle file v0.3
7 1974
2586.568477 0.0 0.0
0.999172 0.006216 0.040209
-0.007904 0.999088 0.041961
-0.039912 -0.042244 0.998310
0.159750 0.175723 -0.090145
2519.493923 0.0 0.0
0.996470 -0.028192 -0.079076
0.028902 0.999551 0.007843
0.078819 -0.010101 0.996838
-0.576690 0.249271 0.607875
3133.654506 0.0 0.0
```

Fig.5.Bundler.out

The first line has two parameters: <num_cameras> and <num_points>, the following lines respectively are the focal length f, followed by two radial distortion coeffs, a 3*3 matrix representing the camera rotation R and a 3-vector describing the camera translation T. Combining the point informations and using formula (1),we can get the point clouds.

As you can imagine, when the number of images is large enough, the reconstruction will be complete .

V. DISCUSSION AND FUTURE WORK

Based on the above experiment, the author analyzed the main problems of SFM and the next step research mainly has the following several aspects:

A. *Selection of Detectors*

The selection of detectors is crucial for high-quality keypoints. Bundler only use SIFT to extract features, we make a qualitative comparison among with the classic detectors, the results are shown in TABLE I:

TABLE1. A QUALITATIVE COMPARISION AMONG WITH THE CLASSIC DECTORS

method	Blur	Scale	Illumination	Rotation	Time
HARRIS	common	none	common	none	good
FAST	common	none	common	none	best
SIFT	good	best	common	best	common
PCA-SIFT	best	good	good	good	good
SURF	good	common	best	common	good
BRISK	best	best	good	common	best
FREAK	common	good	good	common	good

TABLE I shows that the current methods has each advantages on resistance the image translation, rotation and scale changes, but there is a large contradiction between speed and accuracy. So you can select flexibility according to actual needs.

B. *The problems of BA*

1) Depending on the initial value: different initial values have great influence on the effects of reconstruction.

2) BA performance: BA performance improvement is still an open question.In our experiment, our results could be higher efficient and more compelete if we had considered the parallel framework [29~30].

C. *Dense multiple view*

SFM only calculation the three-dimensional coordinates of the matching points. The effect of reconstruction heavily relies on intensity of feature points. Bundler can only get relatively sparse point clouds, as shown in Figure 4b. We use CMVS to cluster the scene and then use PMVS to get point clouds. Dense multiple view is still in the exploratory stage[34].

In addition, SFM itself involves the estimation of so many highly coupled parameters, there will be large amounts of uncertainty, how to eliminate ambiguity[31] is worth developing in the further research.

VI. CONCLUTIONS

SFM is currently the most widely used in the field of 3D reconstruction. In this paper, the basic principle of SFM is introduced, and the SFM implement method is discussed. Through the experimental results has demonstrated the efficiency and intelligent of SFM, but this method is still not perfect, especially with the increasing growth of images and heavily relies on intensity of feature points. In the next stage, it is important for us to solve practical problems listed in the paper.

REFERENCES

[1] Liu Wei, Wu Yihong, Hu Zhanyi. A Survey of 2D to 3D Conversion Technology for Film [J]. Journal of Computer-Aided Design & Computer Graphics, 2012, 24(1): 14-28.

[2] Agarwal,Snavely,Simon et al.Building Rome in a Day[C]. The Twelfth IEEE International Conference on Computer Vision,Kyoto,2009.

[3] Snavely N, Seitz S M, Szeliski R.Photo Tourism: exploring photo collection in 3D[C].In SIGGRAPH Conf.Proc., 2006.835-846.

[4] Kushal,Self,Furukawa,Gallup,Hernandez,Curless.Photo Tours[C].3DImaging,Modeling,Processing,Visualization and Transmission,Japan,2012.

[5] Bill Triggs, Philip McLauchlan, Richard Hartley et al. Bundle Adjustment—A Modern Synthesis[J].Vision Algrorithms:Theory and Practice,2000,13(5):47-71.

[6] Richard Szeliski.Computer Vision:Algorithms and Application[M/OL].2010.http://szeliski.org/Book/.

[7] Building Rome in a Day (iamge) [EB/OL]. 2009(2009-9-17).http://it.sohu.com/20090917/n266789919.shtml.

[8] Snavely, Seitz, Szeliski. Skeletal graphs for efficient structure from motion[C]. IEEE Computer Society Conference on Computer Vision and Pattern Recognition,Anchorage, 2008.

[9] Richard Hartley, Andrew Zisserman.Multiple View Geometry in Computer Vision [M]. 2002.

[10] Harris C, Stephens M. A combined corner and edge Detector[C].Proceedings of the Fourth Alvey Vision Conference. [S.l.]: [s.n.], 1988: 147-151.

[11] Lowe D. Object recognition from localscale-invariant features[C]. Proceedings of the 7th IEEE International Conference on Computer Vision. Greece, 1999: 1150-1157.

[12] Bay H, Tuytelaars T, Van Gool L. SURF: Speeded Up Robust Features[C]. Proceedings of the 9th European Conference onComputer Vision.Austria, 2006: 404-417.

[13] Leutenegger S, Chli M, Siegwart R. BRISK: Binary Robust Invariant Scalable Keypoints[C].Proceedings of the 13th European Conference on Computer Vision.Spain, 2011:2548-2555.

[14] Alahi A, Ortiz R, and Vandergheynst P.FREAK: fast retina keypoint[C]. Conference on Computer Vision and Pattern Recognition.USA,2012: 510-517.

[15] Gao Hong-bo, Wang Hong-yu, Liu Xiao-kai. A Keypoint Matching Method Based on Hierarchical Learning[J]. Journal of Electronics & Information Technology, 2013,35(11):2751-2757.

[16] D.Nister, H.Stewenius.Scalable recognition witha vocabulary tree[C].Computer Vision and Pattern Recognition.USA,2006:2161-2168.

[17] A. Irschara, C. Hoppe, H. Bischof. Efficient structure from motion with weak position and orientation priors[C].Computer Vision and Pattern Recognition Workshops.USA,2011:21-28.

[18] Song Cao and Noah Snavely.Learning to Match Images in Lage-Scale Collections[C]. Europeon Conference on Computer Vision. Italy,2012.

[19] Faugeras O.and Luong Q T.The Geometry of Mulitiple Images[M].[S.l.]:The MIT Press,2001.

[20] Abdel-AzizaYI,KararaHM.Direct linear transformation into object space coordinate sinclose-rang ephotogram-metry.Proc.Symp.Close-Range Photogrammery.1971:1-18.

[21] Tsai R Y.A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses[J].IEEE Journal of Robotics and Automation,1987,3(4):323-344.

[22] Zhang Z.A flexible new technique for camera calibration[J].IEEE Transaction onPattern Analysis and Machine Intelligence,2000,22(11):1330-1334.

[23] Maybank S J and Faugeras O D.A theory of Self-calibration of a Moving Camera[J].International Journal of Computer Vision,1992,8(2):123-151.

[24] Lourakis M I,Deriche R.Camera self-calibration using the singular value decomposition of the fundamental matrix:From pointcorrespondences to 3D measurements[C]. The 4th Asian Conference onComputer Vision. Taipei,2000:403-408.

[25] Pollefeys M,Koch R,Van Gool L. self-calibration and metric reconstruction in spiteof varying and unknown internal cameraparameters[J]. International Journal of Computer Vision,1999,32(1):7-25.

[26] Lourakis M I A, Argyros A A. SBA: a software package for generic sparse bundleadjustment[J]. ACM Transactions on Mathematical Software, 2009, 36(1): 1-30.

[27] Agarwal S, Snavely N, Seitz S M et al. Bundle adjustment in the large[C]. Proceeding of the 11thEuropean Conference on computer Vision.Part II. Crete, Greece:Springer, 2010.29-42.

[28] LIU Xin, SUN Feng-Mei, HU Zhan-Yi. Distributed Bundle Adjustment in 3D Scene Reconstruction with Massive Points[J]. ACTA AUTOMATICA SINICA, 2012,38(9):1428-1438.

[29] Hadoop[EB/OL].2014(2014-4-10).http://hadoop.apache.org.

[30] Spark:Lightning-fast cluster computing[EB/OL]. 2014(2014-3-30). http:// spark.apache.org.

[31] Kyle Wilson, Noah Snavely.Network Principles for SFM:Disambiguating Repeated Structures with Local Context[C]. IEEE International Conference on Computer Vision Sydney, 2013.

[32] Song Cao, Noah Snavely. Minimal Scene Descriptions from Structure from Motion Models[C]. Computer Vision and Pattern Recognition.USA,2014.

[33] Smith B M, Li Zhang, Hailin Jin. Stereo Matching with Nonparametric Smoothness Priors in Feature Space[C]. IEEE Computer Society Conference on Computer Vision and Pattern Recognition.USA, 2009.

[34] Amael Delaunoy, Marc Pollefeys. Photometric Bundle Adjustment for Dense Multi-View 3D Modeling. Computer Vision and Pattern Recognition. USA,2014.

[35] Xiao jian-xiong.Multi-view 3D Reconstruction for Dummies.http://mit.edu/jxiao/Public/software/SFMedu

Smart Agent Based Prepaid Wireless Energy Meter

Au Thien Wan, Suresh Sankaranarayanan and Siti Nurafifah Binti Sait
School of computing and Informatics
Institut Teknologi Brunei
Brunei Darussalam
e-mail: twan.au@itb.edu.bn , suresh.sn@itb.edu.bn, fifahsait@gmail.com

***Abstract* - Prepaid meter (PM) is getting very popular especially in developing countries. There are many advantages to use prepaid meter as opposed to postpaid meter both to the utility provider and to the consumer. Brunei Darussalam has adopted PM but it is not intelligent and not wireless enabled. Reading meters and topping up balance are still done manually. Utility provider does not have information on the usage statistics and has only limited functionalities in the grid control. So accordingly a novel software agent based wireless prepaid energy meter was developed using Java Agent Development Environment (JADE-LEAP) allowing agent from utility provider to query wireless energy meter for energy values for every household. These statistics can be used for statistical computation of the power consumed and for policy and future planning.**

Keywords - Wireless Prepaid Smart Meter, JADE, LEAP, Prepaid Meter

I. INTRODUCTION

The traditional method of electricity billing system involves meter readers to periodically visit every house to take readings. There are many issues to this method such as taking wrong readings, lack of meter readers, and houses in very remote areas, meters in inconvenient location and so forth. Many technological advancement have been carried out and one such is employing software agent replicating human beings to collect energy values by means of power line communication [1][2][3][4][5]. In some cases wireless technologies are used where energy meter embedded with Zigbee sensor making it wireless accessible for retrieving energy values for billing [6][7][8][9][10][11] and some even employ GPRS for retrieving the energy units for billing from Zigbee based wireless remote meter [12].

Prepaid meters (PM) offer many advantages both to the utility provider and to the consumers. To the utility provider, this reduces tremendously many issues arise from meter readers such as delays, wrong and infrequent meter reading resulting in bulk amount of billing that consumers would need to pay and further consequent in not paying, disputes and so forth. One of the main motivation of using PM is energy conservation. Brunei Darussalam for example is very actively heading towards achieving energy conservation and the introduction of PM is one of the initiatives. With the introduction of PM the total electrical power usage in 2012 in Brunei Darussalam in the housing sector was 1,879 GWH, which was a reduction of 12.4 per cent compared to the total usage in 2011, which was 2,145 GWH. The total cost saving as a result was 30 million Brunei dollars [13][14].

However PM does not have wireless feature for meter reading and the utility company does not have real time information on how much energy units are consumed, balance units for each household and how much users spend on energy monthly. The users are unable to monitor their energy usage on a regular basis wirelessly too.

Based on the current gaps and the motivations in relation to using a Wireless Prepaid Meter (WPM), we developed a simulated Smart Agent based Wireless Prepaid Meter (SAWPM) using Java Agent Development Environment (JADE-LEAP) agent development kit [15][16]. This research does not concentrate on wireless prepaid meter design but on development of smart agent in wireless prepaid meter for utility provider and consumers. The system here communicates by means of agents mimicking the job of human to collect the balance from meters periodically for consumers and utility provider.

Section II of the paper provides some background research survey related to energy meter. Section III details the system architecture of Smart Agent based Wireless Prepaid Meter (SAWPM). Section IV describes the energy meter simulating energy values consumed for household. Section V shows the functionalities of DES agent at the energy department. Section VI shows agent in Wireless Prepaid Meter (WPM). Section VII concludes the paper with future work.

II. LITERATURE REVIEW

A. Agent based Remote Energy Meter

The first pioneering research work that was done [1] using software agents have been used to access energy meters in the house remotely for recording the units used and accordingly the amount calculated. These agents have been developed in Java-RMI. The agent then replicates the human agent who visits the house to read the meter.

Research was further done in developing agents for remote energy meter reading for billing, where energy department server is connected to domestic clients by means of the power line [2]. In this system software agent has been employed to move from the energy department to the domestic energy meter which then picks up the energy meter data and returns to the energy department for lodging the data [3].

B. *Wireless Remote Postpaid Energy Meter*

In the Sultanate of Oman for instance, a model of Wireless Automatic Meter Reading System (WAMRS) has been developed [7] in which the wireless communication is based on IEEE 802.15.4 (ZigBee) standard and security is implemented by following the Direct Sequence Spread Spectrum (DSSS) protocol. Successful demonstration of WAMRS prototype has made it possible to be implemented in Oman on a larger scale for meter reading applications. The main goal of this research was to send periodical readings of an electricity meter wirelessly to a server in the billing office of the utility company.

In [10], GPRS communication has been used to retrieve meter values from Zigbee Wireless remote meter of water, electricity and gas for computation. This system has low cost and a little power consumption, while it has great extension and security.

C. *Wireless Prepaid energy Meter*

In [17], a GSM-based Energy Recharge system for prepaid metering has been developed for Nigerian Power Sector towards the metering and billing system. The GSM-based Energy Recharge Interface contains a prepaid card equivalent to a mobile SIM card. The prepaid card communicates with the power utility using GSM communication network. Once the prepaid card value has reached zero, the consumer load is disconnected from the utility supply by a latching relay.

III. SMART AGENT BASED WIRELESS PREPAID METER ARCHITECTURE

The Smart Agent based Wireless prepaid meter architecture we proposed consists of three modules; the first is the wireless energy meter that can generate units consumed by the energy meter which is used further for the calculation of every data. The second handles energy monitoring tasks which includes data such as balance units and total top up amount at Department of Electrical Services (DES). The later will handle the exchange of values between devices for querying tasks from user's mobile handset in providing continuous transfer of balance units, top-up amount, in addition to triggering the alert system when the units hit a certain threshold amount. Fig 1 shows the overall architecture design of a Smart Agent Based Prepaid Wireless Energy Meter System. The architecture design consists of four main modules:

- Energy meter simulator which can generate energy units consumed by the household which is used further for the calculation of balance units and for subsequent actions at the Wireless energy meter as a replacement of physical energy meter.
- The JADE Agent Main container for DES which can destroy or combine all active agents in the same environment serving its purposes.
- JADE-LEAP Agent [18][19] for Zigbee enabled Wireless Energy meter.
- JADE-LEAP Agent [18][19] for User Android Application for monitoring and querying balance units wirelessly. It also handles the exchange of values between devices for the querying tasks.

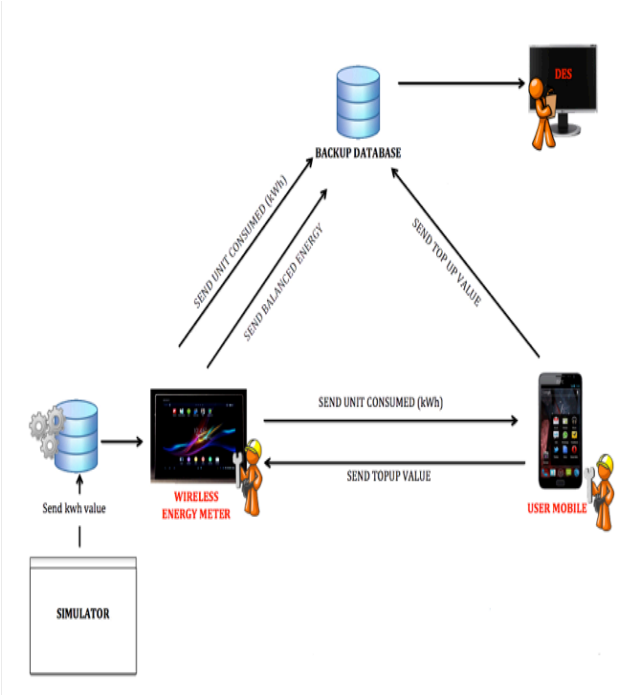


Fig. 1. Agent based Wireless Smart Prepaid Meter System Architecture

A. *System Architecture Design*

The system architecture shown in Fig.1 is aimed to support both the JADE Agents platform and modules of the proposed system in the best possible way, involving all JADE Agents implemented devices and the position of databases in the given environment.

The system started with the energy simulator generating values for energy consumption. These values are saved and will be used for balance unit calculations in wireless energy meter. The values are then supplied to the wireless energy meter and further to all active JADE agent devices as shown above. The DES backup database will become the medium of displaying data in the DES. This is to ensure precise updated records such as units consumed, balance units and total top up for each meter to be revised every 5th of every month. The User Android application and wireless energy meter required logins before entering the system. This is to ensure security authentication for the system

B. *System Database*

There are two different databases involved in this system:

1. Energy Meter database

The purpose of this is to store following information as shown in Fig 2:

- Login information for both energy meter and user Android applications (user logins and meter logins table).
- Units consumed generated by the energy meter simulator (energy units table).

- Data for balance unit calculations

The energy consumption values generated is according to the daily appliances usage of each household and their peak and off peak hour values from DES Brunei.

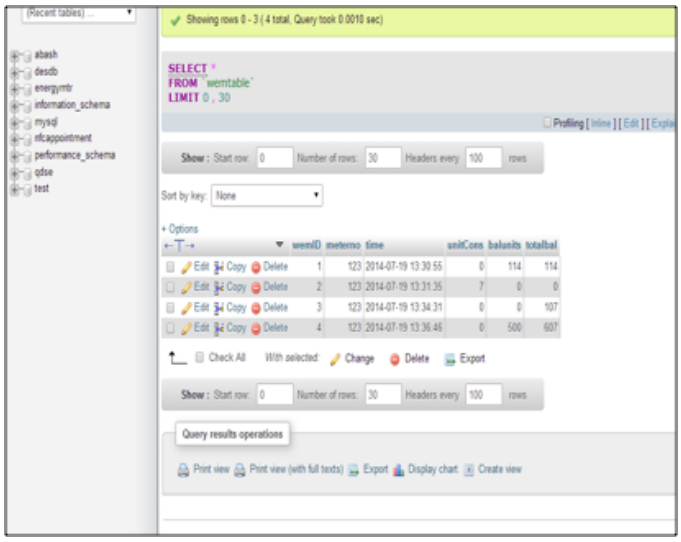


Fig. 2. Database tables for Wireless Smart Prepaid Meter

2. DES Backup Database

This database is for back up any data sent from the user Android application for the calculation of total top-up amount and from the wireless energy meter balance units. These values will only be displayed on DES main page every 5th of the month for monitoring purposes. The table contains these values as shown in Fig.3

- Meter number of each household.
- Timestamp
- Balance units
- Top-up amount
- Total top-up amount

IV. IMPLEMENTATION OF ENERGY METER SIMULATOR

Based on previous work [18] the energy simulator was implemented using JAVA language. Similar to daily energy consumption and current energy meter, the simulator have the features for peak and off-peak hour, the energy units that has been consumed as well as giving a graphical indication of the time and day.

The Energy simulator periodically generates the energy consumption units according to the data and information given by the DES and Singapore Power (Singapore utility provider) as shown in Table 1. Once the units are generated, the values are merged with the peak and off-peak hours given by utilities provider as shown in Fig 4.

In the wireless energy simulator graphical interface, the implementation are divided into two panels shown in Fig 5:

1. The device panel where all the daily household appliances are listed and the values are generated dynamically according to the peak hours and power. These values are saved to the WAMP server database for later use in JADE Agent applications.

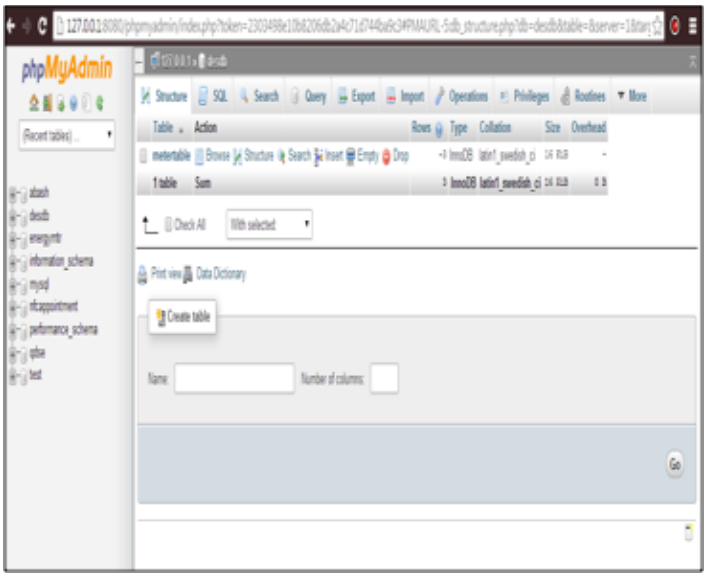


Fig. 3. Database for Department of Electrical Services (DES)

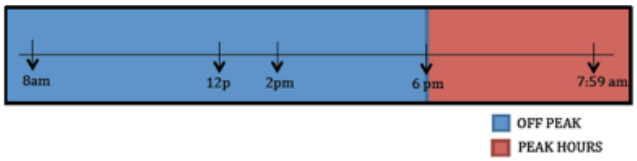


Fig. 4. Peak and Off-peak Hours

TABLE I. ENERGY CONSUMPTION UNITS OF DAILY HOUSEHOLDS APPLIANCES

APPLIANCES	UNITS CONSUMED DAILY IN KWH
Air-Conditioner	288
Aquarium	5
Ceiling Fan	21
Hair Dryer	24
Computer	27
Laptop	9
Lightings (3 units)	10
Radio	1
Refrigerator	96
Rice Cooker	54
Vacuum	59
Washing Machine	41

2. Graphical time-indication panel showing the night and day, off-peak and peak hours.

V. IMPLEMENTATION OF JADE AGENTS IN DES

The DES is responsible for the operation and development in generating, transmitting and distributing electricity to the end users. Hence the DES JADE agent serves as the main container of the agent system environment which has the authority and function to kill the agents and monitor the

transfer of data between all the agents in the system environment. Data such as balance units and total top up amount will go through the DES database to ensure up to date records of each household meter. These values will only be displayed in the DES main page every 5th of every month.

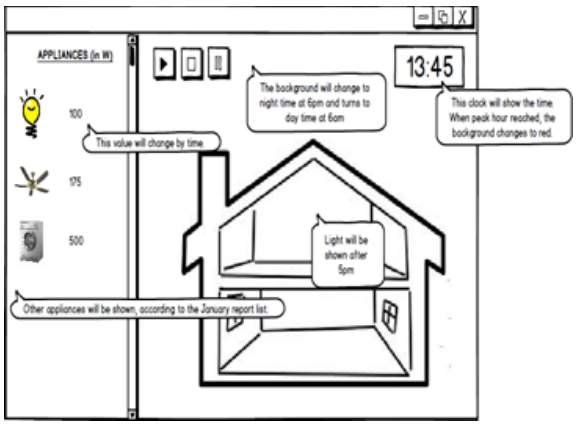


Fig. 5. Energy Meter Simulator

The database in the DES has the function to monitor and access the data given by other JADE agents in the environment such as balance units and top-up value every 5th of the month. To achieve the function, DES backup database will serve to store all the data before the accessed date. It contains only one table which hold the data of meter number, balance units, top-up amount and the total top-up amount to-date. The total top-up amount has the increment function whereby it increases as the user topped up. The DES JADE Agent will act as the main container for all the active agents in the system environment, whereby it has the feature of directory which announces available agents (Directory Facilitator) and controls the agents by creating and destroy the agents (Agent Management System) as shown in Fig 6.

Once the DES JADE Agent Management GUI as shown in Fig.6 is started, the DES main page will also initiate. The DES main page comprises of a table stating the meter number, balance units and total top-up amount to-date.

VI. Implementation of Smart Agent in Wireless Prepaid Energy Meter using JADE-LEAP

The function of the Wireless Prepaid Meter is to display the unit consumed and balance units. The units consumed data are extracted from the Energy simulator which generates energy units based on the usage of appliances at home. The balance units are calculated inside the Wireless energy meter using the given unit consumed by the energy simulator which will be highlighted later in the next section. Login information are required to enter the Smart Agent based Wireless Prepaid energy Meter to ensure the connection to the JADE main container and to provide minimal security to the energy meter.

A. Operation of Smart Agent based Wireless Prepaid Meter

The three main functions are:

- To generate units consumed from the wireless energy meter simulator and send the values to the DES agent and User Android application.
- To calculate the balance units of each household. The balance units are calculated as:

  Balance units = Total balance units after topping up-units consumed

- To send a reminder to User Android application once the balance units hits certain ranges.

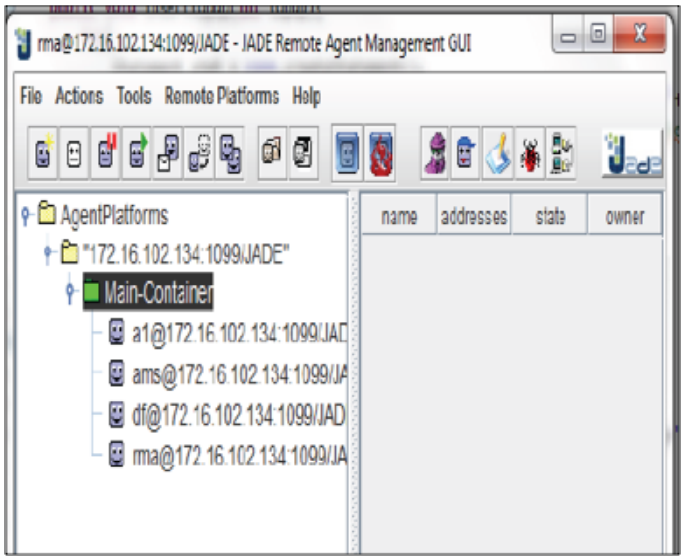


Fig. 6. DES JADE Remote Monitoring Agent

B. Implementation of Wireless Prepaid Meter using JADE-LEAP

The Smart Agent based Wireless Prepaid Meter is compiled and run on Android Emulator and login information is required for user authentication. Once logged in, the Agent based wireless prepaid meter can be connected to the DES main container by setting the energy meter IP to match with the IP address of the DES. Once this is set, the validation process begins.

Once the WAMP database and emulator are connected the unit consumed and balance units in the given period will be displayed on the main page. At the same time, it also communicates with the user Android application and transfers the balance units. These values will be pushed to the main container held by DES as shown in Fig.7. The refresh button will grab energy consumption units of the current time from the database as well as calculating the balance units.

VIII. Conclusion & Future Work

The introduction of prepaid meter (PM) has solved many issues for the utility provider and it also cuts down the hassle of having to visit households to read meters, which can be a big challenge. PM opens up many opportunities for further improvement. For example, providers do not have real time information on how energy units are consumed, balance units

for each household and how users spend on energy usage on a regular basis. And we proposed and developed a prototype Wireless Smart Power Meter using Agent technologies of the JADE and JADE-LEAP running on Android Jelly Bean mobile platform with Eclipse Juno IDE by creating three agents. We were able to demonstrate the capability of the services and the intelligence of this multi-agent environment of the proposed system. The system was able to transfer and query energy units in real time without the help of any human interactions, thus easing the monitoring process. In future the system proposes to integrate other features like sending reminders based on rate of energy units used. In addition computing energy units based on tariff rate, top up amount and taking into amount of energy used towards energy conservation. Last but not least adding the energy units to energy meter from energy department based on top up from users.

Future work will include agent from consumers' mobile devices, which will query the energy meter to study the power consumed, and for topping up the balance. When the meter reaches the threshold, agent at energy meter would also send messages to alert consumers for topping up through mobile devices and failing to do so will lead to power disconnection from the utility provider.

Fig. 7. Communication between Wireless Meter and User Android Application

REFERENCES

[1] Y. Merlin and P. E. Sankaranarayanan, "Accessing Remote Energy Meters Using Software Agents," Proceedings of IEEE Electro/Information Technology. Rochester, USA, 2001.

[2] C D. Suriyakala, and P.E. Sankaranarayanan, "Remote Accessing of Intelligent Energy Meter,"5th International Conference on Trends in Industrial Measurements and Automation, Tiruchirapally, TamilNadu, India, 2007

[3] C.D.Suriyakala and P.E. Sankaranarayan, "Intelligent agent system for accessing remote energy meters," Proceedings of 1st International conference Digital Communications & Computer Applications, Jordan, 2007.

[4] C.D. Suriyakala and P.E. Sankaranarayanan,"Smart Multi Agent Architecture for Congestion control to Access Remote Energy Meters," Proceedings of IEEE International Conference on Computational Intelligence and Multimedia Applications. Sivakasi, Tamilnadu, India, 2007.

[5] C.D. Suriyakala, "Studies on the Application of Software Agents for Accessing Energy Meter Data through Power Line Communication," Ph.d Thesis, Sathyabama University, Chennai, India, 2009

[6] R. Tahboub, D. Lazarescu and V. Lazarescu. "Modeling and Simulation of Secure Automatic Energy Meter Reading and Management Systems using Mobile Agents," International Journal of Computer and Network Security, Vol.7, No.1, pp.244-253, 2007.

[7] T. Jamil. "Design and Implementation of a Wireless Automatic Meter Reading System," Proceedings of world Congress on Engineering, London, UK, 2008

[8] C. Tatsiopoulos and A. Ktena, "A Smart ZIGBEE Based Wireless Sensor Meter System," 16th International Conference on Systems, Signals and Image Processing (IWSSIP), Chalkida, Greece, 2009.

[9] N. Kommu, P.Nagamani and M.Kollam, "Designing of an Automated Power Meter Reading with Zigbee Communication," International Journal of Computer and Communication Technology, Vol.2(7), pp.13-16, 2011

[10] A.H. Primicanta, M.Y Nayan and M Awan, "ZigBee-GSM based Automatic Meter Reading system," 2010 International Conference on Intelligent and Advanced Systems (ICIAS), Kuala Lumpur, Malaysia, 2010

[11] H.C. Chen and L.Y Chang, "Design and Implementation of a ZigBee-Based Wireless Automatic Meter Reading System," PRZEGLĄD ELEKTROTECHNICZNY (Electrical Review), 2012

[12] L. Q. Xi and Li. Gang. "Design of remote automatic meter reading system based on ZigBee and GPRS," Proceedings of Third International Symposium on Computer Science and Computational Technology(ISCSCT '10), Jiaozuo, P. R. China, 2010

[13] http://www.bt.com.bn/business-national/2011/12/12/installation-prepaid-electricity-meters-new-accounts-will-be-charged

[14] http://www.theborneopost.com/2013/03/13/30m-in-energy-savings-since-introduction-of-new-electricity-tariff/

[15] F. Bellifemine, G. Caire, G and D. Greenwood," *Developing multi-agent system with jade*. Chichester, UK: John Wiley & Sons, 2007.

[16] F. Bellifemine, G. Caire, A. Poggi and G. Rimassa, "Jade: A white paper. *EXP in Search of Innovation, 3*(3), 6-19, 2003.

[17] B.O.Omijeh and G.I.Ighalo. "Modelling of GSM based Energy Recharge Scheme for Prepaid Meter," IOSR Journal of Electrical and Electronics Engineering, Vol.4(1), 46-53, 2013

[18] Sankaranarayanan, S. and A.T. Wan. "*ABASH, Android based smart home monitoring using wireless sensors*." IEEE Conference on *Clean Energy and Technology (CEAT),* Langkawai, Malaysia, Nov 18-20 2013. Pp 494-499.

Towards A Hosted Private Cloud Storage Solution for Application Service Provider

Hsin Tse Lu, Chia Hung Kao, Po Hsuan Wu
Cloud System Software Institute
Institute for Information Industry
Taipei, Taiwan
Email: {oliu, chkao, paulwu}@iii.org.tw

Yi Hsuan Lee
School of Informatics
University of Edinburgh
Edinburgh, United Kingdom
Email: s1401456@sms.ed.ac.uk

***Abstract*—Private cloud storage solutions, such as Oxygen Cloud and COSA (Cloud Object Storage Appliance), which provide file synchronization and sharing with high performance and integral security for enterprises. Those solutions have well-defined functions for integrating with enterprise environment. However, not all enterprises have enough resource to deploy, maintain and manage their own services. On the other hand, application service providers are searching for efficient and automatic service based on these storage solutions for more profits. In this paper, we propose a hosted private storage architecture based on COSA (Cloud Object Storage Appliance) and CAKE (Cloud Application Kernel Environment). In this architecture, each appliance hosted in a data center with Internet connectivity, storage pool and hypervisor(s); and a portal machine can provide self-service and monitoring features to enterprise and application service provider. The design and the implementation of the hosted private cloud storage solution are also described in this paper.**

I. Introduction

Nowadays, since information increases rapidly, enterprises need a way to manage data efficiently. Public cloud storage services like Dropbox [1] and Google Drive [2] allow users to store their data remotely and access them at anytime via any devices; the service could even provide high performance and integral security if enterprises host their own service, such as Oxygen [3] Cloud and COSA (Cloud Object Storage Appliance) [4] [5]. However, when enterprises host their own cloud service, they need human resource to manage the network connectivity, power consumption, machine location and availability; they also need a system administrator with information technology knowledge, such as backup restore solution, log archiving and E-mail service, for filling enterprises' SLA (service level agreement). On the other hand, to increase more profits, efficiency of service delivery and to reduce [6] [7], the traditional application service provider transforms the network-delivered business service from customize an application for individual customer to shared service model.

As what mentioned in pervious section, public cloud storage service offers great functionality let end users to manage their data cross multiple devices. However, many enterprises does not allow their employee access the public cloud storage service in the office network cause the security issue. There are many researches explained the security issue on the public cloud service [8] [9]. In the enterprise deployment, the enterprise administrator needs functions such as: auditing, users and groups management, control privileges and security access.

In this paper, to address issues from enterprises and application service providers, the cloud storage service should be transformed to logic-isolation approach [10] [11] [12] of multi-tenant architecture and hosted by service provider. To that end, the service should be a network-delivered, pre-build and self-configurable service.

II. Component in the Multi-Tenancy Architecture

As shown in Fig 1, we put serval components together to build a hosted private storage architecture for multi-tenant service.

A. Application: COSA (Cloud Object Storage Appliance)

COSA provides EFSS (Enterprise File Sharing and Synchronize) function, it is a network-delivered application. Users can store their data remotely and can access them at anytime from any devices. COSA is also a virtual file system, which allows users operate files with versioning, recycle bin, share link and additional features.

B. Computer: CAKE (Cloud Application Kernel Environment)

CAKE [5] is a KVM-based hypervisor, which offers an environment for hosting virtual machines. It is a network-delivered application and also exposes virtual machine management API to allow a third-party software to control the lifecycle of the virtual machine. CAKE also provides high availability and migration feature for filling enterprises' SLA requirements.

C. Networking

Network contains three different layers, including WAN layer, DMZ layer and LAN Layer.

1) WAN Layer: WAN layer includes first tier firewall and connection to ISP (Internal Service Provider), all end users connect to their application service from Internet through serval public IP addresses provided by ISP. The application service provider could control the overall network bandwidth and configure the network port policy for filling individual customer's need in this layer.

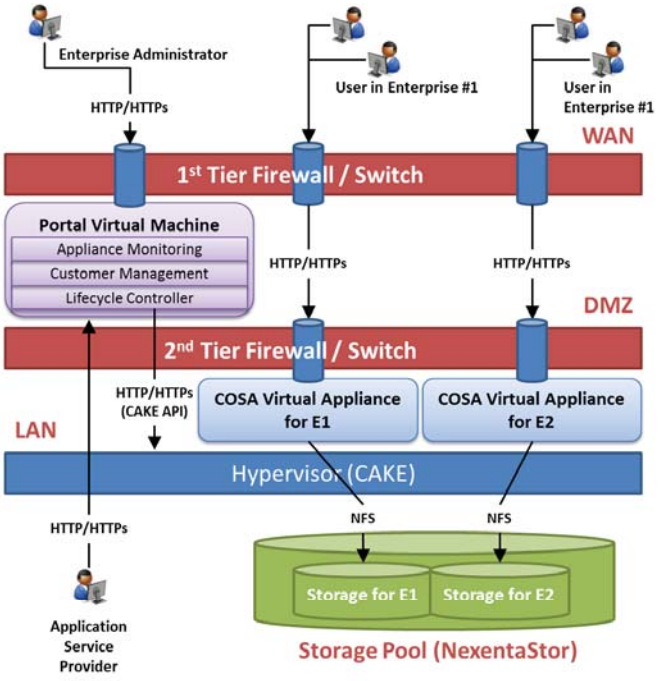


Fig. 1. Architecture of the Multi-tenant Cloud Storage Service

2) DMZ Layer: DMZ layer includes second tier firewall, it provides connection from WAN to LAN. The application service provider could control the network's bandwidth by customer in this layer. The port policy in this layer will only allow specific HTTP/HTTPs traffic for end users.

3) LAN Layer: For the security reason, shared component, such as computer node and storage pool, which does not allow users access directly should locates in this layer.

D. Storage: NexentaStor

NexentaStor [13] is an enterprise class SDS (Software-Define-Storage), which provides functionality and capability such as SSD/HDD hybrid configuration for high performance, high availability, capacity scalability, disaster recovery, block-level encryption and provisioning. It enables application service provider easily manage storage hardware and provide file storage for application store the physical data.

III. Design and Implementation of the Muti-tenancy Architecture

As shown in Fig 1, to approach the multi-tenant architecture, there are two virtual appliances on top of the CAKE.

A. COSA Virtual Appliance

When enterprises host COSA themselves, the application service provider will deploy all COSA modules, including web container and database, into one physical machine with amount HDD storage; the whole management function will be enabled in this case, such as power management and system monitoring. To approach the multi-tenant architecture, the application service provider separates web container and database to COSA virtual machine and storage pool, service states won't be leaved into COSA virtual machine; and also disables serval service management functions, such as power management. After the implementation, application service provider translates the virtual machine as a template and stores it at the storage pool.

B. Portal Virtual Machine

For efficient and automatic service goal, there are three major functions in Portal Virtual Machine.

1) Appliance Monitoring: Appliance Monitoring function provides information such as storage usage and network bandwidth to enterprise's administrator and application service provider. This function is integrated with SNMP agent which hosts at firewall and COSA Virtual Machine.

2) Customer Management: Customer Management provides self-registration function to enterprises which require a cloud storage service, and also lets application service provider manage their customer list. Application service provider can disable service by customer in this function if there is any abnormal usage notification.

3) Lifecycle Controller: Lifecycle Controller has integrated with CAKE API, NexentaStor API and firewall SNMP agent. When an enterprise require a cloud storage service, Lifecycle Controller will follow four steps to enable the service: (1) Bootup COSA virtual appliance from VM template, (2) Configure firewall policy and IP address for the COSA virtual appliance, (3) create a disk volume on NexentaStor, and (4) publish the COSA virtual appliance information on the Portal Virtual Machine let customer start to use.

IV. A Case Study: Hosted Private Cloud Storage in Real World

The first hosted private cloud storage was deployed in III (Introduction of Institute for Information [14]) since 2011. There are two departments engaged into this show case, the first one is III MIS, they are the application service provider; the second one is III CSSI (Cloud System Software Institute), they are end users and also the enterprise administrator. The COSA service was hosted in III IDC (Internet Data Center) and deployed manually by CSSI engineer. The COSA service was installed as a virtual appliance over CAKE and connect to a traditional NAS through NFS protocol. The ISP (Internet Service Provider) provides a public IP address for this service which allowed end users and the enterprise administrator connect to the service via Internet.

A. End User Scenario

COSA is an object storage and provides EFSS(Enterprise File Sharing and Synchronize) function to end user [4]. in this show case, most end users install COSA desktop agent in their Windows and MAC devices for access the file and folder. For the usability issue, user will install COSA mobile APPs on their iOS and Android device to access the file and folder. In order to enhance the office collaboration performance,

Fig. 2. COSA Management Dashboard

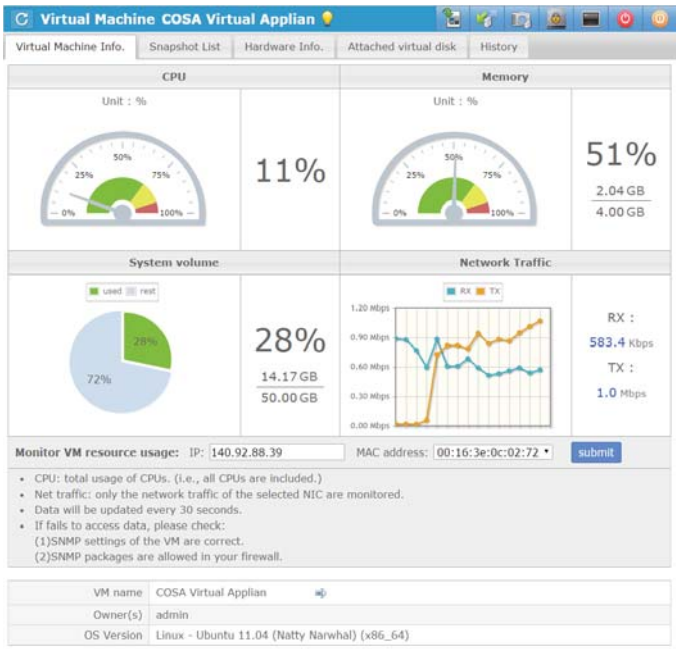


Fig. 3. CAKE Virtual Appliance Monitoring

the enterprise administrator enabled the share permission to everyone, the user can share the file and folder to anyone whatever if the receiver have COSA account or not.

B. Enterprise Administrator Scenario

As shown in Fig 2, it is the COSA management dashboard, which provides functionality for enterprise administrator to manage user account and permission control. Enterprise administrator can get service information such as storage capacity, user's active log and status through these dashboard.

C. Service Provider Scenario

As shown in Fig 3, it is the CAKE monitoring dashboard. The performance of COSA service can be monitored by application service provider, which measure the service's network bandwidth, CPU and memory. Application service provider can also enable / disable service through the virtual appliance management interface.

V. Conclusion

The logic-isolation approach provides advantage of efficiently and automatically service for saving application service provide costs. This solution was designed and implemented to enable the hosted storage cloud service for enterprise to manage data efficiently and for application service provider to create new business mode. Future work includes design and implementation of billing system, workflow and metering.

Acknowledgment

This study is conducted under the "Cloud computing systems and software development project(3/3)" of the Institute for Information Industry which is subsidized by the Ministry of Economy Affairs of the Republic of China .

References

[1] Dropbox. A public cloud storage service. [Online]. Available: http://dropbox.com
[2] Google. A public cloud storage service. [Online]. Available: https://drive.google.com
[3] A cloud storage service for private cloud solution. [Online]. Available: http://home.oxygencloud.com
[4] C. H. Kao and S. T. Liu, "A prototype system for object management in private cloud," in *Cloud and Service Computing (CSC), 2011 International Conference on*, Dec 2011, pp. 348–353.
[5] G. Wang and J. Unger, "A strategy to move taiwan's it industry from commodity hardware manufacturing to competitive cloud solutions," *Access, IEEE*, vol. 1, pp. 159–166, 2013.
[6] S. H. Kim and D. Kim, "Multi-tenancy support with organization management in the cloud of things," in *Services Computing (SCC), 2013 IEEE International Conference on*, June 2013, pp. 232–239.
[7] X. Zhang, P. Sun, Y. Huang, and W. Sun, "A model-driven framework for enabling self-service configuration of business services," in *Web Services, 2008. ICWS '08. IEEE International Conference on*, Sept 2008, pp. 497–504.
[8] K. Ren, C. Wang, and Q. Wang, "Security challenges for the public cloud," *Internet Computing, IEEE*, vol. 16, no. 1, pp. 69–73, Jan 2012.
[9] L. Liu, R. Moulic, and D. Shea, "Cloud service portal for mobile device management," in *e-Business Engineering (ICEBE), 2010 IEEE 7th International Conference on*, Nov 2010, pp. 474–478.
[10] Q. Shen, X. Yang, X. Yu, P. Sun, Y. Yang, and Z. Wu, "Towards data isolation #x0026; collaboration in storage cloud," in *Services Computing Conference (APSCC), 2011 IEEE Asia-Pacific*, Dec 2011, pp. 139–146.
[11] M. Factor, D. Hadas, A. Hamama, N. Har'el, E. Kolodner, A. Kurmus, A. Shulman-Peleg, and A. Sorniotti, "Secure logical isolation for multi-tenancy in cloud storage," in *Mass Storage Systems and Technologies (MSST), 2013 IEEE 29th Symposium on*, May 2013, pp. 1–5.
[12] W. Lloyd, S. Pallickara, O. David, J. Lyon, M. Arabi, and K. Rojas, "Service isolation vs. consolidation: Implications for iaas cloud application deployment," in *Cloud Engineering (IC2E), 2013 IEEE International Conference on*, March 2013, pp. 21–30.
[13] nexenta. (2014, jun) A software define storage service. [Online]. Available: http://www.nexenta.com/products/nexentastor
[14] A non-governmental organization in taipei, taiwan. [Online]. Available: http://web.iii.org.tw

Performance Analysis of Parallel Smoothed Particle Hydrodynamics on Multi-core CPUs

Chen Wenbo, Yucheng Yao, Yang Zhang
School of Information Science and Technology, Lanzhou University
Lanzhou, China
{chenwb, yaoych12, zhyang}@lzu.edu.cn

***Abstract*—This paper presents a parallel SPH implementation on multi-core CPUs. The implementation uses a hash table to store particles data and divides the program code into 2 parts for parallelization. The first part has no data race, but the second part has data race. Then, the paper compares the running time and parallel speedup of each part to find the bottleneck of the parallel SPH program. The results show that the program can achieve linear speedup just with the first part to be parallelized when the search radius is large. And the second part has become a performance bottleneck only when the search radius is small enough (for each cell only contains one or two particles on average). We present a method to parallelize the second part without affecting the performance of the first part. The results show that our method can ease the performance bottleneck when the search radius is small.**

Keywords—SPH; Multi-core CPU; Neighbor Search; SMP; Parallel

I. INTRODUCTION

Smoothed Particle Hydrodynamics method is a Lagrangian approach to simulate complex motion of fluid by using large amount of particles. It was first proposed to simulate astrophysical phenomenon by Bob Gingold, Joe Monaghan and Lucy [1]. The SPH method has been successfully applied to many fields, such as computational fluid dynamics, materials science, heat conduction, etc.[2-4]. Meanwhile, the SPH method requires a large number of particles to simulate physical phenomena accurately, which needs very high CPU performance.

In the past few decades, the number of transistors integrated on a chip increasing in accordance with Moore's Low, which due the frequency of CPU to double every 18-24 months. Thus, the same program will run faster with CPU updating. But now, the CPU technology has entered the multicore era for the power wall [5]. If you want your program to get improved performance from multi-core CPU upgrade (for the number of cores increases), the programmer needs to use explicit parallelization technology in the development stage [6].

Therefore, the parallelization for the SPH method has great significance. The major problem in the SPH parallelization is the performance loss caused by data race in multi threads. The data race is caused by accessing shared data with modification operation in multi threads environment. This is also called threads safety which needs to use the locking mechanism of the operating system. But thread blocking lock would due to more loss of performance. In multicore system, the spin lock is an alternative [7]. Another technique called Lock-Free just uses atomic operation to guarantee exclusive access to shared data [8]. Whether it is a spin lock or Lock-Free method, the program can't avoid atomic operations. That means the threads still need to race in the atomic level which would reduce the efficiency of CPU's cache mechanism.

This paper presents a parallel SPH program based on hash table. The implementation does not use any lock or atomic operation to protect the shared data, thereby avoiding performance loss by threads race. We firstly divided the program code into 2 parts for parallelization. The first part has no data race, but the second part has data race. Then, we found that the first part was with high load and the second part with low load. Our strategy is priority to parallelize the high load part. And the parallelization of low load part is depends on the situation.

II. RELATED WORK

There have been many studies about acceleration technology for SPH. Adams et al. used particles with different radii to construct an adaptive SPH simulation [9]. Their method could reduce the number of particles needed which greatly improved the performance of dense fluid's simulation. Considering the radii variable, they used a kd-tree as the spatial partitioning scheme for neighbor search. As a result, their method needs to rebuild the kd-tree in each simulation time step.

Although SPH has a natural parallel features with itself, the memory updating caused by moving of particles will cause data race in parallel or concurrent threads. Amada et al. used GPU to accelerate the particle-based fluid simulation. They implemented the neighbor search on the CPU, and used the GPU to calculate SPH computing. This is a very good idea. But the current GPUs cannot access the host memory directly, which led to data transfer between the host and the device through the PCI bus in each time step. Using GPU to accelerate SPH has been a research focus. And the mainstream approach to deal data race is sorting the particles in each time step [11-13].

The difficulty of implementing SPH on SMP and multi-core system also lies in efficient neighbor search and data race. Wróblewski et al. compared the performance of SPH based on OpenMP and MPI [14]. Their results show that MPI achieve higher efficiency. In their model, a uniform grid is used for neighbor search. And they did not use a hash table to store data. Therefore, their system is more suitable for uniform particle distribution scenarios. Markus et al [15] implemented a multi-core SPH system which avoids the data race by sorting the particles in each time step. The sorting method can save memory, but with additional computing resource.

Sven Ganzenm et al. have presented a parallel object-oriented framework for particle simulations written in C++. Their implementation of parallel I/O improved the performance greatly [16]. In the work of David et al., a parallel numerical simulation framework has been presented. They presented a domain distribution methodology which can take advantage of shared memory machine [17].

In the paper of Prashant et al., they introduced a parallel and interactive SPH simulation and rendering method on the GPU [18]. Their neighbor search method is based on Z-indexing and sorting method on the GPU. Brandon et al. presented a method to accelerate SPH by adaptively constructing and reusing particle pairing information [19].

III. SPH MODEL

SPH-based interpolation is a grid-free method. Its basic idea is to transform the equation of continuous physical motion into interpolation and integral of discrete particles. For example, to calculate the value of function *f(r)*, you can define a kernel function *W (|r-r' |, h)* as the weight of the value *f (r')* at the position *r'*. And then, you can use the values of function f in the neighbor field of r to estimate the value of *f(r)* using the following equation.

$$< f(r) >= \int f(r')W(|r-r'|,h)dr' \qquad (1)$$

And it is discretized as follows:

$$< f_i >= \sum_{j=1}^{n} f_j \frac{m_j}{\rho_j} W(|r_i - r_j|,h) \qquad (2)$$

In the equation (2), m_j stands for neighbor particle's mass, ρ_j for the density, h for the size of neighbor field. Mostly, the neighbor field is a circle area with radius *2h* as showing in Fig. 1.

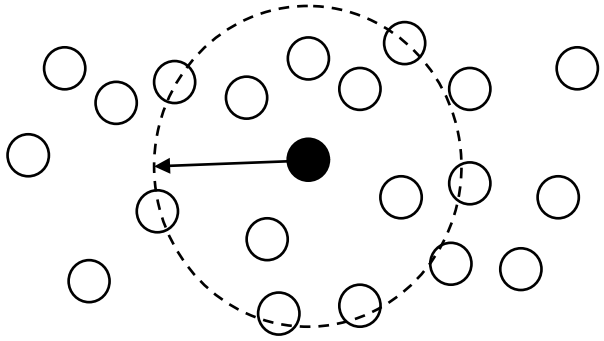


Fig. 1. Neighbor field of a particle

The interpolation kernel function $W(|r-r'|,h)$ is also called smoothing function which has a great influence on the simulation result. The mostly used interpolation kernel function is the B-spline interpolation function. It is described as the following equation (3).

$$W(|r-r'|,h) = \frac{1}{\pi h^3}\begin{cases} \left(1-\frac{3}{2}\gamma^2\right)+\frac{3}{4}\gamma^3 & 0<\gamma<1 \\ \frac{1}{4}(2-\gamma^3) & 1\le\gamma<2 \\ 0 & otherwise \end{cases} \qquad (3)$$

In SPH, the gradient of a function can be obtained easily using the gradient of the interpolation kernel function, as show in the following equation (4).

$$< \nabla f_i >= \sum_{j=1}^{n} f_j \frac{m_j}{\rho_j} \nabla W(|r_i - r_j|,h) \qquad (4)$$

This paper uses the famous Neville - Stokes equation as the control function of fluid motion. The density is calculated with the basic SPH interpolation method.

$$\rho_i = \sum_j m_j W_{ij} \qquad (5)$$

The pressure is calculated with concerning of symmetry.

$$f_i^{pressure} = -\sum_j m_j \frac{p_i + p_j}{2\rho_j} \nabla W_{ij} \qquad (6)$$

And the viscous force is calculated using the following interpolation function.

$$f_i^{viscosity} = \mu \sum_j m_j \frac{v_i - v_j}{\rho_j} \nabla^2 W_{ij} \qquad (7)$$

IV. NEIGHBOR SEARCH

SPH is based on particle system. To calculate particles' physical property, it needs to search their neighborhood.

There are many method to do neighbor search. This paper uses the uniform grid space partitioning method to do neighbor search, which is widely used. The basic idea of uniform grid method is to divide the whole space into N*N child cells with the same size. So that, the neighbor search for

a particle is limited in a few cells (27 cells in 3D scene) instead of the whole space. Figure 2 shows the searching field in a 2D scene. In order to find out all the neighbor particles of the black particle, we just need to traverse the neighboring gray cells.

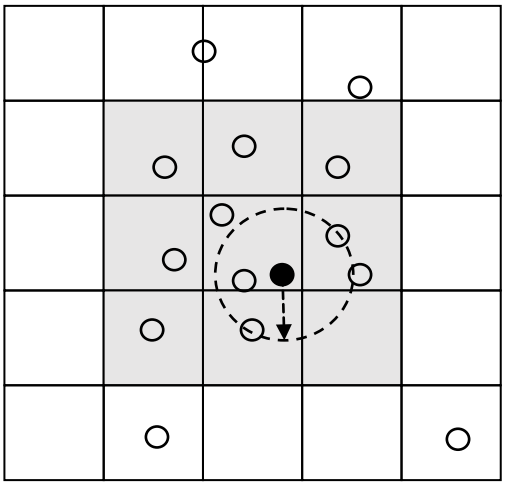


Fig. 2. Neighbor search for a particle

With uniform grid, we can find particles with specific location quickly. In our implementation, each cell is marked with an integer ID generated by its space location. And the particles in a cell are organized as a linked list. The uniform grid could be implemented as an array or a hash table. With an array, we could gain faster running speed but with more memory needed. This paper uses a hash table to store cells by using cell IDs as hash keys.

To find neighbors of the particle P, we can do like this:

```
1   set <x, y, z> = cell_coordinate(P.position)
2   set S as a particle who's each property is 0
3   for ix = -1 to 1 do
4       for iy = -1 to 1 do
5           for iz = -1 to 1 do
6               K = cell_id(<x+ix, y+iy, z+iz>)
7               L = map(K)
8               S += doSPH(P, L)
9           end for
10      end for
11  end for
12  return S
```

The function cell_coordinate(r) in the pseudo-code is to calculate the coordinate of the cell which contains the position r. The function cell_id(r) is used to calculate the cell ID through its coordinate. The function map(K) returns the linked list which contains all the particles in the cell marked by K. And The function doSPH(P, L) returns a partial value of the SPH interpolation result.

V. PARALLELIZATION

Our implementation is a simple simulation of fluid. It is based on time iteration which looks like this:

```
1 for t<T do
2   for each particle P do
3           do the neighbor search and SPH interpolation for P.
4   end for
5   for each particle P do
6           if P's cell ID is changed
7                   move P from the current linked list to the right list.
8           end if
7    end for
8    t += Δt
8 end for
```

In the outer for-loop, in each iteration the progress reads the data which is updated in the previous iteration. So in this loop, the data is dependent and the code cannot be parallelized. The parallelization is applied on the inner for-loop.

To parallelize the inner loop, we need to handle the data race. Even though the data of each particle is interpolated from its neighbor particles, the computing process of each particle can run independently. Because each process updates only one particle which it right handles. The problem is on moving particles from one cell to another. As a particle moving, it may leave from the current cell and enter into a new cell. In the simulation program, this is mapped to moving data from one linked list to another. The data race occurs when two threads try to access the same linked list with writing operation.

To avoid data race, we see the code in the inner loop as two parts. The first part does the neighbor search and SPH interpolation without moving data. The second part moves data to the right lists. The first part is just the first inner for-loop which can be parallelized easily. The second part is the next inner for-loop with data race. The parallel pseudo-code for the first part looks like this:

```
1   parallel for each particle P do
2           do the neighbor search and SPH interpolation for P.
3   end parallel for
```

The "parallel" keyword indicates each iteration of the loop can be run in parallel.

A simple way to parallelize the second part is to search all the lists in parallel, and then moving the data in a single thread. There is no problem if the linked list references are stored in an array. It is easy to access data in an array by indexes. But our implementation uses a hash table. A hash table stores data as key-values. You may have no way to access a value by absolute index for different data may mapped to a same

position. In our program, the hash table used is provided by the gnu c++ extension library. This hash table has no way to access the value by a position index in the memory. Our parallel pseudo-code looks like this:

```
1 parallel for each particle P do
2       calculate the new_key from P's new position.
3       if old_key != new_key do
4               old_list = map(old_key);
5               new_list = map(new_key)
6               old_list.remove(Pi);
7               new_list.add(Pi);
8       end if
9 end parallel for
```

VI. Experiment Result

Our program was tested on a SMP machine with two Intel® Xeon® CPU X5650 processors. The Intel® Xeon® CPU X5650 CPU has 6 cores with 12 threads. The platform has 12 cores with 24 threads. The simulation space size is set to 1.0 * 1.0 * 1.0. The test used 1 million particles with 1000 time iterations.

Firstly, we tested the sequence code with different search radius to find its influence on the performance. We gathered the running time (hours) of each part. The results are shown in Fig. 3.

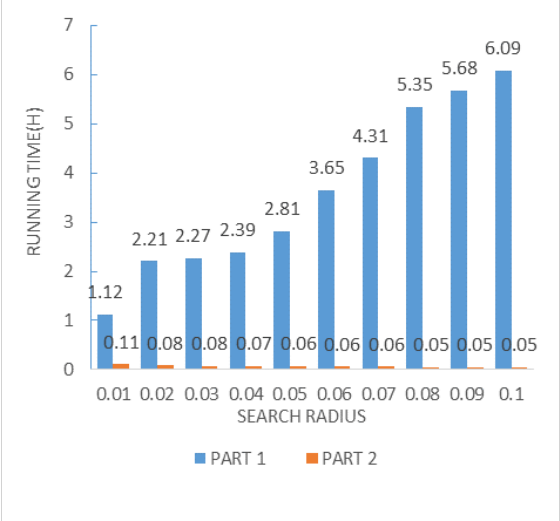


Fig. 3. Running time of sequence code with different radii

In our SPH implementation, the resolution of the grid is decided by the search radius R. The size in each dimension is 1.0/ R. So, the larger the search radius is, the smaller the grid resolution, the larger each cell, the more the number of particles in each cell. As a result, the SPH interpolation calculation time of part 1 increases. From the Fig. 3, we can see that only the running time of part 1 increases as the search radius grows bigger. The increasing is obvious when the radius changes from 0.01 to 0.02. The value almost doubles. We think even though the average particles' count is 1 when the search radius equals 0.01, the actual situation is that there are many empty cell. In the program, the empty cell is ignored without any executing time.

Another feather is that the running time of part 1 is just opposite compared with part 1. But it just declined a little as the search radius increases. The most important thing is that the running time of part 2 is much short than part 1. And the gap increases as the search radius grows bigger. We think the part 2 should not be a bottleneck when the search radius is bigger enough. To prove the opinion, we then tested the program only with part 1 parallelized as well as set the search radius to be 0.1. Fig. 4 shows the running time of part 1 and part 2 with different number of parallel threads. Fig.5 shows the parallel speedup of part 1 and the overall program.

As shown in Fig.4, the running time of part 2 is much short than part 1. Fig. 5 shows that the overall program's parallel speedup is as well as part 1. Even though when the threads' number increases to 24, the overall speedup is 17.18 which is a little smaller than part 1's 19.80. The results show that we can gain a nice speedup only with parallelization on part 1 as the search radius is relatively big.

Then, we choose the search radius of 0.01 and tested again with only part 1 parallelized. The results are shown in Fig. 6 and Fig. 7. The running time of part 2 is also an important fact influenced the performance as shown in Fig. 6. It is clearer as shown in Fig. 7 that the speedup of overall program is only half of the speedup of part 1 when the threads' number is more than 12. But it also shows that the speedup of part 1 increases little when the number of threads is bigger than 12. In fact, when the search radius becomes relatively smaller, the average number of particles in each cell also becomes smaller. As a result, the executing time of numeric computing of SPH interpolation declines, but the time of accessing the hash table increases, which leads to the CPU accesses the memory more frequently. It seems that this will lead to Intel's Hyper-Threading technology has no effect.

Finally, we parallelized the code of part 2. The results are shown in Fig. 8 and Fig. 9. It shows that the running time of part 2 also declines as the threads' number increases. Fig. 9 shows that our method to parallelize part 2 really can ease the performance bottleneck.

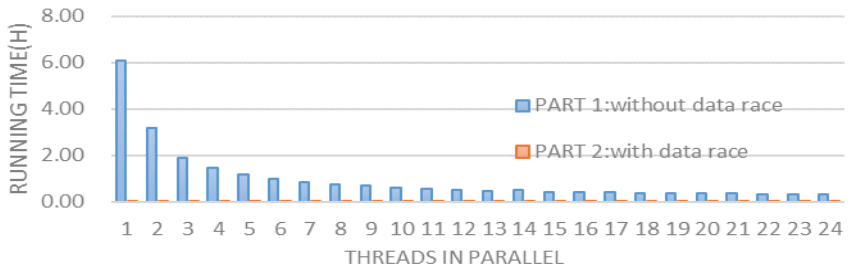


Fig. 4. Running time of parallel code with search radius of 0.1 (only with part 1 parallelized)

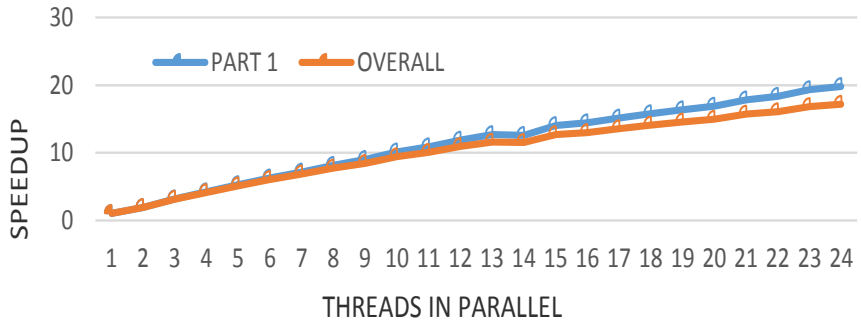


Fig. 5. Parallel speedup with search radius of 0.1 (only with part 1 parallelized)

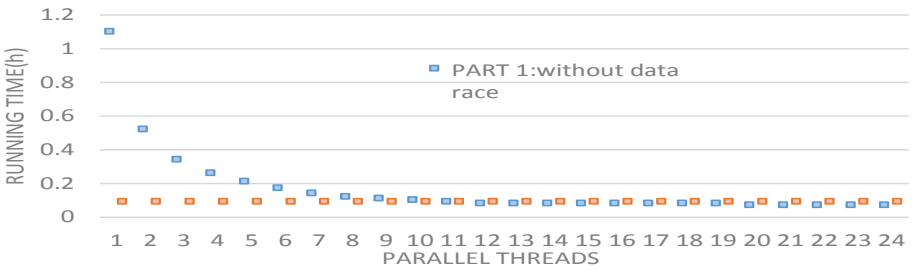


Fig. 6. Running time of parallel code with search radius of 0.01 (only with part 1 parallelized)

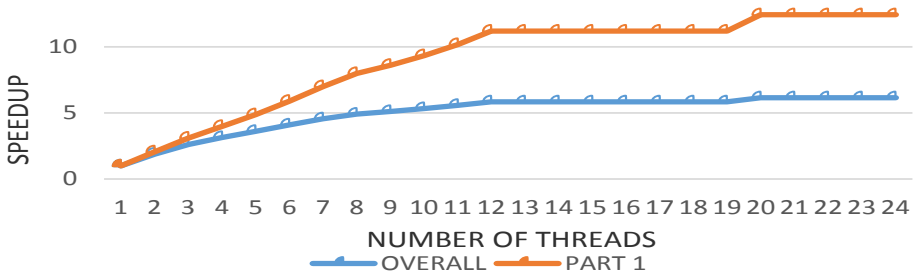


Fig. 7. Parallel speedup with search radius of 0.01 (only with part 1 parallelized)

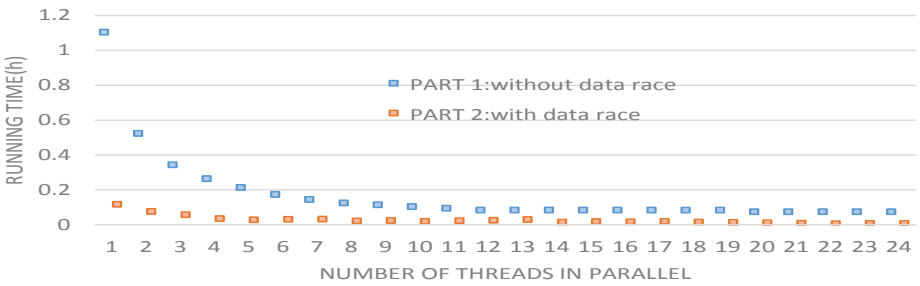


Fig. 8. Running time of parallel code with search radius of 0.01

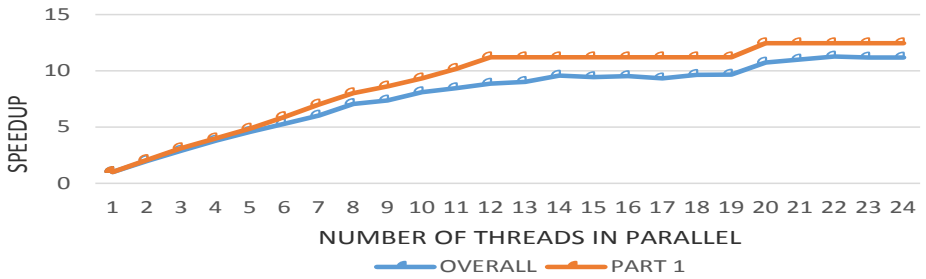


Fig. 1. Fig. 9. Parallel speedup with search radius of 0.01

VII. CONCLUSION

The results show that it is enough to achieve a nice speedup only with the first part parallelized in most situation. However, the second part will be a bottleneck if the search radius is small enough to achieve one or two particles in a cell on average. The method to parallelize the code in second part is effective if the search radius is really small.

As the results shown, the SPH program may not gain benefit from the Hyper-Threading technology if the search radius is small.

The method to parallelize the code of second part has a flaw. It needs to search a linked list to find the right particle which will waste a lot of time when the list is large. In our test, the search radius is small, which means the linked lists are short on average. The flaw is not a problem with short lists. But the method cannot apply to the situation with a big search radius, or it will slow down the overall program. In the future, we want to find a method of the best of both worlds.

ACKNOWLEDGMENT

This paper is supported by the Network Center of Lanzhou University. Thanks for the Network Center to provide the experimental equipment.

REFERENCES

[1] R. A. Gingold and J. J. Monaghan, "Smoothed particle hydrodynamics: Theory and application to non-spherical stars," Mon. Not. Roy. Astron. Soc., vol. 181, p. 375, 1977.

[2] M. Ihmsen, J. Cornelis, B. Solenthaler, C. Horvath, and M. Teschner, "Implicit Incompressible SPH," Visualization and Computer Graphics, IEEE Transactions on, vol. 20, pp. 426-435, 2014.

[3] C. Liu, J. Zhang, and Y. Sun, "The optimization of SPH method and its application in simulation of water wave," in Natural Computation (ICNC), 2011 Seventh International Conference on, 2011, pp. 2327-2331.

[4] R. C. Batra and G. M. Zhang, "Modified Smoothed Particle Hydrodynamics (MSPH) basis functions for meshless methods, and their application to axisymmetric Taylor impact test," Journal of Computational Physics, vol. 227, pp. 1962-1981, 1/10/ 2008.

[5] G. Pam Frost, "Multicore Processors for Science and Engineering," Computing in Science & Engineering, vol. 9, pp. 3-7, 2007.

[6] A. Ebnenasir and R. Beik, "Developing parallel programs: A design-oriented perspective," in Multicore Software Engineering, 2009. IWMSE '09. ICSE Workshop on, 2009, pp. 1-8.

[7] T. Johnson and K. Harathi, "A prioritized multiprocessor spin lock," Parallel and Distributed Systems, IEEE Transactions on, vol. 8, pp. 926-933, 1997.

[8] M. M. Michael, "High performance dynamic lock-free hash tables and list-based sets," presented at the Proceedings of the fourteenth annual ACM symposium on Parallel algorithms and architectures, Winnipeg, Manitoba, Canada, 2002.

[9] B. Adams, M. Pauly, R. Keiser, and L. J. Guibas, "Adaptively sampled particle fluids," ACM Trans. Graph., vol. 26, p. 48, 2007.

[10] T. Amada, M. Imura, Y. Yasumuro, Y. Manabe, and K. Chihara, "Particle-based fluid simulation on gpu," in ACM Workshop on General-Purpose Computing on Graphics Processors and SIGGRAPH, 2004.

[11] K. Hegeman, N. Carr, and G. P. Miller, "Particle-Based Fluid Simulation on the GPU," in Computational Science – ICCS 2006. vol. 3994, V. Alexandrov, G. Albada, P. A. Sloot, and J. Dongarra, Eds., ed: Springer Berlin Heidelberg, 2006, pp. 228-235.

[12] A. Hérault, G. Bilotta, and R. A. Dalrymple, "SPH on GPU with CUDA," Journal of Hydraulic Research, vol. 48, pp. 74-79, 2010/01/01 2010.

[13] P. Goswami, P. Schlegel, B. Solenthaler, and R. Pajarola, "Interactive SPH simulation and rendering on the GPU," presented at the Proceedings of the 2010 ACM SIGGRAPH/Eurographics Symposium on Computer Animation, Madrid, Spain, 2010.

[14] P. Wróblewski and K. Boryczko, Parallel Simulation of a Fluid Flow by Means of the SPH Method: OpenMP vs. MPI Comparison, 2012.

[15] M. Ihmsen, N. Akinci, M. Becker, and M. Teschner, "A Parallel SPH Implementation on Multi-Core CPUs," Computer Graphics Forum, vol. 30, pp. 99-112, 2011.

[16] S. Ganzenm, S. Pinkenburg, and W. Rosenstiel, "SPH2000: a parallel object-oriented framework for particle simulations with SPH," presented at the Proceedings of the 11th international Euro-Par conference on Parallel Processing, Lisbon, Portugal, 2005.

[17] D. W. Holmes, J. R. Williams, and P. Tilke, "A framework for parallel computational physics algorithms on multi-core: SPH in parallel," Adv. Eng. Softw., vol. 42, pp. 999-1008, 2011.

[18] P. Goswami, P. Schlegel, B. Solenthaler, and R. Pajarola, "Interactive SPH simulation and rendering on the GPU," presented at the Proceedings of the 2010 ACM SIGGRAPH/Eurographics Symposium on Computer Animation, Madrid, Spain, 2010.

[19] B. Pelfrey and D. House, "Adaptive neighbor pairing for smoothed particle hydrodynamics," presented at the Proceedings of the 6th international conference on Advances in visual computing - Volume Part II, Las Vegas, NV, USA, 2010.

The Realization of Green Storage in Hadoop

Qiao Zhu
Department of Computer Science and Engineering, Hunan University
Changsha, China
zhuqiao641127@163.com

Li Miao
Department of Computer Science and Engineering, Hunan University
Changsha, China
miaoli2000@163.com

***Abstract*—Hadoop has been successful at harnessing expansive data-centers resources for large-scale data analysis. However, their effect on data-centers energy efficiency has not scrutinized completely. The energy consumption of Hadoop Distributed File System in data-centers accounts for a great part of total cost of ownership and disk is the main storage media for clusters. Analysis of the interactions between clusters with disks when running a Hadoop application showed a disk idle time when shuffle to memory, which can be used to guide a simple green storage algorithm for Hadoop cluster. The algorithm simulation results with Terasort for 10G data in a cluster with 11 nodes can save 2.47WH.**

Keywords—Disk, Energy, Green storage, Hadoop.

I. INTRODUCTION

Few technologies have gotten more attention over the past few years than Hadoop,there are three main reasons for this: First, from a volume perspective, many tools start being impractical when thresholds in the dozens of terabytes are exceeded. Second, from a variety perspective, traditional analytic tools only work well with structured data, which represents, at most, 20 percent of the data in the world today [1]. Finally, there is the issue of how fast the data are arriving at your organization's doorstep—Big Data velocity. Considering the pressing need for technologies that can overcome the volume and variety challenges for data at rest, it is no wonder that Hadoop become the main tool for large-scale data analysis and is widely used in data-centers, social media analysis, log analysis and other big data applications.

Much of the work in building Hadoop was done by Yahoo!, which reportedly has over 40,000 nodes spanning its Hadoop clusters and can store over 40PB of data [2]. While in the pursuit of extreme performance, the trade-off between performance and power consumption needs to be considered. In addition, the storage power consumption constitutes a significant part of the TCO (total cost of ownership) of data-centers. Hence, the energy-conservation of the large-scale cluster has become a priority. Here we consider how to improve energy efficiency in Hadoop cluster.

Hadoop, a classic implementation of MapReduce, which has many characteristics for energy efficiency. First, MapReduce frameworks implements a distributed data-store composed of the disks in each node, which enables affordable storage for multi-petabyte data-sets with good performance and reliability. So it needs to ensure data availability. There is significant amount of research literature about improving Hadoop energy efficiency, such as [3], which found that disable some nodes totally can save energy. Although it is a truly way to save energy, these benefits are based on decreasing performance, which means the running time of task will increase when taking off some nodes . [4] presented GreenHDFS, an energy-conserving, hybrid, logical multi-zoned variant of Hadoop's compute cluster, which simulation results show that GreenHDFS is capable of achieving 26% savings in the energy costs of a Hadoop cluster in a three-month simulation run. Nevertheless, partitioning needs more CPU resource. In the above research, which benefits are based on that the performance is affected. Thus, we believe how to save energy without decreasing the performance should be the focus of our work.

The remainder of this paper is as follows. Section II analyzes the interactions between Hadoop cluster with disks when running a Hadoop application, and we found that there is a disk idle time when shuffle to memory in Reduce phase. Section III conducts research on green disk storage based on that the principle-setting disk to standby when it is idle, and achieves four interfaces to managing disk states. Section IV conducts research on how to monitor real disk IO with Blktrace, presents our simple disk-energy-control algorithm. Section V makes a conclusion and highlights our future work.

II. INTERACTIONS BETWEEN HADOOP CLUSTER WITH DISKS

During the Map phase and Reduce phase, there could be several reads and writes of the data from and to the memory. This may result in a disk idle. To achieve a better understanding of the two procedures, some details of the implementation are provided as the following.

A. Map Phase

We divide the map phase into three phases: the reading, the buffering and the writing phase.

Reading Phase

During reading, the map task reads the input, which is called split. Hadoop, following the data locality optimization, does its effort to run the map task on a node where the input data resides in HDFS. There is no disk interaction since the input is directly fetched from the HDFS (Hadoop Distributed File System) and here is a short time that disks are idle. However, reading is so much faster, this short idle time can be negligible.

Buffering Phase

During buffering, three procedures take place: partitioning, sorting and spilling (to disk). Buffering is the phase during which the map output is serialized and written to a circular buffer. The available buffer size is defined by the configuration property io.sort.mb and by default is 100MB.When the map () function emits a record, it is serialized into the main buffer and meta-data are stored into accounting buffers. The accounting buffer can store meta-data for a predefined number of records before the spill thread is triggered. Once reached, a thread will begin to spill the contents to disk in the background. After spilling has been triggered, the data are divided into partitions corresponding to the reducers that they will ultimately be sent to. Within each partition, the background thread performs an in-memory sort by key and finally writes the data in a file to disk. Partitioning and sorting take place in the memory buffer, so there is an idle time for disks. After the map () function has processed all the input records, a flush () function will be called to write the remaining records of buffer to disk, there would be a number of spill files on disks. Partitioning and sorting is very quickly, so the disks are usually busy.

Writing Phase

The last phase is writing phase where the map output is written to disk as a single file. The flush () function calls mergeParts() function to merge all spill files. The mergeParts () function calls the user-define-combine () function according to needs. After the merge operation is completed, Map is basically completed, entering the wait submitted state. This procedure is the busiest one.

B. Reduce Phase

The reduce phase is more complicated than the map phase, with a high degree of parallelization and overlap in the operations of a single reduce task. The reduce task includes three phases: the shuffle/copy, sort and reduce phases.

Shuffle /Copy Phase

When a copier thread retrieves one of the map outputs, a decision is taken on where this output will be written/copied, in memory or on disk. Every reduce task is given a memory of a prespecified size, which is availMem. However, only a percentage of this memory is provided for the purposes of the copy/shuffle phase. The map output is copied in memory only if its size is less than the threshold. The max single shuffle segment fraction is the percentage of the in-memory limit that a single shuffle can consume and by default is equal to 0.25. A size of less than maxMapSize will allow map output to be written in memory; otherwise, it is propagated to disk. As soon as the copiers are assigned with a map output, they start writing it in a file in memory. When the accumulated size of the copied in memory outputs reaches the threshold. A background thread calls doInmemMerge () function to merge the stored in memory files yielding a new file on disk. As the shuffling phase continues, map outputs accumulate in memory, in-memory merge is triggered and new files are stored on disk. Therefore, shuffle stage has two situations: shuffle in memory and shuffle on disk. When shuffling in memory, there is no disk operation until the merge thread is triggered. Terasorting 10G data generated by Teragen on a cluster with 11 nodes shows us: when shuffling in memory, MERGE thread will wait 50 seconds to carry out a merge. Meanwhile, the disks are idle.

For the shuffle stage, some Hadoop applications have no reduce phase, so it has no shuffle stage, such as Teragen. But Hadoop application almost have no idle time in map phase, so to achieve green storage in map phase is impracticable. In addition, if reduce input block size is greater than maxMapSize, the block will shuffle to disk directly and the disk is always busy. So, we only consider these Hadoop applications that have both map phase and reduce phase which reducing both on memory and disk for our following green storage analysis, such as Terasort, Wordcount. TABLE I shows the shuffleinmemory () function infos when the above two applications deal with 10 GB data generated by Teragen on a cluster with four nodes. We find that: the maximum access time is 26.9s and 63.9s respectively, which means they have how long disk idle time.

In addition, some parameters, such as mapred.child.java.opts, maybe have some effects on the shuffleinmemory () function. Fig.1 shows the changes of shuffleinmemory () function infos when the parameter sets the value to 200M, 512M and 1024M. With the increase of the parameter value, the related time will decrease. It is because the available memory becomes larger when the parameter sets to a high value, but it will speed up the efficiency of data processing, the shuffleinmemory () function will be more efficiency.

Sort Phase

The sort phase includes the last merge of all the files, a phase before they are propagated to the actual reduce function. A merging algorithm can result in the creation of several temporary files before the last one is produced, the algorithm saves a trip to disk by implementing the final merge on the fly and directly feeding the reduce function with the results. This final merge can come from a mixture of in-memory and on-disk segments. Particularly, a decision is made whether the remained in-memory outputs will be kept in memory or will be

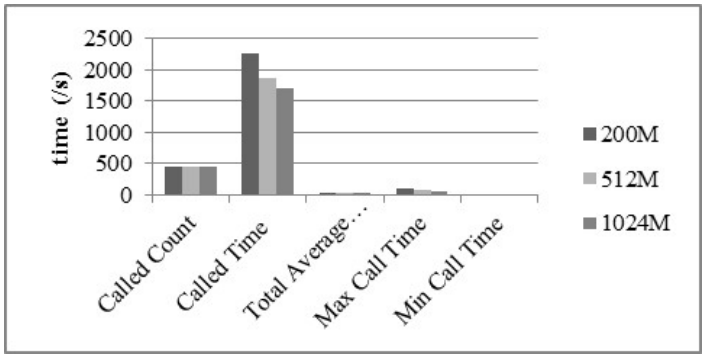


Fig. 1. The changes of shuffleinmemor() function info when the parameter has different setting.

TABLE I. THE SHUFFLEINMEMORY() FUNCTION INFO OF TERASORT AND WORDCOUNT.

Application	Called Count	Called Time	Total Average Call Time	Max Call Time	Min Call Time
Terasort	450s	1739.7s	3.9s	26.9s	0.0s
Wordcount	450s	1700.0s	3.8s	63.9s	0.0s

merged in one file and written to disk before the last final merge takes place. This decision aims at feeding the reduce the in-memory segments are spilled to disk, otherwise they are kept in memory and take part in the last merge as a separate file each. At the time, disks are usually very busy.

Reduce Phase

The reduce function processes the merge results based on the user's implementation and the final output is directly written to the HDFS. Consequently, no bytes are read or written locally, so disks are idle.

Based on the above analysis, we can find the disk is not always busy when running a Hadoop job, which can guide us to achieve Hadoop green storage. In order to realize it in user mode, we start the following research.

III. DISK ENERGY-SAVING TECHNOLOGY

A. Development of Disk-energy-saving Technologies

Power consumption has become increasingly important, so varieties of disk-energy-saving technologies are rapid emergence, such as using storage mediums that consume smaller energy, disk storage space-allocation optimization, physical-drive-level energy saving and Hierarchical storage management technologies [5].

Currently, magnetic tape after leaving the drive does not consume any energy, so the data are no longer used will be archived with tape to realize green storage. In addition, disk includes two main kinds: SSD(solid-state disk) and HDD (hard disk driver). SSD has not any internal movement of mechanical components, so its power consumption is much smaller than HDD. Although SDD has a good performance, the cost is large. According to a research from iSuppli, a professional statistics company. If data-centers worldwide gradually upgrade the HDD to SSD, data-centers will be able to save total 166,643 MWH electricity from 2008 year to 2013 yeah. Meanwhile, the daily power consumption of a 15,000-rpm SCSI (Small Computer System Interface) HDD is 14WH, while that of SSD is only 7WH, which can save 50% energy. Therefore, SSD will be more popular. Physical-drive-level energy saving is mainly suitable for high-capacity SATA (Serial Advanced Technology Attachment) HDD, which is realized by MAID (Massive Array of Idle Disks). Disk-storage-space-allocation optimization is mainly refers to deduplication (Data De-duplication) technology, which retain one copy of the same data only, the other is replaced by a pointer. Hierarchical storage management technologies (such as ILM Information Lifecycle Management) is mainly based on the frequency of data access and stratify the frequently accessed data to store on fast FC/SAS (Fibre Channel/serial-attached SCSI) disk drives, infrequently accessed data is stored in a relatively slower SATA disk drives and the archived data will be moved to tape. These energy-saving technologies are different, but now the MAID principle gets more and more attention. For HDD, there is a relatively mature technology to achieve energy saving, such as head Reset: reset the head when disk is idle to save energy, which is used in the latest Western Digital hard drive [6].

B. Disk Energy

A typical disk consists of the following main components: head, permanent magnet and spindle. It has special structure, logical structure and some related parameters. When responding to a disk IO request, the process can be divided into the following four phases. (1)Tracking phase, the head moves to the corresponding cylinders. (2)Rotationing phase, the disk slice is waiting to rotate to a matched position. (3) Transfering phase, the head reads data from the disk or writes data to disk. (4)Idle phase, the time is from current IO request is responded completely to the start of the next IO request. As Fig.2 shows: the disk states are the following three: active, idle and standby[7]. When a disk is active, the disk rotates at high-speed and the head is seeking, positioning or accessing data. Here, the disk power consumption is maximum. When the disk is idle, the disk keeps rotating, the head arm is stopped and most of the other electronic components are in a closed state, the disk power consumption is lower than the active case.When the disk is in standby mode, the electronic components are turned off fully and the disk stops spinning, the disk is the lowest energy consumption [8].

These disk states can be controlled by the related interface, which defines trigger conditions for each state and realizes the transferring between disk states. Currently, there are a lot of hard disk standard interfaces, such as IDE (ATA), SCSI, SATA and SAS. For energy management, the control theories of different interfaces are much same. Now the main supports for power management are APM (Advanced Power Manager) [9] and ACPI (Advanced Configuration and Power Interface) [10] , both of which cannot be detached from the support of the operating system and the related hardware. They enable us to manage disk energy consumption from the user mode, so there are some classic tools to control disk states, such as Hdparm [11] and smartmontools [12]. Hdparm is often with Linux, which can detect, display and set parameters for

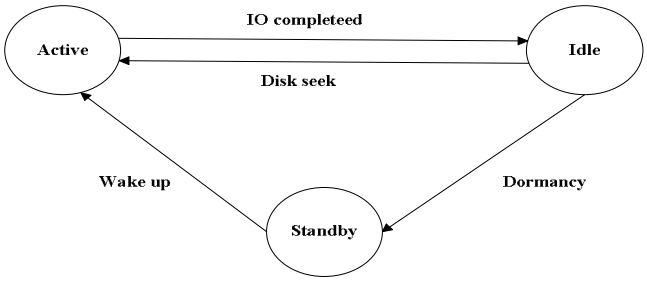


Fig. 2. Three disk states.

IDE and SCSI hard disks. We can use some commands to view, modify the disk states from the version 3.3 or later.

IV. A SIMPLE DISK ENERGY CONTROL ALGORITHM

A. Disk States Control Interface and Energy Measurement

Hdparm enable us to view and modify the disk states. With the inspiration of its source code, we realize how to get the disk status and how to set for the disk status.

We use hdparm-9.43 [13], which is the latest version. Here, we realize four interfaces,as mentioned above,we use these interfaces to set for variable disk status, use the dd command to test direct IO and buffer IO and watch the changes of disk status, calculating the time from the active to the standby and the time from the standby to the active. In the test, we draw three valuable experiences: First, there is the operating system on the primary disk, which is likely to wake up the disk if the test is standby or sleep. Thus, we need to reformat as ext3 file system for another disk, which details [14] are shown as Table three. In addition, the disk service-smartctl will defaultly conduct a self-test for the disk every 30min, which also can wake up the disk. To ensure these interfaces feasibility, this service should be closed in advance. Then, we found that buffer IO does not necessarily lead to the changes of the disk states, but direct IO can does it. Here, buffer IO means the IO between the cache and the buffer, and the IO between the buffer and disk. Direct IO refers to the IO between the cache and disk. If we do not make a distinction between these IO, then we may make wrong judgments for the buffer IO when disk is idle. Finally, these tests show us: the time of a classic SCSI disk needs to spin-up and spin-down are 1.45s, 0.49s.

In addition, we use an electric energy meter-QINGZHI 8775A [15], a node with four disks as shown in TABLE II and an energy-meter-monitoring platform installed on the windows XP to have a test. When there is absence of job, the power changes of the test disk are shown as TABLE III. And we find that the disk can save 5W power from active to standby in a relatively stable environment.

B. Algorithm Introduction

Based on the above analysis, we can design a simple disk-energy-control algorithm to realize Hadoop green storage. The principle of the algorithm is standby when disk is idle, which is, when there is not disk IO and the disk is active, we can set it to standby. In detail: if there is not disk IO in the 10s and the current disk state is active, disk state will be set to idle. We choose 10s because the time of a typical SCSI disk spin up (rotating preparation) takes 1.5s, the maximum is 10s (depending on the actual spin time as disks vary).

TABLE II. THE MODEL INFORMATION OF THE TEST DISK.

Model	**Capacity**	**Speed**	**Data Transmission rate**	**CHS**
ST31000524NS	1TB	7200 RPM	300MB/s	16383 /16 /63
spinup time	**spindown time**	**Average lantecy**	**Energy (age)**	
10s max	20s max	4.17ms	Idle4.61W Sleep0.59W	

TABLE III. THE RESULTS OF MEASUREMENT.

Sdd	**Sdc**	**Sdb**	**Sda**	**Total average power**
Standby	Standby	Standby	Standby	96.5W
Standby	Standby	Standby	Active/idle	101.5W
Standby	Standby	Active/idle	Active/idle	106.7W
Standby	Active/idle	Active/idle	Active/idle	112.3W
Active/idle	Active/idle	Active/idle	Active/idle	117.3W

standby. In detail: if there is not disk IO in the 10s and the current disk state is active, disk state will be set to idle. We choose 10s because the time of a typical SCSI disk spin up (rotating preparation) takes 1.5s, the maximum is 10s (depending on the actual spin time as disks vary).

C. Algorithm Introduction

Based on the above analysis, we can design a simple disk-energy-control algorithm to realize Hadoop green storage. The principle of the algorithm is standby when disk is idle, which is, when there is not disk IO and the disk is active, we can set it to standby. In detail: if there is not disk IO in the 10s and the current disk state is active, disk state will be set to idle. We choose 10s because the time of a typical SCSI disk spin up (rotating preparation) takes 1.5s, the maximum is 10s (depending on the actual spin time as disks vary).

Algorithm implementation

The implementation requires get the time of disk IO requests and calls the disk-states-control interfaces. At first, we need to know how to monitor disk IO, and distinguish it from other IO. With a deep study about various IO monitoring tools, Blktrace [16] is undoubtedly the best choice, which can get the disk IO request information in user mode. The results tell us: For Blktrace, buffer IO reading and writing has only once unplugged operation, but direct IO has the same unplugs operations as the times of writing or reading. Moreover, we can further to know the buffer IO has once unplugged operation because there is an IO between the buffer and disk. Therefore, we can use the unplug events to determine when there is a disk IO. The main code is on the following page.

Algorithm for Hadoop Case

First, Hadoop clusters are usually run on some ordinary nodes, these nodes do not use disk arrays or hybrid disks storage technology. A node may have two more disks, we can set the other unused disks to standby fully to achieve energy conversation. For the primary disk, we can use our algorithm to save energy.

For the test on a Hadoop cluster with 11 nodes, which run the Terasort benchmark with 10GB data generated by the Teragen.The simulation shows us: the Map stage is always reading each 4KB data to the memory. The completion time of a Map block is 18.3s and block size is 64M, so the process time of 4KB data is millisecond level, while the time of a typical SATA disk spin-up is 1.45s, so to achieve energy-efficient for disk in Map stage is infeasible. In Reduce phase, when shuffling to memory, there are 32 times waiting of merge process threads and every waiting time is 55 seconds. When we use our algorithm, the energy consumption of the cluster can be saved 2.47WH. For the biggest Hadoop cluster with 42,000 nodes, the energy savings is significant.

Algorithms:Simple Disk energy control algorithm

Input:current_timestamp,last_io_time[i],disk_state

Output: Disk state set command

algorithms demonstration**:**

```
1  Last_io_time[ ];
   //the last disk_io request timestamp array of each disk
2  setitimer(ITIMER_REAL,&timer,NULL);
   //Timer:refresh last_io_time array every one second
3  signal(SIGALRM,alarm_handler);
4
   current_timestamp=current_time_stamp(start_timestam
   p,current_timestamp);
   //current_timestamp
   //energy consumption control
5  for(int i=0;i<n;i++)
6   {
7   If(current_timestamp-last_io_time[i]>10
    &&disk_state=active)
8   set_status("y",dev_name[i]);
   //disk state into standby immediate
9   }
```

V. CONCLUSION AND FUTURE WORK

As data-centers are intensive, the energy consumption is increasingly important. The disk as the most important storage device, how to achieve green storage has great significance in the massive Hadoop cluster with thousands of disks. In this paper, we analyze the interactions between disks and cluster when running a Hadoop job, and we found there is a long idle time when shuffle to memory in Reduce phase, which can guide us to do research about disk energy-saving technologies and achieved four interfaces for disk states. Then,we used Blktrace to monitor the real disk IO and implemented a simple disk-energy-control algorithm, which has achieved to save energy without affecting Hadoop performance. However, the algorithm is very simply one, so there is still much room for optimization. In our test, we did not consider the latency caused by our algorithm. In addition, we use Blktrace for background monitoring, which also requires certain resources to collect data, so it is necessary to quantify costs and benefits of the monitoring program. Then, we hope use some ways such as using data format to predict next operations in the future, which is very useful that set the disk to standby when shuffle to memory with the interfaces directly and this will eliminate the overhead of monitoring. Alternatively, we can achieve disk spin-up before the arrival of disk IO requests, which is the real significance of disk-energy-control algorithm.

REFERENCES

[1] Zikopoulos P, Parasuraman K, Deutsch T, et al. Harness the Power of Big Data The IBM Big Data Platform[M]. McGraw Hill Professional, 2012.

[2] Hadoop .[EB/OL],http://Hadoop.apache.org/.

[3] Leverich J, Kozyrakis C. On the energy (in) efficiency of Hadoop clusters [J]. ACM SIGOPS Operating Systems Review, 2010, 44(1): 61-65.

[4] Kaushik R T, Bhandarkar M. GreenHDFS: Towards an Energy-Conserving Storage-Efficient, Hybrid Hadoop Compute Cluster[C]//Proceedings of the USENIX Annual Technical Conference. 2010

[5] Mainstream disk storage energy-saving technologies introduction. [EB/OL],

[6] http://emc.ofweek.com/2013-01/ART-8320058-11000-28660821.html.

[7] Tian Lei,Feng Dan, Yueyin Liang, Wu Suzhen,Mao Bo.Survey on Power-saving Technologies for Disk based Storage Systems [J].computer science,2010,37(9).

[8] Seagate. ATA Interface Reference Manual. 36111-001, Rev. C, 1993-05-21.

[9] Information technology -AT Attachment 8 - ATA/ATAPI Command Set (ATA8-ACS). Working Draft Project American National Standard.T13/1699-D. Revision 4a, May 21, 2007.

[10] Advanced power management. [EB/OL], http://en.wikipedia.org/wiki/Advanced_power_mana-gement,2012-11-14.

[11] Hdparm (8) - Linux man page. [EB/OL], http://linux.die.net/man/8/hdparm.

[12] smartctl- monitoring hard disk status uses smartmontools to monitor disk health status. [EB/OL], http://blog.csdn.net/smartmz/article/details/6031742.

[13] Hdparm-9.43.[EB/OL], http://sourceforge.net/projects/hdp-arm/develop?source=navbar.

[14] Seagate.Product Manual.Constellation® ES Serial ATA.

[15] 8775A-Electric energy meter. [EB/OL], http://www.mgd17.com/product_detail.asp?pid=01060721.

[16] Blktrace use Brief. [EB/OL], http://blog.sina.com.cn/s/blog_48c95a190100dp5w.html.

Multi-core based Parallelized Cooperative PSO with Immunity for Large Scale Optimization Problem

Zhao-Hua Liu
School of Information and Electrical Engineering, Hunan University of Science and Technology, XiangTan, China.
Email:zhaohualiu2009@hotmail.com

Xiao-hua Li
School of Information and Electrical Engineering, Hunan University of Science and Technology, XiangTan, China

Jing-XingZhao
School of Information and Electrical Engineering, Hunan University of Science and Technology, XiangTan, China

Wen Tan
School of Information and Electrical Engineering, Hunan University of Science and Technology, XiangTan, China

***Abstract*—A parallelized cooperative multiple particles swarm optimization algorithm with immunity mechanism based on the multi-core architecture is proposed for large scale optimization problem in this paper, named M-PCPSO-I. A novel memory information sharing scheme is designed for particles and facilitates communication among different swarms in the population space. The global best individuals selected from sub-swarms are saved in the leader set and promoted by using the improved immune clonal selection operator. The M-PCPSO-I algorithm is paralleling implementation on a share-memory computer system through the multi-core architecture. The high dimension problem results validated the proposed algorithm have good computational performance, and also the computational efficiency is greatly enhanced by multi-core parallelization.**

Keywords: particle swarm optimization (PSO); artificial immune system (AIS); information sharing;parallel;high dimension problem.

I. INTRODUCTION

In reality , the problems have become more and more complex ,especially in practical engineering applications and society economic management. It is a big challenging for finding the global optima of those large scale problems as it has the feature of higher dimension and bigger storage. Particle swarm optimization (PSO) was firstly put forwarded by Kennedy and Eberhart in 1995[1]-[2], which imitates the behavior of swarms in nature. The PSO has been widely used for solving problems in science and engineering [3]-[5], as it has a potential for complex problem solution with the development of computer. However, since the PSO is also based on swarm iterative computation which may cause to lose diversity and suffer from trapping in local optima at the later stage of evolution.

This work was supported in part by Key Projects in the National Science and Technology Pillar Program (2012BAH09B02), National Natural Science Foundation of China (61174140,51374107), Doctoral Fund of Ministry of Education of China (20110161110035), China Postdoctoral Science Foundation Funded Project (2013M540628, 2014T70767), and National Natural Science Foundation of Hunan Province (14JJ3107).

Recently, authors developed a lot of improved version algorithm to improve solution performances of the PSO. For instance, a hybrid PSO with mutation is proposed by Ahmed et al in [6]. Juang et al. proposed a hybrid PSO algorithm associating with the genetic operator in [7], the diversity of PSO is significantly enhanced by using the GA operators .Compared with basic PSO, the hybrid PSO with wavelet dynamic muta which can obtain dynamic optimization effect by using the wavelet function dynamic advantages. Lian flying direction of particle and proposed a comprehensive learning PSO in [9] .Although the method in [9] is superior in keeping diversity; it does not design a scheme to jump out of the local optima when the whole population is losing diversity. A parameter adaptive regulation scheme and an elitist learning strategy are introduced into the PSO, Zhan et al. in [10] proposed an adaptive PSO which can accelerate the convergence speed and jump out of the local optima. Some researchers proposed to use multi-population scheme to improve the diversity of PSO [11]-[14], it is because the spatial distribution of multiple populations are broader compared to single population. For instance, Bergh et al. in [11] proposed a multiple PSO algorithm, the whole population was divided into many small swarms, each part swarm are to optimize different parts of the problem, which showed a better performance compared to single PSO. Every version of PSO has different merits when face to different complex problem. The existing intelligent algorithms are mostly based on serial computing using center processor. The computational costs are still a great constraint for intelligent optimization algorithm, especially in optimization problems with mass data processing or high dimensions, which may influence on the algorithm convergence performance.

In this paper, a parallelized immune cooperative PSO computational framework using OpenMP based on multi-core architecture is proposed. The proposed M-PCPSO-I is to combine the co-evolution theory [15], artificial immune system (AIS) mechanism [16]-[17], with multi-core parallel computing technique [18].The framework of M-PCPSO-I consists of one leader population and several normal swarms based on parallel collaborative computing model. In each

generation of the algorithm, the global best individual of normal swarms will be selected into leader set. The memory is updated by the improved immune clonal selection operator. The information sharing mechanism can assist the inter-swarm communication. And also, the proposed method is parallelized on the multi-core by using the OpenMP. Compared to other hybrid PSOs, the performance of the proposed M-PCPSO-I is tested and verified by some standard benchmark functions, which show better performance in global search, solution accuracy, and convergence speed. In addition, the results shows that the proposed approach offers high speed-up and a considerable time cost.

II. PARALLEL IMMUNE COOPERATIVE MULTIPLE PSO

A. Principle of PSO Algorithm

Assuming in a *d*-dimensional solution space, each particle *i* is composed of two vectors, the vector $V_i = \{V_{i1}, V_{i2}, \ldots, V_{id}\}$ and the position vector $X_i = \{X_{i1}, X_{i2}, \ldots, X_{id}\}$. The searching procedure can be given by:

$$V_{id}(t+1) = \omega V_{id} + c_1 * rand_1()(Pbest_{id}(t) - X_{id}(t)) + c_2 * rand_2()(gBest_d(t) - X_{id}(t)) \quad (1)$$

$$X_{id}(t+1) = X_{id}(t) + V_{id}(t+1) \quad (2)$$

In (1)-(2), $Pbest_{id}$ represents the best position found by *i*-th particle up to now and $gBest_d$ is the best particle among the entire population. c_1 and c_2 are the acceleration coefficients, ω is the inertia weight factor decreasing linearly, $rand_1$ and $rand_2$ are two uniformly distributed numbers generated randomly in the range of [0,1], respectively.

B. Principle of M-PCPSO-I Algorithm

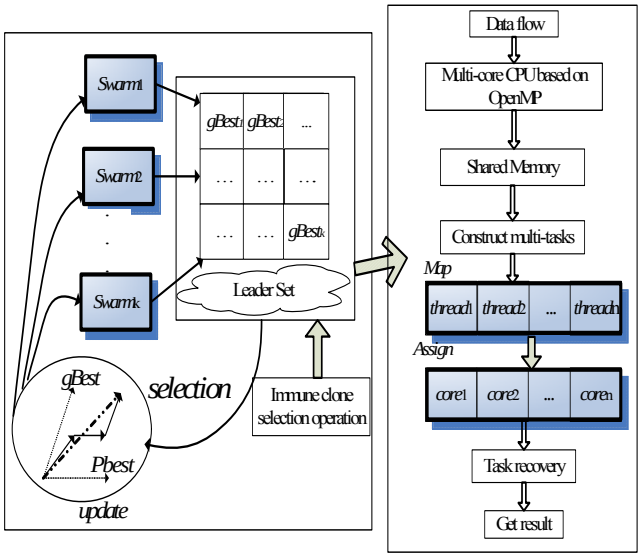


Fig. 1. The model of M-PCPSO-I

The M-PCPSO-I consists of two level architectures, the upper leader population and the bottom normal subpopulations. The multi-population cooperative scheme can enhance the population diversity since involves a population exploiting the current optimal solution, while others populations are encouraged to explore potential excellent solution in the search space. Further, an immune evolutionary mechanism which can help the algorithm escape from local minima; moreover, the parallel computing technique can help speed up the proposed method. The OpenMP (open multiprocessing) programming interface is widely used in multi-core CPUs in the field of engineering application and scientific computing. Additionally, the OpenMP provides a structured interface for multithreaded standard which does not require a large amount of code restructuring and also not required to handle communication issues for parallelization. So, it is very simple and flexible for developing parallel applications on multi-core computing platforms on the OpenMP platform. The model and flow of M-PCPSO-I are as shown in Fig.1. In Fig.1, $gBest_i$ represents the muti-*gBest* individual selected from the different normal swarms, where $Swarm_k$(K>0) represents the K-th normal swarm of the whole population. Furthermore, the memory is treated as the Leader set in immune system.

Algorithm: M-PCPSO-I

Step1:Initialize parameters , normal subpopulations $swarm_i$ and leader set
Step2:parallel execution based on OpenMP
Executing main thread
fork();// parallel regions
Step3:for i=1 to I //1≤i≤I , I is the number of normal sub-swarm;
Perform the process of PSO for sub-population *Pi* ;
update $particle_i$ velocity using the equation (1)
update $particle_i$ position using the equation(2)
Evaluate the fitness value of $particle_i$;
end for
Step4: The best individuals of each normal sub-swarm are selected into memory and constitute *leader set.*
Step5: Perform the process of immune clonal selection with adaptive wavelet hypermutation for leader set based on (4)-(6).
Step6: Set barrier ().
Step7:Join();//sequential region
Step8:Executing main thread
Step9: Until a terminate-condition is met, or else, returns to step3.
Step10: output the result.

The main procedure of M-PCPSO-I algorithm as follows:

1) Memory information -sharing cooperative PSO

To promote the excellent search information comminuting between the particles among themselves, an excellent information sharing scheme based on immune memory scheme is designed for the PSO. In the personal level, the particle will follow its best experienced behavior in its history. In the global level, the entire population will follow the best particle randomly select from leader set. This modified paradigm of PSO is formulated as follow.

$$V_{id}(t+1) = \omega V_{id} + c_1 * rand_1()(Pbest_{id}(t) - X_{id}(t)) + c_2 * rand_2()(gBest_{\varphi i[d]}(t) - X_{id}(t)) \quad (1)$$

$$X_{id}(t+1) = X_{id}(t) + V_{id}(t+1) \quad (2)$$

where $\varphi i = \lfloor rand*k \rfloor$, the leader set is *Leader*= {$gBest_1$, $gBest2$,…,$gBest_k$} is composed of the different global particles selected from the different swarms. The leader set is enhanced by immune clonal selection. The whole process is detailed as follows.

1) *Clone*. In each generation, the *gBest* individuals are clonal and their scale is proportional to its fitness. The clonal scale of the whole *leader* set is shown in (3).

$$N_c = \sum_{i}^{N} round(\frac{\beta N}{i} + C) \tag{3}$$

where N is the memory size, β is within (0, 1).C is the fairness factor, it usually larger than one.

2) *Adaptive dynamic wavelet hymutation.* After the clonal expansion and generate a temporary clonal population, employing the wavelet mutation operator to improve the dynamic performance of antibodies as the wavelet mutating space is dynamically varying during the search process. The Morlet wavelet features (as shown in Fig. 2) [8] is an example of the mother wavelet.

$$\sigma = \frac{1}{\sqrt{a}} e^{-(\frac{\varphi}{a})^2/2} \cos(5(\frac{\varphi}{a})) \tag{4}$$

In term of (4), where a is the dilation parameter for mother wavelet and $\varphi \in [-2.5a, 2.5a]$ and will be randomly generated during the iterative process.

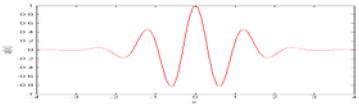


Fig. 2. Morlet wavelet.

The amplitude of the function can be adjusted through controlling the dilation parameter a. So, an adaptive nonlinearly decreasing strategy for the parameter a is proposed, which given by equation (5).

$$a = a_{\max} - (a_{max} - a_{min}).(\frac{T-t}{T})^2 \tag{5}$$

where a_{max} , a_{min} are the upper and lower boundaries of a, respectively. By using (4) and (5), all the antibody temporary clonal population will be operated by (6) in each generation

$$gBest_d^{new} = gBest_d + \sigma * gBest_d \tag{6}$$

3) *Immune selection*. After the hyper-mutation and generated new offspring, the new generated individuals will be reselected and updated the *leader* set according to the fitness, the details as in [16].

2) Algorithm parallelization implementation based on OpenMP

The OpenMP architecture provides fork-join programming model for multithreaded standard which does not require a large amount of code restructuring for parallelization as well as not required to handle communication issues. Based on the above principle, the M-PCPSO-I is run in multi-core processors based on the OpenMP as shown in Fig.3. Meanwhile, the workload can be equally or fairly distributed among cores should be considered.

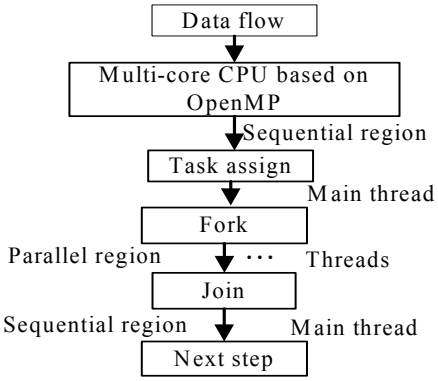


Fig. 3. the M-PCPSO-Irunning on multi-core CPU based on OpenMP.

C. Testing for the M-PCPSO-I with high dimension problem

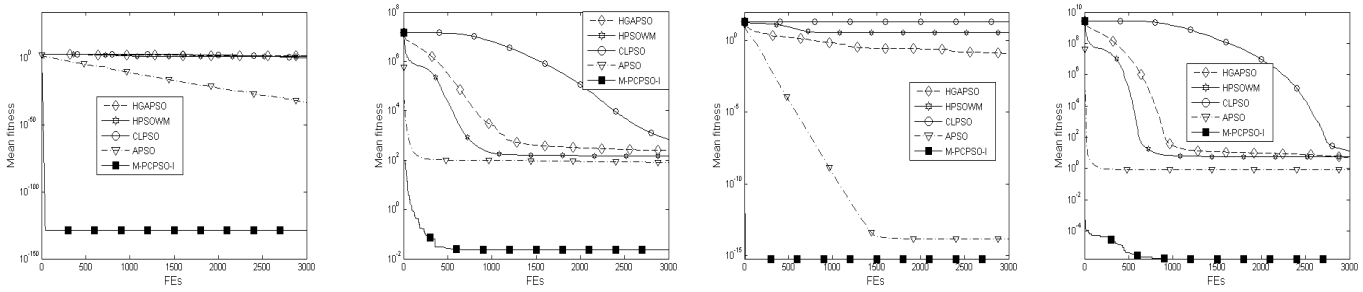
The performance of M-PCPSO-I is validated by a set of standard benchmark test functions which listed in Table 1.All these functions are tested 30 times and with one hundred dimensions. The comparisons are in term of the mean results, standard deviation and the *t-test* value. A series of hybrid PSOs are used for comparison with the M-PCPSO-I.The existing hybrid PSOs as follows: HGAPSO (hybrid PSO with genetic algorithm) [7], HPSOWM (hybrid PSO with Wavelet Mutation) [8], CLPSO (comprehensive learning PSO) [9], APSO (adaptive Particle Swarm Optimization) [10], which have better performance in unmodal and multimodal problem solution. The value of "acceptance" in Table Ⅰ is defined to judge whether a solution predefined found by the PSOs would be acceptable or not as follow in [10]. The parameters for M-PCPSO-I as follows: The inertia weight w in term (4) is set to be w in the range of [0.90, 0.4] and linearly decreases as in [10], the acceleration coefficients c_1,c_2 are both set to be 1.49445 as given in [10]. In equation (6), β is set to be 0.8 and b is setup to be 5. All algorithms are tested use the same number of 3000 FEs(function evaluations) for each test function, All the hybrid PSOs with the same population size of 30 in each subpopulation. All experiments are tested on the same computer with AMD Athlon (tm) Ⅱ X2 250 multi-cores Computer with four processors. The software computing platform is visual studio 2010.

TABLE I. TEST FUNCTION

Test Function	Domain Range	Acceptance	Global Optimal

$f_1(x) = \sum_{i=1}^{100} \lvert x_i \rvert + \prod_{i=1}^{100} \lvert x_i \rvert$	$-10 \le x_i \le 10$	0.00001	0
$f_2(x) = \sum_{i=1}^{100} [100(x_{i+1} - x_i^2)^2 + (x_i - 1)^2)]$	$-10 \le x_i \le 10$	50	0
$f_3(x) = -20\exp(-0.2\sqrt{\frac{1}{30}\sum_{i=1}^{100} x_i}) - \exp(\frac{1}{30}\sum_{i=1}^{100}\cos(2\pi x_i)) + 20 + e$	$-32 \le x_i \le 32$	0.001	0
$f_4(x) = \frac{\pi}{30}\{10\sin(\pi y_1) + \sum_{i=1}^{100-1}(y_i - 1)^2[1 + 10\sin^2(\pi y_{i+1})] + (y_{30} - 1)^2\} + \sum_{i=1}^{100} u(x_i, 10, 100, 4)$ $where, y_i = 1 + \frac{1}{4}(x_i + 1), u(x_i, a, k, m) = \begin{cases} k(xi-a)^m ..., xi > a \\ 0 ..... , -a \le xi \le a \\ k(-xi-a)^m, xi < -a \end{cases}$	$-50 \le xi \le 50$	0.0001	0

1) Accuracy comparison of different PSOs



(a) (b) (c) (d)

Fig.4. The comparisons of the five different PSOs on the four test problems with 100 dimensions：a(*f*1),b(*f*2),c(*f*3),d(*f*4)

TABLE II. THE COMPARISON ON FIVE PSOS

Function		HGAPSO	HPSOWM	CLPSO	APSO	M-PCPSO-I
f_1	mean	35.084	9.869	0.625	4.389×10^{-34}	2.257×10^{-129}
	Std.dev	27.323	6.991	0.108	7.68×10^{-34}	8.289×10^{-129}
	t-value	9.079	9.982	40.921	4.041	0
f_2	mean	247.687	150.973	695.582	81.953	0.023
	Std.dev	73.306	55.156	113.34	5.579	0.032
	t-value	23.889	19.352	43.395	103.788	0
f_3	mean	0.124	3.624	20.087	1.421×10^{-14}	5.887×10^{-16}
	Std.dev	0.606	0.521	0.029	2.60×10^{-15}	0
	t-value	0.001	0.049	4.898	0.036	Na/N
f_4	mean	4.912	5.348	13.141	0.816	1.662×10^{-6}
	Std.dev	2.396	1.668	2.167	0.0056	2.964×10^{-6}
	t-value	14.496	22.672	42.880	984.862	0

The comparison results are shown in Table Ⅱ, which are in terms of mean fitness and standard deviation (Std. Dev) of the solutions obtained from five different algorithms. The graphically presents convergence features in solving four test functions between these hybrid PSOs as shown in the Fig.4. As can be seen from the Table Ⅱ and Fig.4., The M-PCPSO-I can significantly improve the optimizing performance for solving most of these test functions and provide a good performance in solving both unimodal functions and multimodal problem in terms of the solution accuracy and convergence speed.

The *t-test* is used to judge the differences between the M-PCPSO-I and other hybrid PSOs .The *t-test* (*t-value*) [8] is a statistical method to evaluate how significant difference between two methods. The results of *t*-values between the M-PCPSO-I and other hybrid PSOs are shown in Table Ⅱ. From the Table Ⅱ, most *t*-values are higher than 1.645, so, the performance of the M-PCPSO-I is significantly better than that of other PSOs with a 95% confidence level. As can be seen, the M-PCPSO-I shows better performances in convergence, global search and dynamic performance. The reasons are that the proposed M-PCPSO-I is based on multiple populations which can activate the diversity of the whole population. The dynamic optimization performance is greatly enhanced by the improved immune clonal selection with adaptive wavelet mutation operator.

2) Computational complexity Comparison of Different PSOs

TABLE III. THE COMPARISON ON COMPUTATIONAL COMPLEXITY OF FIVE ALGORITHMS

	Function	HGAPSO	HPSOWM	CLPSO	APSO	M-PCPSO-I
f_1	Mean Fes	4673.68	1377.43	3990.97	562.8	401.17
	Time(s)	1.01	0.319	0.628	0.123	0.061
	Ratio(%)	63.3	100	100	100	100
f_2	Mean Fes	2006.83	1388.57	3593.68	1063.33	175.13
	Time(s)	0.453	0.319	0.536	0.233	0.026
	Ratio(%)	96.7	100	93.3	100	100
f_3	Mean Fes	2490.4	1305.03	_	1341.4	316.13
	Time(s)	0.69	0.461	_	0.338	0.208
	Ratio(%)	100	100	_	100	100
f_4	Mean Fes	2106.2	1606.35	3836.5	1158.13	673.66
	Time(s)	1.217	1.211	1.703	0.439	0.301
	Ratio(%)	100	56.7	100	100	100

The convergence speed of an optimization algorithm is also an important feature to prove its superior to other algorithms. Table Ⅲ shows that M-PCPSO-I generally has comparatively fewer iteration times such as the mean number of FEs or the mean cost of CPU time for searching an acceptable solution (list in Table Ⅰ). The actual cost of CPU time is an important feature to describe the computational cost of an algorithm, for many existing hybrid PSOs have added extra CPU computational time as given in TABLE Ⅲ. For example, for reaching an acceptable solution, tests on f_1 show that the average numbers of FEs with 4673.68, 1377.43, 3990.97, 3972.23, and 562.8 are required for every particle in the HGAPSO, HPSOWM, CLPSO, and APSO algorithms, respectively.

However, each particle within the M-PCPSO-I only uses 401.17FEs on average whereas its CPU compute time is 0.061s, which is the small compute time among the five hybrid PSOs algorithms. To sum up, the M-PCPSO-I spends the smallest number of FEs and the least CPU time to reach acceptable solutions on four typical different test functions, the convergence is faster. The reasons that the proposed M-PCPSO-I is parallelization and running on multi-core CPU based on OpenMP, experiments illustrated that the computational efficiency is greatly improved with the increasing core number.

III. CONCLUSION

A novel parallel immune cooperative multiple particles swarm optimization algorithm based on multi-core is proposed. The framework of M-PCPSO-I consists of one upper memory population and bottom normal swarms. The memory is promoted through the improved immune clonal selection with adaptive wavelet mutation operator. Moreover, the proposed method running on multi-core CPU using OpenMP and the computational efficiency is greatly enhanced.

The performance of the proposed M-PCPSO-I is validated by some standard benchmark functions and show better performance in convergence speed and global search. From the tests, we can see that the proposed M-PCPSO-I outperforms the other hybrid PSOs. The proposed method provides better performance and fast convergence. In the future, we will put the proposed method executing on map/reduce architecture and develop a map/reduce–based scalable parallel M-PCPSO-I method.

REFERENCES

[1] R. C. Eberhart and J. Kennedy, "A new optimizer using particle swarmtheory," in Proc. 6th Int. Symp. Micromachine Human Sci., Nagoya, Japan, 1995, pp. 39–43.

[2] J. Kennedy and R. C. Eberhart, "Particle swarm optimization," inProc. IEEE Int. Conf. Neural Netw., Perth, Australia, 1995, vol. 4, pp. 1942–1948.

[3] Z H Liu, J Zhang, S W Zhou, XH Li, K Liu. "Coevolutionary Particle Swarm Optimization Using AIS and its Application in Multiparameter Estimation of PMSM" IEEE Transactions on cybernetics,vol 43,no.6,pp.1921-1935,Dec,2013.

[4] F.J. Lin, L.T. Teng, J.W. Lin,S.Y.Chen, "Recurrent Functional-Link-Based Fuzzy-Neural-Network-Controlled Induction-Generator System Using Improved Particle Swarm Optimization, " IEEE Trans. Ind. Electron.,vol.56, no.5,pp. 1557 - 1577 , May, 2009 .

[5] C.H. Liu and Y.Y. Hsu," Design of a Self-Tuning PI Controller for a STATCOM Using Particle Swarm Optimization ," IEEE Trans. Ind. Electron.,, vol.57,no.2, pp. 702 – 715, Feb.2010

[6] A. A. E. Ahmed, L. T. Germano, and Z. C. Antonio, "A hybrid particle swarm optimization applied to loss power minimization," IEEE Trans.Power Syst, vol. 20, no. 2,pp. 859–866, May ,2005.

[7] C. F. Juang, "A hybrid of genetic algorithm and particle swarm optimization for recurrent network design," IEEE Trans. Syst., Man, Cybern. B,Cybern, vol. 34, no. 2, pp. 997–1006, Apr. 2004.

[8] S.H .Ling, H.H.C .Iu; K.Y. Chan,H.K .Lam, B.C.W .Yeung and F.H. Leung," Hybrid Particle Swarm Optimization With Wavelet Mutation and Its Industrial Applications ", IEEE Trans. Syst., Man, Cybern. B,Cybern., vol.38,no.3, pp. 743 – 763, Jun. 2008.

[9] J. J. Liang, A. K. Qin, P. N. Suganthan, and S. Baskar, "Comprehensive learning particle swarm optimizer for global optimization of multimodal functions," IEEE Trans. Evol. Comput., vol. 10, no. 3, pp. 281–295,Jun. 2006.

[10] Z.H.Zhan, J. Zhang, Y. Li, and H. S.H. Chung." Adaptive Particle Swarm Optimization," IEEE Trans. Syst., Man, Cybern. B,Cybern., vol. 39, no. 6, pp.1362-1381.Dec.2009.

[11] F. Van den Bergh and A. P. Engelbrecht, "A cooperative approach to particle swarm optimization," IEEE Trans. Evol. Comput., vol. 8, no. 3,pp. 225–239, Jun. 2004.

[12] R. A. Krohling and L. S. Coelho, "Coevolutionary particle swarm optimization using Gaussian distribution for solving constrained optimization problems," IEEE Trans. Syst., Man, Cybern. B, Cybern., vol. 36, no. 6, pp. 1407–1416, Dec. 2006.

[13] G.G. Yen, and W.F. Leong."Dynamic Multiple Swarms in Multiobjective ParticleSwarm Optimization," IEEE Trans. Syst. Man Cybern. Part A-Syst. Hum., vol. 39, no. 4, pp.890-911,Jul. 2009.

[14] Z H Zhan , JJ Li , JN Cao , J Zhang , H.S.-H. Chung , and YH Shi. "Multiple Populations for Multiple Objectives: A Coevolutionary Technique for Solving Multiobjective Optimization Problems," IEEE Transactions on Cybernetics, Vo. 43 , no. 2 ,pp. 445 – 463, Apr .2013

[15] M. A. Potter, K. A. De Jong. Cooperative coevolution: An architecture for evolving coadapted subcomponents. Evol. Comput., vol.8,no.pp.1-29,Janu.2000.

[16] H W Ge, L Sun, Y C Liang,etal.An effective PSO and AIS-based hybrid intelligent algorithm for Job-Shop Scheduling. IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans, vol.38,.no.2,pp.358-368.Mar.2008.

[17] A.M. Whitbrook, U. Aickelin, J.M. Garibaldi, "Idiotypic Immune Networks in Mobile-Robot Control," IEEE Trans. Syst., Man, Cybern. B, Cybern.,vol.37,no.6,pp. 1581 – 1598, Dec.2007

[18] Y Shi, and CH Chan , "An OpenMP Parallelized Multilevel Green's Function Interpolation Method Accelerated by Fast Fourier Transform Technique," IEEE Transactions on antennas and propagation, vol. 60, no. 7, pp.3305-3313.July. 2012.

Small File Access Optimization Based on GlusterFS

Xie Tao, Liang Alei
School of Software Engineering
Shanghai Jiao Tong University
Shanghai, China
Foxterran@163.com, liangalei@sjtu.edu.cn

***Abstract*—This paper describes a strategy to optimize small file's reading and writing performance on traditional distributed file system. Traditional distributed file system like GlusterFS stores data within local file system (XFS, EXT3, EXT4, etc.), which shows a significant bottleneck on file metadata lookup. We try to re-design metadata structure by merging small file into large file, thus to reduce size of metadata, so we can store the whole files' metadata inside main memory. We design and implement the whole strategy on GlusterFS, test results show a great performance optimization on small file operation.**

Keywords—small file; distributed file system; metadata; optimization; glusterfs;

I. INTRODUCTION

Every day, billions of data files are transferred along the Internet, distributed storage techniques or so called cloud storage systems make processing and archiving petabytes magnitude data possible. In our daily life, we use web service tools such as Drop box, Box or iCloud to sync or store files. By analyzing size or formats of these files, we notice that a large part of those files' size is less than 10MB, such as photos, travel notes or worksheets. According to the statistic data comes from Facebook, it currently stores over 260 billion photos, which equals about 20 petabytes amount of data. There are one billion new photos uploaded to its server every week and Facebook serves over one million images per second at peak [2]. Performance optimization on small files like photos or notes has a huge impact on front end user experience [1]. As these numbers definitely will keep increasing in the future, small file access optimization poses a significant challenge for distributed storage system infrastructure.

Modern distributed file system such as GFS(Designed by Google), HDFS, Swift(OpenStack), GlusterFS, did a quite good job on data location, failure recovery and detection, linear scale out and consistent view maintaining [6-7], but at the bottom of their whole architecture, data storage still rely on local file systems, which usually are XFS, EXT3 or EXT4. In our test and analysis, those traditional local file systems have some disadvantages: directories and per file metadata. For small file access, these features are not necessary. For example, in the scenario of photo READ or WRITE, owner or attributes of metadata are not used and thereby waste storage capacity [1].

But the biggest disadvantage of traditional local file system is found inside the process of file looking up, the file's metadata called inode must be read from disk during the whole lookup process, meanwhile the real application scenario we deal with is write once, read often, never modified and rarely deleted, this small delay caused by metadata disk operation multiplied by billions of file access per time, becomes main throughput bottleneck [1]. File looking up in local file system requires at least three disk operations: one for translating file name into inode number, one for reading inode information from disk to memory, and the last one for reading file content.

In this paper, we try to reduce times of disk operations during file lookup to complete performance optimization. The whole strategy is to redesign the structure of metadata by merging small files into large file, reorganize architecture of file system's disk operation related part. We design and implement this optimization method on GlusterFS, one popular, open source, and well-designed distributed file system.

We organize the remainder of this paper as follows. Section II provides background of base file system and its advantages, Section III describes overview of the design and details of real implementation, Section IV demonstrates performance optimization, Section V discusses about future work and concludes the whole paper.

II. BACKGROUND

A. Overview

Gluster is a set of open source solutions, it can run on commodity hardware. Gluster uses consistent hashing algorithm and decentralized cluster to enable linear scale-out as much as possible, both performance and capacity benchmark are quite excellent. Its base file system AKA GlusterFS uses single namespace to access multiple standards. It constructs a pool of storage including disk and memory [14].

GlusterFS is the core of Gluster Solutions, distributed in the form of RPM and Debian packages. It is supported on most major linux distribution, running in user space gives it privilege to be exported as NFS, CIFS, FTP, and HTTP(S) which are all user level applications. GlusterFS can be locally attached, and can be managed by one single command. With elastic hash algorithm and stack organized architecture, GlusterFS is highly scalable (more than 1000 servers), and reliable (failure recovery, RAID inside and error detective encoding). Integrated with other Gluster solutions,

GlusterFS's method of storing and managing large amount of unstructured data is simple. Figure 1 illustrates basic architecture of GlusterFS. Client-Server model is also another great feature.

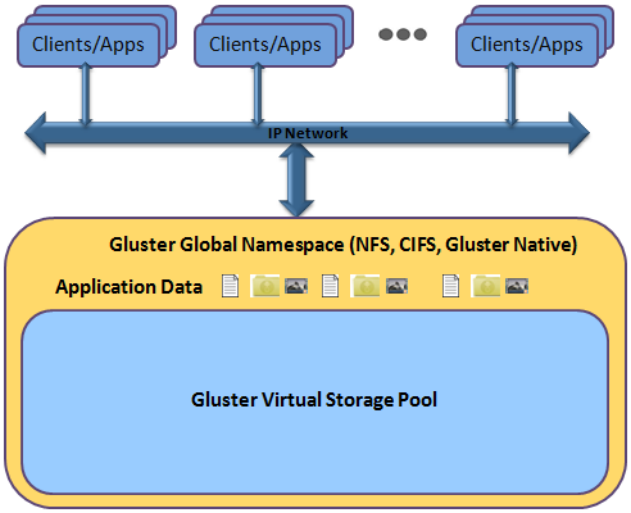


Fig. 1. GlusterFS Architecture

B. *Advantages of GlusterFS*

A lot of enterprise level cloud storage services are built on GlusterFS. Redhat, Box, Dropbox and other famous cloud storage service providers all choose GlusterFS as their key archiving bottom level components, performance optimization could bring great benefits. As an open source project, GlusterFS has complete documentations and full source code access. Optimizing performance on such platform makes work sufficient and possible.

GlusterFS compiles its source code with so called translator mechanism [13], which looks like function stack architecture. This is a loose coupling structure. Modifications of each layer of components will not affect each other, thereby makes importing a brand- new implementation much easier.

III. Design And Implementation

The base infrastructure has a lot of advantages: linear scale out of scalability, failure tolerance and recovery, decentralized metadata management and flexible interface [3-5]. Our design should not compromise all these features. Optimization should not harm original distributed system's architecture, which is quite loose coupling [6]. Our goals are accelerating random read speed and maintaining reliability and scalability.

A. *Design Overview*

GlusterFS's stack organized components structure makes modification convenient and easy, principle of indirection requires layer independency. Write and random read only interact with hard disk, we can simply add an extra layer between file parsing and disk operations. As our bottleneck's analysis shows, redesigning metadata structure is on top of our task schedule.

Prime strategy is merging small files into large file, as to small file's metadata, delete original inode's unnecessary elements, keep track of small file's offset in large file and its actual size. The layout of new version metadata is illustrated in Figure 2.

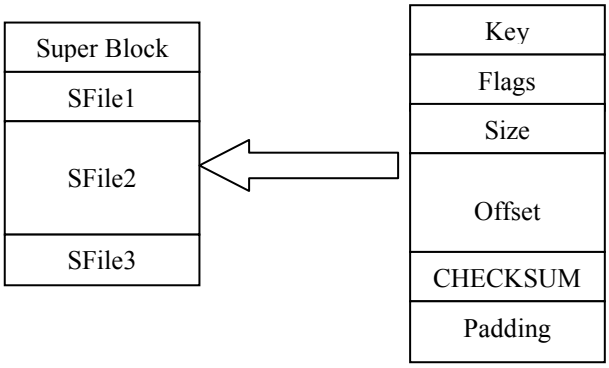


Fig. 2. Layout of SFile metadata

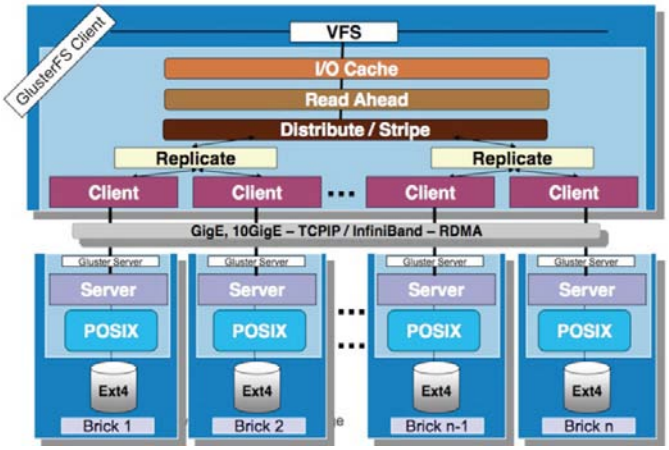


Fig. 3. Stack of translators in GlusterFS

Key stands for 64-bit file ID, Flag is used for deletion mark, Size is data size, Offset is actual file data location, CHECKSUM is used to check data integrity. Padding stores extra data to align SFile.

Volume management keeps open file descriptors for each physical disk sector and also a copy of small file ID to file metadata in-memory mapping.

B. *Details Of Implementations*

GlusterFS constructs its components with stack, it's called Translator in Gluster Domain. Figure 3 demonstrates all the Translators.

Add an extra layer above POSIX translator, called Acceleration translator. Acceleration layer handles upper layer's file name parse request, maintains in memory mapping of file ID to metadata. Figure 4 demonstrates Acceleration translator's component details. We divide this translator into four components: Request Router, Request Sender, Mapping Manager, File Compactor.

Request Router handles upper layer's file operations, parse request, file deletion request will be sent to File Compactor, file reading or writing request will be sent to both Request Sender and Mapping Manager. Balancing request work load is the most import task for Request Router.

Request Sender is responsible for sending requests to POSIX translator. It encapsulates POSIX APIs into simple READ\WRITE\DELETE operations. Mapping Manager and File Compactor fulfill their file related operation by calling Request Sender.

Mapping Manager maintains in memory mapping data. We stores metadata information in one specific log file, after system booting up, the first job of Mapping Manager is reading log file, initializing in memory mapping, and maintaining consistent view of files. Log file will be stored with backup replicas in every store machine.

File Compactor handles file deletion. Our application scenario is "Write once, Read often, never Modified and rarely delete", if we delete each file flagged to be deleted right after DELETE request handling, there will be a lot of holes inside big file, which is some kind of internal storage fragmentation. It's hard to maintain data integrity. Instead of doing this, we use File Compactor to delete file when the whole system's work load is not heavy, or when system is not busy. DELETE request will be fulfilled by change SFile's metadata Flag data sector.

When it comes to WRITE request, Acceleration layer initialize file name to id mapping, write its offset, data size metadata to SFile metadata node, determine which large file this small file is about to merge into.

When it comes to READ request, Acceleration layer translate filename into file ID, then it looks up the whole in memory map, gets offset and large file descriptor, if Flag field returns 1, indicates that this file has been deleted, then returns Not Found information to up layer, if it returns 0, sends read request to POSIX layer with large file descriptor and offset, reads actual data from disk. Thus file metadata look up requests at most one disk operation.

When it comes to DELETE request, Acceleration layer just modify the SFile's Flags field into value 1, when the whole system is not busy, Acceleration layer scans in memory map, compacts all the file to be deleted.

IV. EVALUATION

We evaluate our optimization in two sides: READ and WRITE. Figure 5 and 6 demonstrate the test result as follows. Our optimization targets at multiple files READ and WRITE. In the READ test, we try to get files' contents from GlusterFS Cluters, file amount increases from 10000 thousands to 60000 thousands, all the files' size is under 2 MB. Pure GlusterFS or the original one shows significant READ delay when file amount increases, while Optimized GlusterFS stay the same. In the WRITE test, we try to write data into GlusterFS Clusters, file amount increases from 10000 thousands to 60000 thousands. All files' size is under 2 MB. Pure GlusterFS delays with the increasing file amount, almost linear, and Optimized GlusterFS display horizontal line.

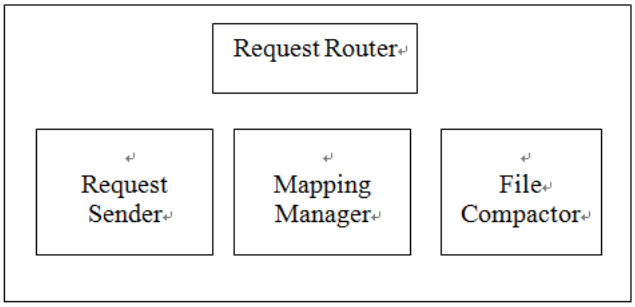


Fig. 4. Accelerator translator components overview

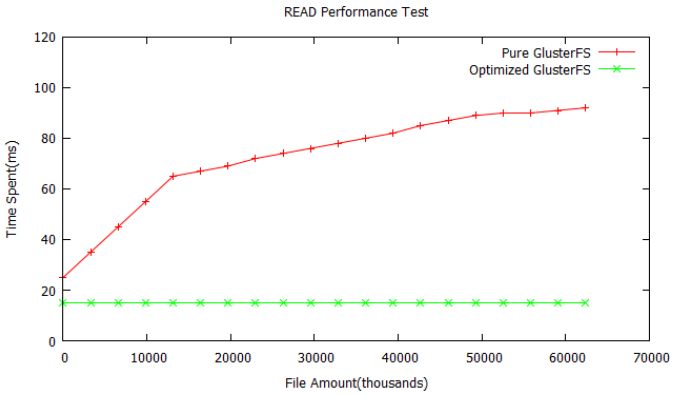


Fig. 5. READ Performance Test

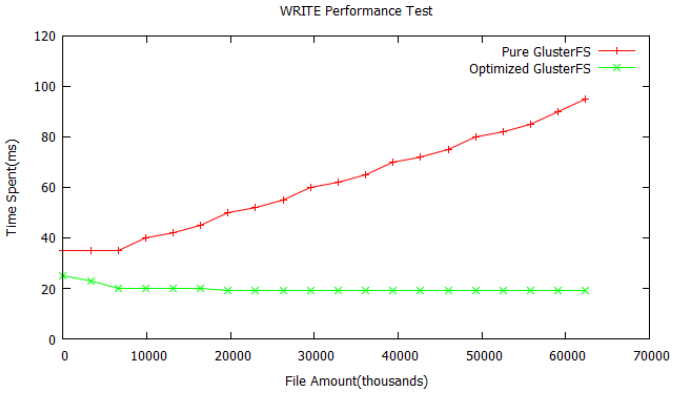


Fig. 6. WRITE Performance Test

Test results confirmed our assumptions, original GlusterFS stores data files directly into local file system, small amount (< 10000000) of files will not affect too much, but since each file needs at least one specific metadata, which is called inode under certain context, each inode consumes about 256 bytes of storage capacity, when file amount gets to some threshold, part of inode data will be swapped into disk, thus file looking up will require more disk operations, just like the performance test graph shows, after put more than 10000000 files into storage cluster, the speed of reading or writing file starts to slow down. While optimized GlusterFS shows quite different behavior, Acceleration layer merge small file into large file, even with more than 10000000 files stored in cluster, the speed of reading or writing file is barely changed. File amount

in cluster will not affect READ and WRITE performance with the help of Optimization.

V. Conclusion And Future Work

The paper describes a strategy aims at optimizing small file access performance in distributed file system. We redesigned structure of file metadata, minimized its size and merged small file into large file. This mechanism were implemented within GlusterFS, evaluation experimented on GlusterFS cluster shows a great acceleration on both WRITE and READ side. The performance is steady and reliable.

But there are some disadvantages of our work, which we should pay more attention to. Our implementation only applies for GlusterFS based platform, while there are still a lot of other distributed storage systems out there, such as Swift, FastDFS, Lustre, etc. Modifying all those system costs huge work load, and sometimes might compromise original good feature for distributed file system. As the matter of fact, bottleneck is local file system's file looking up, our next job will be extending local file system, like XFS, importing the whole strategy into source code, so that all the other distributed file systems could just exploit this specific feature of local file system to complete small file reading or writing performance's optimization.

Also, there is some other way to cut down the cost of metadata looking up. For example, instead of looking up in a table, we could use generating function like the one used in Ceph, so file data storage location will be outputted by one simple arithmetic calculation. This mechanism will certainly accelerate speed of file reading or writing, we should explore it in the future.

Acknowledgment

This work is sponsored by China National Power Grid Research Project 52272313507D.

References

[1] Lim. Hyeontaek, Bin. Fan, David G. Andersen, and Michael. Kaminsky. "SILT: A memory-efficient, high-performance key-value store." Proceedings of the Twenty-Third ACM Symposium on Operating Systems Principles. ACM, 2011.

[2] Doug. Beaver, Sanjeev. Kumar, Harry C. Li, Jason. Sobel, Peter. Vajgel. "Finding a Needle in Haystack: Facebook's Photo Storage." OSDI. Vol. 10. 2010.

[3] M. K. Aguilera, A. Merchant, M. Shah, A. Veitch, and C. Karamanolis. "Sinfonia: a new paradigm for building scalable distributed systems." ACM SIGOPS Operating Systems Review. Vol. 41. No. 6. ACM, 2007.

[4] Yu Ge, Kaneko. K, Bai. G, Makinouchi. A. "Transaction management for a distributed object storage system WAKASHI-design, implementation and performance." Data Engineering, 1996. Proceedings of the Twelfth International Conference on. IEEE, 1996.

[5] Mesnier, Mike, Gregory R. Ganger, and Erik Riedel. "Object-based storage." Communications Magazine, IEEE 41.8 (2003): 84-90.

[6] You, Lawrence L., Kristal T. Pollack, and Darrell DE Long. "Deep Store: An archival storage system architecture." Data Engineering, 2005. ICDE 2005. Proceedings. 21st International Conference on. IEEE, 2005.

[7] Mathur, Gaurav, P. Desnoyers, D. Ganesan. "Capsule: an energy-optimized object storage system for memory-constrained sensor devices." Proceedings of the 4th international conference on Embedded networked sensor systems. ACM, 2006.

[8] S. He and D. Feng. "Design of an object-based storage device based on i/o processor." SIGOPS Oper. Syst. Rev., 42(6):30–35, 2008.

[9] Rowstron, Antony, and Peter Druschel. "Pastry: Scalable, decentralized object location, and routing for large-scale peer-to-peer systems." Middleware 2001. Springer Berlin Heidelberg, 2001.

[10] R Sears, C Van Ingen, J Gray. "To BLOB or not to BLOB: Large object storage in a database or a filesystem?." arXiv preprint cs/0701168 (2007).

[11] Kim, Won. Introduction to object-oriented databases. Vol. 90. Cambridge, MA: MIT press, 1990.

[12] D. Zeinalipour-Yazti, S. Lin, V. Kalogeraki, D. Gunopulos, and W. A. Najjar. "MicroHash: An efficient index structure for flash-based sensor devices." In Proc. 4th USENIX Conference on File and Storage Technologies, Dec. 2005

[13] Darcy, Jeff. "Extending GlusterFS with python." Linux journal 2012.223 (2012): 2

[14] Developers, GlusterFS. "The Gluster web site." (2008).

An Adaptive Framework For Personalized Recommendation Algorithms

Jianchang Tang
School of Software Engineering
Shanghai Jiao Tong University
Shanghai, China
tjc0411@163.com

Xinhuai Tang
School of Software Engineering
Shanghai Jiao Tong University
Shanghai, China
tang-xh@cs.sjtu.edu.cn

***Abstract*—Different personalized recommendation algorithms are suitable for different scenarios. In this paper, we use artificial neural networks to implement an adaptive framework. When we add different recommendation algorithms into it and train it with the data from a given scenario, it can calculate the weight of each algorithm, choose suitable algorithms and give a more accurate prediction rating.**

Keywords—personalized recommendation algorithm; adaptive framework; artificial neural networks;

I. INTRODUCTION

With the explosive growth of the Internet and the e-commerce, recommender system is becoming more and more important. Recommender system can help users find the information or products they are interested in. It can be used in many different scenarios such as online shopping, watching movies, searching for information and so on.

Now there are so many personalized recommendation algorithms. In the Netflix million dollar grand prix, the champion used more than one hundred kinds of algorithms [1]. Different recommendation algorithms may get different efficiencies and different accuracies in different scenarios. There is no algorithm suitable for all situations.

Accurate and efficient recommender system can discover the potential consumption tendency of users, and improve the adhesion of users. When we need to design a recommender system, how should we choose the suitable algorithms to let the recommender system accurate?

In this paper, we design an adaptive framework for personalized recommendation algorithms. After adding the recommendation algorithms into this framework, it can automatic choose the suitable recommendation algorithms and help you build the recommender system.

You can first add the personalized recommendation algorithms you need into this framework. Then use the data from the field you choose to train it. After training, this framework can calculate the weight of each algorithm. And this framework has become a hybrid recommender system, as it automatic chooses the suitable algorithms and give a more accurate prediction rating.

We primarily use artificial neural networks to implement our framework. The output of each recommendation algorithm is the networks input. The networks output the prediction rating by using the chosen algorithms, and output the weight of each algorithm. Algorithm with the weight less than the threshold won't be chosen. Then we can use the prediction rating to recommend.

The rest of the paper is organized as follows. Section II introduces the hypothesis in this paper. Section III explains the method we used, the design and the implementation details of our framework. In Section IV, we provide the evaluation methodology and the results of the experimental evaluation. Finally, Section V contains some concluding remarks.

II. HYPOTHESIS

We assume that all the personalized recommendation algorithms add into the adaptive framework output the rating results. Because many algorithms those give a set of N recommendations also mark the items during runtime and choose the top N rating items [2]. Furthermore, a set of N recommendations can turn to N items with rating. We can use normal distribution or other methods to do this.

In this paper, we use prediction accuracy to evaluate the results output by recommendation algorithms [3-4]. As accuracy is the first one to be considered. And all of the output results are assumed ratings, prediction accuracy is effective. What's more, prediction accuracy corresponds to artificial neural networks. And it can be computed quickly. The evaluation beyond accuracy cannot be easily got.

III. AN ADAPTIVE FRAMEWORK FOR PERSONALIZED RECOMMENDATION ALGORITHMS

When we design our adaptive framework, our goal is to make this framework automatic choose the suitable recommendation algorithms and give a better recommendation. Framework can't do anything without info. So we need to train it with data. And we select ANN (artificial neural networks) to do this.

Why we choose ANN to implement our framework? Firstly, the input of this learning model is real-valued vector.

And the target function output is a real value. Secondly, users usually make a rating decision from feelings, so training data may contain errors. ANN learning method is very robust to the training data with noise. Although ANN requires a long training time, it's is typically fast to apply it to the subsequent instance. Above all any function can be approximated with arbitrarily small error by a network with three layers of units [5-6]. So after training ANN can get the target function.

A. Design Overview

Our learning task is to build a network which can give a more accurate prediction rating by using the recommendation from each algorithm. We use the back propagation algorithm to learn the weights of this multilayer network [7]. Figure 1 illustrates the design of the network graph structure. And we describe this structure below.

The first layer is an input layer. It is composed of input units. We can first train the algorithms we add into our framework before we start training ANN. Each input unit can get the prediction rating from the corresponding algorithm. And then input units transfer ratings to the next layer. Each input unit represents an algorithm. Input layer doesn't contain the weights need to be trained. It is only a transport layer.

The second layer is a hidden layer. It consists of three sigmoid units. Each unit gets the inputs from all of the input units. Then it calculates the output value and transfers it to the next layer.

The third layer is also a hidden layer. It consists of two sigmoid units. Each unit gets the inputs from all of the hidden units of the second layer. And it outputs the calculated value to the last layer.

Sigmoid unit calculates it output value by computing a linear combination of its inputs and applying the result to the logistic function $f(y) = 1 / (1 + e^{-y})$.

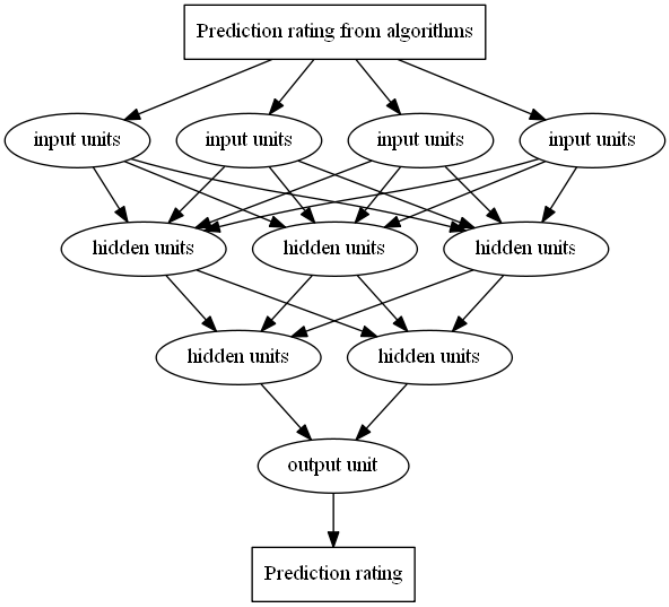


Fig. 1. Network Graph Structure

The fourth layer consists of only one unit. This unit first outputs the prediction rating with using all the given recommendation algorithms. And it calculates the weight of each algorithm. If there is a weight less than the threshold, it modifies the configuration file to set the corresponding algorithm invalid, and do the learning iterations again. After that, it only uses the chosen algorithms to calculate the prediction rating.

B. Implementation Details

In this part, we describe the details of the gradient descent weight update rule used by the back propagation algorithm.

We define o_i to be the output value of unit i, and tv_o to be the target value of the output unit. The input from unit i to unit j is denoted $x_{j\text{-}i}$, and the weight from unit i to unit j is denoted $w_{j\text{-}i}$. η is the learning rate, and α is a constant called the momentum.

First we create the feed-forward network as Figure 1 illustrate. Initialize all the weights to small random numbers. We set the random range from -0.05 to 0.05.

One iteration, we use one training data (may be repeating), a vector of prediction ratings given by the algorithms. We propagate the input forward through the network and calculate the output value o_i of every unit i. Then we propagate the errors backward through the network.

For the output unit o, its error term δ_o

$$\delta_o = tv_o - o_o \tag{1}$$

For each hidden unit h_{3i} from the third layer, its error term δ_{h3i}

$$\delta_{h3i} = o_{h3i}\,(1 - o_{h3i})\,w_{o\text{-}h3i}\,\delta_o \tag{2}$$

For each hidden unit h_{2j} from the second layer, its error term δ_{h2j}

$$\delta_{h2j} = o_{h2j}\,(1 - o_{h2j})\,\sum w_{h3i\text{-}h2j}\,\delta_{h3i} \tag{3}$$

The gradient with respect to the error, delta weight of *n*th iteration $\Delta w_{j\text{-}i}(n)$

$$\Delta w_{j\text{-}i}(n) = \eta\,\delta_j\,x_{j\text{-}i} + \alpha\,\Delta w_{j\text{-}i}(n-1) \tag{4}$$

And the weight is updated by adding the delta weight

$$w_{j\text{-}i} = w_{j\text{-}i} + \Delta w_{j\text{-}i}(n) \tag{5}$$

We will stop the iterative calculation when the error of two iterative results is less than the permissible error. If the iteration time is greater than the maximum permissible iteration time, we will also stop the iteration.

The weight of algorithm k W_{Ak} is calculated by using the input vector $(x_0, x_1, x_2, x_3 \ldots x_n)$, $x_i = 0$, $x_k = 3$. Then the output of the network is the weight of algorithm k.

If the $|W_{Ak}|$ less than the threshold 0.1, we will modify the configuration file and do the learning iteration again.

We set the permissible error to 0.0001, and set the maximum permissible iteration time to 12000. We set these

two parameters to avoid the overfitting problem. We can also use weight decay approach (decrease each weight by some small factor in each iteration) or cross-validation approach to overcome this problem [8-9].

We set the learning rate η to 0.1, and set the momentum α to 0.25. α is used for going through the small local minima and increasing the step size of the update in the same gradient. If these two parameters are set lower, it will take a longer training time. If we set these parameters too high, training will fail to converge with the acceptable error.

We set the number of hidden units to five. The second layer has three of them and the third layer has two of them. Because the number of the test algorithms we add to the framework is four. We test and find that such network structure performs best.

If we want to get the number of hidden units automatically, we can start with a network containing little units. Then add the number of hidden units until the network residual error is reduced to an acceptable level. We can use the cascade correlation algorithm or other methods to do this [10-11].

IV. EVALUATION

In this section we experimentally evaluate our framework. We use three recommendation algorithms from mahout and one random rating function to test our framework.

First algorithm is the user based recommender with Pearson correlation similarity. Second algorithm is the item based recommender with Tanimoto coefficient similarity [12]. Third algorithm is the SVD recommender with factorizes the rating matrix using ALS with Weighted-λ-Regularization [13]. The random rating function returns the rating by random with the seed itemID ∗ 1000 + userID.

We use the datasets from MovieLens. We divide the training data into three parts. The first part contains 90 percent of the data. This part is used to train the three recommendation algorithms. The second part contains 5 percent of the data. This part is used to train the ANN. The third part contains the rest of the data. This part is used to test and evaluate the ANN.

We use RMSE (Root Mean Square Error) to evaluate the prediction rating, because it is sensitive to the rating errors. With larger rating error, it has a larger influence than Mean Absolute Error.

We set the iteration time to 15000, and pick the output data per 100 iterations.

We want to know whether our framework can find the random rating function and modify the configuration file to set it invalid. Whether our framework can give a better result or not after removing the random rating function. And we want to see if our framework can give a more accurate prediction rating than the three recommendation algorithms.

Figure 2 shows the weights of three recommendation algorithms from mahout and one random rating function. They change with iteration. When iteration stops, our framework can find the random rating function.

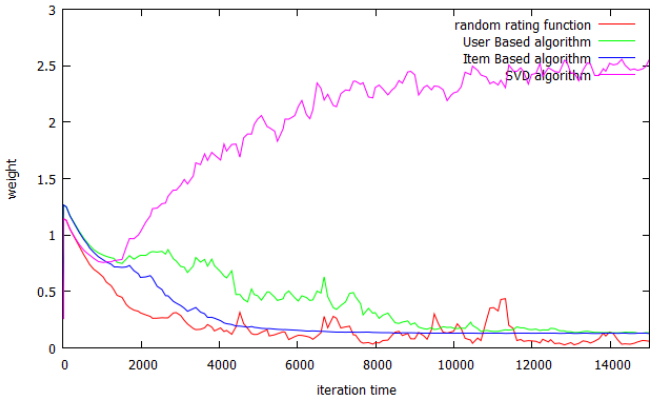


Fig. 2. Weight Of Each Algorithm

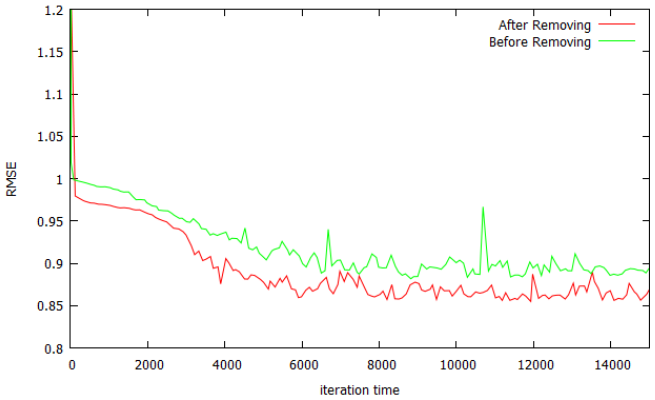


Fig. 3. Rating RMSE Before And After Set Invalid

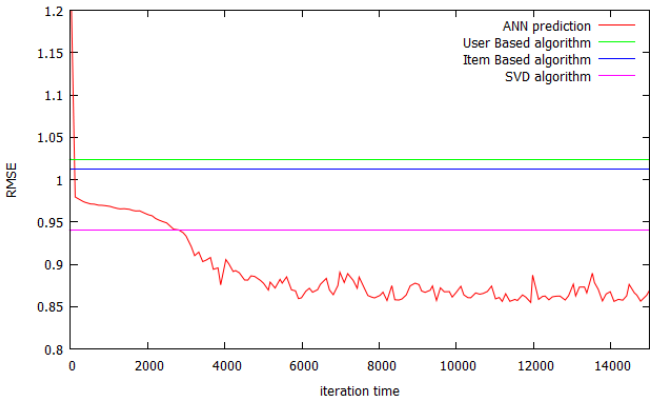


Fig. 4. Framework And Algorithms Rating RMSE

Figure 3 shows the rating RMSE before and after our framework removing the random rating function. After modified the configuration file and set the random rating function invalid, our framework performs better.

Figure 4 shows that after training our framework can output a better prediction rating than the three valid algorithms using in our test.

V. CONCLUSION

In this paper, we use the ANN to implement the adaptive framework. And we experimentally evaluate the framework. Our results shows that our framework can output a more accurate prediction rating for the next recommend step. And our framework can find the useless algorithm, and make our framework more succinct and efficient.

REFERENCES

[1] Bell R M, Koren Y, Volinsky C. The bellkor solution to the netflix prize[J]. 2007.

[2] Deshpande M, Karypis G. Item-based top-n recommendation algorithms[J]. ACM Transactions on Information Systems (TOIS), 2004, 22(1): 143-177.

[3] Herlocker J L, Konstan J A, Terveen L G, et al. Evaluating collaborative filtering recommender systems[J]. ACM Transactions on Information Systems (TOIS), 2004, 22(1): 5-53.

[4] Geyer-Schulz A, Hahsler M. Evaluation of recommender algorithms for an internet information broker based on simple association rules and on the repeat-buying theory[C]//proceedings WEBKDD. 2002: 100-114.

[5] Huang G B. Learning capability and storage capacity of two-hidden-layer feedforward networks[J]. Neural Networks, IEEE Transactions on, 2003, 14(2): 274-281.

[6] Selmic R R, Lewis F L. Neural-network approximation of piecewise continuous functions: application to friction compensation[J]. Neural Networks, IEEE Transactions on, 2002, 13(3): 745-751.

[7] Hecht-Nielsen R. Theory of the backpropagation neural network[C]//Neural Networks, 1989. IJCNN., International Joint Conference on. IEEE, 1989: 593-605.

[8] Krogh A, Vedelsby J. Neural network ensembles, cross validation, and active learning[J]. Advances in neural information processing systems, 1995: 231-238.

[9] Amari S I, Murata N, Muller K R, et al. Asymptotic statistical theory of overtraining and cross-validation[J]. Neural Networks, IEEE Transactions on, 1997, 8(5): 985-996.

[10] Hirose Y, Yamashita K, Hijiya S. Back-propagation algorithm which varies the number of hidden units[J]. Neural Networks, 1991, 4(1): 61-66.

[11] Fahlman S E, Lebiere C. The cascade-correlation learning architecture[J]. 1989.

[12] Lipkus A H. A proof of the triangle inequality for the Tanimoto distance[J]. Journal of Mathematical Chemistry, 1999, 26(1-3): 263-265.

[13] Zhou Y, Wilkinson D, Schreiber R, et al. Large-scale parallel collaborative filtering for the netflix prize[M]//Algorithmic Aspects in Information and Management. Springer Berlin Heidelberg, 2008: 337-348.

An Improved Online Multiple Kernel Classification Algorithm Based on Double Updating Online Learning

Yulin Xiao
College of Mathematics and Computer Science
Fuzhou University
Fuzhou,350108,China
e-mail:N130320048@fzu.edu.cn

Shangping Zhong
College of Mathematics and Computer Science
Fuzhou University
Fuzhou,350108,China
e-mail:spzhong@fzu.edu.cn

***Abstract*—Online multiple kernel classification(OMKC) algorithm is a promising algorithm in machine learning. Because of low error rate and relatively fast training time, it has been sucessfully applied to many real-world problems. However, in the phase of learning a single classifier for a given kernel, the OMKC adopts the perceptron algorithm, which significantly limits the performance of the algorithm. In this paper, we adopts the double updating online learning(DUOL) algorithm to learn the single classifier. Compared to the perceptron algorithm, the DUOL algorithm not only assigns a weight to the misclassified example, but also updates the weight for one of the existing support vectors, which significantly improves the classification performance. Then we use the hedge algorithm to combines these classifiers. The experimental results show that the proposed algorithm is more effective than the OMKC algorithm, the state-of-the-art algorithms, and single kernel learning algorithm.**

Keywords—online learning; multiple kernel learning; DUOL; OMKC

I. INTRODUCTION

Online learning and kernel learning are two active research topics in machine learning. The goal of online learning is to learn a prediction model from a sequence of data examples with time stamps. In general, online learning has the advantages of fast, simple, and often make few statistical assumption, etc[1]. And the goal of kernel learning is to learn an effective kernel function from training data. Because of the training time relatively faster and lower error rate compared to other machine learning method, kernel-based methods have been successfully applied to various real-world problems[2, 3]. However, in some complicated cases, the single kernel method is not enough to meet many practical requirements, such as heterogeneous information, unnormalised data, non-flat distribution of samples, etc. So it is significant to find the optimal combination of multiple kernels to achieve higher flexibility and better capability in solving real-world problems.

In recent years, a lot of multiple kernel learning method have been proposed[4-7]. One of the latest multiple kernel learning algorithm is the Online Multiple Kernel Classification(OMKC)[7], it combines the advantages of online learning and multiple kernel learning. It has higher classification performance than many conventional multiple kernel learning algorithms. It fuses two kinds of online learning techniques: the Perceptron algorithm[8] and the Hedge algorithm[9]. The goal of the Perceptron algorithm is to learn a classifier for a given kernel function, and the goal of the hedge algorithm is to combines the multiple classifiers.

However, the performance of the perceptron algorithm is not enough. For a misclassified example, the perceptron algorithm only simply assign to it a fixed weight, and the weight remains unchanged during the whole learning process, which has significant limitations. This is because when add a new misclassified example to support vector pool, the weights of the support vector in the pool may be no longer optimal, and they should be updated to fit the new misclassified example[10]. The Double Updating Online Learning(DUOL)[10] exactly solve the problem. For a misclassified example, the DUOL algorithm not only assigns a weight to the example, but also updates the weight for one of the existing support vector in the pool.

Motivated by the above observations, we propose an improved online multiple kernel learning algorithm. The algorithm is based on the DUOL and the Hedge algorithm. The DUOL algorithm is employed to learn a classifier for a given kernel, and the Hedge algorithm combines the multiple classifiers. Test on 10 datasets, the proposed algorithm shows promising performance compared to the OMKC algorithms and the state-of-the-art algorithms for online kernel learning.

The rest of this paper is organized as follows. Section 2 briefly introduces the online multiple kernel learning algorithms and the double updating learning algorithm. Section 3 presents our proposed algorithm for online multiple kernel learning. Section 4 presents the experimental results on 10 datasets. Section 5 concludes this paper.

II. A BRIEF REVIEW OF OMKC AND DUOL

This section briefly introduces the online multiple kernel classification(OMKC) algorithm and the double updating learning algorithm(DUOL).

A. The Online Multiple Kernel Classification(OMKC)

Compare to single kernel methods, multiple kernel methods is a more flexible learning model. The recent theory and application have proved that using multiple kernel will enhance the interpretability of the decision function, and using multiple kernel methods often can get better performance[11-13]. In multiple kernel model, the commonest way is to find the optimal combination of multiple predefined kernels. Because of the multiple kernel methods use the mapping ability of all basic kernel, it solves the problem of selecting the kernel parameters for kernel target alignment and kernel function selection well.

Among all multiple kernel methods, one of the latest multiple kernel methods is the Online Multiple Kernel Classification (OMKC)[7]. It is more effective than many other multiple kernel learning algorithms. In general, it is more challenging than typical online learning algorithms, because both the classifiers and the subset of selected kernels aren't know, and more importantly, the solutions to the kernel classifiers and their combination weights are correlated[7]. To solve this problem, OMKC based on two learning algorithms: the perceptron algorithm and the Hedge algorithm. The perceptron algorithm is used to learn a kernel classifier with some selected kernel, and the Hedge algorithm is used to combine these learned classifiers. Algorithm 1 shows the detailed steps of the OMKC.

Algorithm 1: The general framework for OMKC

1: **INPUT**:
   --- kernels: $k_i(\cdot,\cdot): x \times x \to R, i = 1,...,m$
   --- Weights $w_i(1) = 1, i = 1,...,m$
   --- Discount weight $\beta \in (0,1)$.
2: **Initialization**: $f^1 = 0, w^1 = 1, \theta^1 = \frac{1}{m}\vec{1}$
3: **for** t=1,2, ..., T **do**
4: Receives a new instance: $x_t \in X$;
5: Predict: $\tilde{y}_t = sign(f_{t-1}(x_t; w_t))$;
6: Receive its true label $y_t \in Y$;
7: **for** $i = 1,2,...,m$ do
8: Set $z_i^t = I(y_t f_i^t(x_t) \le 0)$
10: Update $f_i^{t+1}(x) = f_i^t(x) + z_i^t y_t k_i(x_t, x)$
11: **end for**
12: $\theta_i^{t+1} = \frac{w_i^t}{w_t}, i = 1,...,m$, where $W_t = \sum_{i=1}^m w_i^t$
13: **end for**

where $w_i(t)$ denotes the combination weight for the *i-th* kernel classifier at round *t*. It is updated by Hedge algorithm.

B. The Double Updating Online Learning(DUOL)

A lot of online learning algorithms have been proposed in recent years [8][14-17]. The Perceptron algorithm[8] is one of the most well-known online learning algorithm, but it has significant limitations. The Double Updating Online Learning(DUOL) [10] exactly solve the perceptron's defect.

For a general online learning algorithm, according to [10], when add a misclassified example to the classification function, the improvement of the objective function denote by Δ_t, if an online learning algorithm is designed to ensure that for all t, Δ_t is bounded from below by a bounding constant Δ, then the number of mistakes made by the algorithm denoted by M is upper bounded by:

$$M \le \frac{1}{\Delta}(\min_{f \in H_k} \frac{1}{2}\|f\|_{H_k}^2 + C\sum_{i=1}^{T} l(y_i f(x_i)) \qquad (1)$$

Where $l(y_i f(x_i)) = \max(0, 1 - y_i f(x_i))$ is the hinge loss function. According to [18, 19], the bounding constant $\Delta = 1/2$ when only update the classifier with the newly misclassified example.

In contrast to the above, the DUOL algorithm not only assigns a weight to the example, but also updates the weight for one of the existing support vector in the pool, which significantly improves the classification. According to [10]'s analysis, the number of prediction mistakes M made by DUOL on a sequence of examples is bounded by:

$$2(\min_{f \in H_k} \frac{1}{2}\|f\|_{H_k}^2 + C\sum_{i=1}^{T} l(y_i f(x_i)) - \frac{\rho^2}{2} M_d^w(\rho) - \frac{1+\rho}{1-\rho} M_d^s(\rho) \quad (2)$$

where $\rho \in [0,1)$.

It is obvious that the number of mistakes made by the DUOL algorithm is smaller than the general online learning algorithm that only performs a single update in each trial, such as the Perceptron algorithm.

III. OUR PROPOSED METHOD

A. Theory Analysis

OMKC algorithm has a good performance on classification task, its classification accuracy is better than many classic online learning algorithms, such as single kernel perceptron algorithm, the state-of-the-art online multiple kernel algorithms, etc. However, in the phase of training classifier, OMKC use the perceptron algorithm, which is really not enough. Because for a misclassified example, the perceptron algorithm only simply assign a fixed weight to the example. The weight will not change during the whole process. Although such method has advantages in computational efficiency, it has significant limitations. Because when add a new misclassified example to the support vector pool, the weights of support vector in the pool may no longer optimal, so they should be updated to fit the new misclassified example.

The DUOL algorithm exactly solves the above problem. When add a new misclassified example to the support pool, the DUOL algorithm not only updates the weight of the misclassified example, but also adjusts the weight of one existing support vector which is most seriously conflicts with the new support vector. Therefore, for any kernel function, the performance of the classifier trained by DUOL algorithm

are better than the classifier trained by the perceptron algorithm. Then, using the Hedge algorithm combine the classifier trained by DUOL, the performance of our proposed algorithm is better than OMKC algorithm.

B. The Proposed Method

Motivated by the above observations, we combine advantages of the DUOL algorithm and the $OMKC_{(S,D)}$, proposed a new online multiple kernel algorithm. The algorithm fuses the DUOL algorithm and the Hedge algorithm. The algorithm can be divided into two phases: the first phase is using the DUOL algorithm to learn a classifier for a given kernel, and the second is using the Hedge algorithm to combine all classifier. Algorithm 2 shows the detailed steps of the proposed algorithm.

Algorithm 2: The proposed algorithm

INPUT:
--- kernels: $k_i(\cdot,\cdot): x \times x \to R, i = 1,...,m$
--- Weights $w_i(1) = 1, i = 1,...,m$
--- Discount weight $\beta \in (0,1)$.
Initialization: $f^1 = 0, w^1 = 1, \theta^1 = \frac{1}{m}\vec{1}$
for t=1,2, ..., T **do**
  Compute $q_i(t) = w_i(t) / [\max_{i \le j \le m} w_j(t)], i = 1,...,m$;
  Compute $p_i(t) = (1-\delta) q_i(t) + \delta / m, i = 1,...,m$
  Receives a new instance: $x_t \in X$;
  %Use the DUOL algorithm to lean a classifier for a given kernel
  for $i = 1,2,...,m$ **do**
    Predict: $\tilde{y}_i^t = sign(f_i^{t-1}(x_t))$;
    Receive its true label $y_t \in Y$;
    Calculate: $l(y_t, \tilde{y}_t) = \max\{0, 1 - y_t^t f_i^{t-1}(x_t)\}$;
    if($l_t > 0$)
      Search for auxiliary example;
      According to formula to calculate the optimal value of γ_a and d_{γ_b};
      Update the support vectors and the corresponding weights;
    end if
  end for
  %Use Hedge algorithm to update the weight of each classifier
  for $i = 1,2,...,m$ **do**
    Sample $s_i(t) = Bernoulli\_Sampling(p_i(t))$;
    Set $z_i^t = 1$ if $I(y_t f_i^t(x_t) \le 0$ and 0 otherwise;
    Update $w_i^{t+1} = w_i^t \beta^{z_i^t s_i(t)}$;
    Update $f_i^{t+1}(x) = f_i^t(x) + s_i(t) z_i^t y_t k_i(x_t, x)$
  end for
end

IV. EXPERIMENT

In this section, we first describe the experimental setups. Then we show that our method outperforms previous methods.

A. Experimental Setups

To evaluate the effectiveness of our method, ten sets of two-class problem experiments are conducted. The ten two-class standard datasets are obtained from LIBSVM[20] and UCI machine learning repository[21]. These datasets were chosen quite arbitrarily. In our experimental, we predefine a pool of 16 kernel functions, including 13 Guassian kernels of kernel width parameter σ in [2^{-6}, 2^{-5}, …, 2^6], and 3 polynomial kernels (i.e., $k(x_i, x_j) = (x_i^T x_j)^P$) of degree parameter p=1, 2, and 3.

To obtain the stable average results, all the experiments were conducted over 20 random permutations for each dataset, and all the reported results were averaged over these 20 runs. Our experiments were evaluated on a PC with Intel(R) Core(TM) i5-2310 @ 2.90GHz CPU and 16GB RAM by Matlab.

B. Comparison with Previous Methods

We compare the proposed algorithm with the following baseline algorithm:
----DUOL: a single kernel online learning algorithm in [10].
----OM-2: a state-of-the-art online learning algorithm for multiple kernel learning[22].
----OMKC: It includes four algorithms, they are $OMKC_{(D, D)}$, $OMKC_{(D, S)}$, $OMKC_{(S, D)}$, $OMKC_{(S, S)}$[7].

Table I summarizes the average experimental results of the proposed algorithm and others. We can draw several observations from the results: First, the proposed algorithm can significantly improve the performance of the DUOL algorithm. Second, the proposed algorithm has a better performance on classification task than other multiple kernel learning algorithms, including OM-2 algorithm and the family of OMKC algorithm. Third, as for running time, since the proposed algorithm requires double updates, it is less efficient than OM-2 and OMKC algorithm; And since the proposed algorithm need to combine multiple kernel classifier, it is less efficient than DUOL algorithm. But in general, the proposed algorithm has the best classification performance among all kernel learning algorithm.

In fig. 1, we report the classification performance of DUOL algorithm, OM-2 algorithm, OMKC algorithm and our proposed algorithm on ten datasets. It is obvious that, for all ten datasets, our proposed algorithm has lower error rate than other kernel method.

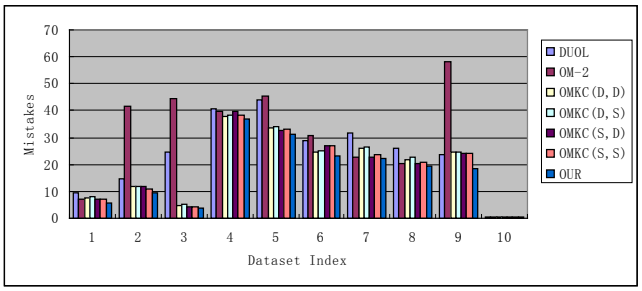


Fig. 1. The comparison of the experiment results

TABLE I. THE EXPERIMENT RESULTS OF OUR PROPOSED ALGORITHM AND OTHER ONLINE LEARNING ALGORITHMS

DataSets			DUOL	OM-2	$OMKC_{(D,D)}$	$OMKC_{(D,S)}$	$OMKC_{(S,D)}$	$OMKC_{(S,S)}$	Our
1	Votes84	Mistake(%)	9.560	7.210	7.370	8.210	6.940	7.010	**5.790**
		Times(s)	0.007	0.060	0.199	0.210	0.192	0.128	0.579
2	Ionoshpere	Mistake(%)	14.680	41.700	11.700	11.800	11.840	11.110	**9.380**
		Times(s)	0.011	0.102	0.305	0.314	0.252	0.146	1.017
3	Breast	Mistake(%)	24.800	44.330	4.850	5.410	4.110	4.370	**3.820**
		Times(s)	0.021	0.132	0.321	0.336	0.294	0.175	0.701
4	Australian	Mistake(%)	40.620	39.620	37.670	38.300	39.550	38.510	**37.090**
		Times(s)	0.028	0.134	0.405	0.417	0.372	0.280	3.870
5	Diabetes	Mistake(%)	43.800	45.350	33.690	34.000	32.510	33.060	**31.330**
		Times(s)	0.035	0.153	0.438	0.453	0.410	0.315	4.020
6	Splice	Mistake(%)	28.810	30.790	24.590	25.120	26.860	26.980	**22.940**
		Times(s)	0.039	0.200	0.588	0.596	0.524	0.365	6.106
7	Svmguide3	Mistake(%)	31.830	22.840	26.000	26.550	22.820	23.750	**22.000**
		Times(s)	0.061	0.339	0.718	0.732	0.663	0.486	4.516
8	A3a	Mistake(%)	26.060	20.230	21.960	22.480	20.330	20.660	**19.420**
		Times(s)	0.196	1.303	2.445	2.486	1.929	1.317	25.458
9	Spambase	Mistake(%)	23.550	58.160	24.360	24.780	24.170	24.100	**18.240**
		Times(s)	0.359	2.921	4.999	5.027	3.406	2.191	37.578
10	Mushrooms	Mistake(%)	0.260	0.370	0.330	0.380	0.290	0.350	**0.240**
		Times(s)	0.118	4.997	5.057	5.206	3.421	1.472	6.533

V. CONCLUSION

In this paper, we propose an improved online multiple kernel learning algorithm based on OMKC. It combines two types of online learning algorithms: the DUOL algorithm and the Hedge algorithm. The experiment results show that, for all ten datasets, the performance of our proposed algorithm are better than DUOL algorithm, OM-2 algorithm, and OMKC algorithm.

REFERENCES

[1] Hoi, Steven CH, Jialei Wang, and Peilin Zhao. "LIBOL: a library for online learning algorithms," The Journal of Machine Learning Research , vol.15.1, pp. 495-499, 2014.

[2] Schölkopf, Bernhard, and Alexander J. Smola. “Learning with kernels: support vector machines, regularization, optimization, and beyond,” MIT press, 2002.

[3] Shawe-Taylor, John, and Nello Cristianini. “Kernel methods for pattern analysis,” Cambridge university press, 2004.

[4] Rakotomamonjy, Alain, et al. "SimpleMKL," Journal of Machine Learning Research, vol.9, pp.2491-2521, 2008.

[5] Sonnenburg, Sören, et al. "Large scale multiple kernel learning," The Journal of Machine Learning Research, vol.7, pp.1531-1565, 2006.

[6] Xu, Zenglin, et al. "An extended level method for efficient multiple kernel learning," Advances in neural information processing systems, 2009.

[7] Hoi, Steven CH, et al. "Online multiple kernel classification," Machine Learning, vol. 90.2, pp. 289-316, 2013.

[8] Rosenblatt, Frank. "The perceptron: a probabilistic model for information storage and organization in the brain," Psychological review , vol. 65.6 , pp. 386., 1958.

[9] Freund, Yoav, and Robert E. Schapire. "A desicion-theoretic generalization of on-line learning and an application to boosting," Computational learning theory. Springer Berlin Heidelberg, 1995.

[10] Zhao, Peilin, Steven CH Hoi, and Rong Jin. "Double updating online learning," The Journal of Machine Learning Research, vol.12, pp 1587-1615, 2011.

[11] Jin, Rong, Steven CH Hoi, and Tianbao Yang. "Online multiple kernel learning: Algorithms and mistake bounds," Proc. 21st Int’l Conf. Algorithmic Learning Theory, vol.6331, pp.390-404, 2010.

[12] Zhuang, Jinfeng, Ivor W. Tsang, and Steven Hoi. "Two-layer multiple kernel learning," The Journal of Machine Learning Research, vol.15, pp. 909-917, 2011.

[13] Zhuang, Jinfeng, et al. "Unsupervised multiple kernel learning," The Journal of Machine Learning Research, vol. 20, pp.129-144, 2011.

[14] Crammer, Koby, and Yoram Singer. "Ultraconservative online algorithms for multiclass problems," The Journal of Machine Learning Research, vol.3, pp. 951-991, 2003.

[15] Cesa-Bianchi, Nicolo, Alex Conconi, and Claudio Gentile. "On the generalization ability of on-line learning algorithms," IEEE Transactions on Information Theory, vol.50.9, 2050-2057, 2004.

[16] Crammer, Koby, et al. "Online passive-aggressive algorithms," The Journal of Machine Learning Research, vol.7, pp.551-585, 2006.

[17] Fink, Michael, et al. "Online multiclass learning by interclass hypothesis sharing," Proceedings of the 23rd international conference on Machine learning. ACM, 2006.

[18] Shalev-Shwartz, Shai, and Yoram Singer. "Online learning meets optimization in the dual," In Proceedings of the 19th Annual Conference on Learning Theory, vol.4005, pp.423-437, 2006.

[19] Shalev-Shwartz, Shai, and Yoram Singer. "A primal-dual perspective of online learning algorithms," Machine Learning, vol.69.2-3, pp.115-142, 2007.

[20] Chih-Jen Lin, LIBSVM Data: Classification, Regression, and Mult-label. http://www.csie.ntu.edu.tw/~cjlin/libsvmtools/datasets.

[21] UCI Machine Learning Repository. http://www.ics.uci.edu/~mlearn/.

[22] Jie, Luo, Francesco Orabona, Marco Fornoni, Barbara Caputo, and Nicolo Cesa-Bianchi. "OM-2: An online multi-class multi-kernel learning algorithm," In Proc. Of the 4th IEEE online learning for computer vision workshop, 2010.

Comprehensive Evaluation of Cross-platform TV Shows Research

Lu Lu,Yin Fulian, Chai Jianping
Communication University of China
Beijing
b-lucy-ll@hotmail.com

Lin Jiecong
Beijing Institute of Computer Application
Beijing
linjiecong@163.com

***Abstract*—The traditional ratings survey method has various problems under the environment of 'media integration', such as the form unitarily and no in-depth analysis. For these problems, the paper presents a comprehensive cross-platform television program evaluation system. This comprehensive evaluation system has the characteristics of multi-service targets, multi-platform, multi-method, multi-dimensional, multi-evaluate content. It can evaluate TV shows more scientific and comprehensive compared with the existing program evaluation method.**

Keywords—evaluation; TV shows; index; cross-platform

I. INTRODUCTION

In the era of 'media integration', TV media will encounter opportunities and challenges at the same time [1]. As an important criterion for evaluation of TV shows, audiences rating can only be obtained through simple user survey as traditional broadcast television is unidirectional coverage. China Central Television issued < CCTV program evaluation system plan> in 2002. The system is characterized by 'three indicators, a ruler' and is commonly known as 'elimination system' [2]. In 2011 CCTV introduced a new program evaluation system, which is divided into four categories: guiding force, influence, communication ability and professionalism [3]. Britain is the first country to start using qualitative research methods research program quality. BBC launched an evaluation system with the goal of 'Public value' in 2004 and the 'Public value' is decomposed into value elements, broadcasting mission and measure way. The measures of public value are the daily conduct of "continuous performance evaluation system", concluding touch of rate, quality, influence, and investment value [8, 9, 10]. From the late 1960s, Communication Research Office of the United States also began a series of programs to promote qualitative research projects. American commercial television evaluation system consists of two parts: evaluation before and after broadcasting.

With the increasing of audience's network viewing behavior, the defects of traditional method of investigation on ratings in the new media environment is slowly rendering, and it puts forward new requirements of television program evaluation system. Current international mainstream program evaluation methods still concentrate on single platform of broadcasting and television. As one of the most important aspects of the triple play, interaction and fusion of broadcasting and television, the Internet and telecommunications network will enable the television industry to undergo tremendous changes. Under this background, China should rely on the successful experience of the evaluation system at home and abroad and urgently needs to introduce new monitor perspectives and metrics in order to evaluate TV program more comprehensive and deeply.

II. EVALUATION SYSTEM ARCHITECTURE

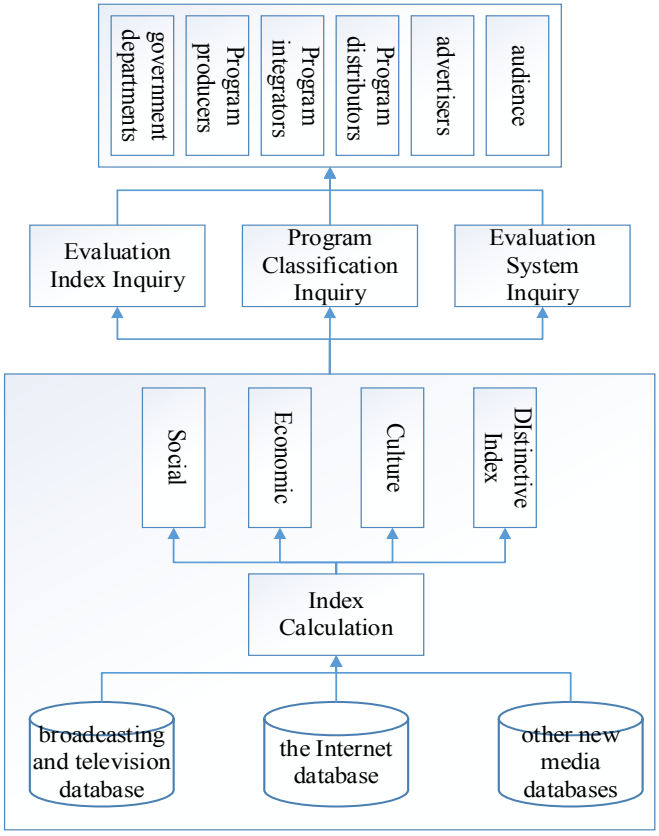


Fig 1. Program evaluation system architecture diagram

Program evaluation system architecture includes user layer, application layer, technology layer and data layer.

A. User Layer

User layer describes the object of service. Traditional broadcast and television industry chain consists of government departments, advertisers, program producers / integrators, program distributors and audience. From the perspectives of industry chain, service object of TV shows evaluation can be government departments, program producers / integrators, program distributors, advertisers and audience.

Government departments publicize for the party and the government's voice, lead the development trend of advanced culture, respond to sudden and major events and guide public opinions. They need to make sure whether the public opinion of the program is correct.

The main product of broadcasting and television industry is the program producers / integrators provided through plan and production. Program producers / integrators need to collect audience's feedback of program content and achieve a market-oriented program production process.

Program distributors' responsibilities are to integrate programs and present to the audience. At present the main program distributors are CCTV, provincial and municipal TV stations premium channels publishing units. TV stations play an important role of choosing, scheduling and displaying programs.

As a derivative of modern media development, advertising has occupied a pivotal position in the field of broadcasting and television industry. Advertisers need to find the target groups, determine the viewing time and platform according to the characteristics of their products in order to achieve targeted product launches.

Audience is the only beneficiaries of broadcasting and television industry. And the ultimate goal of broadcasting and television industry is to serve audience.

B. Application Layer

Application layer describes available application services. The evaluation system can provide inquiry service of evaluation index, program classification and program evaluation system.

We designed a tertiary-level evaluation index system for TV shows evaluation. The first level of index is social, economic and cultural. Under each first level of index, there are refined secondary level of indexes. For example, under the index of 'social ', there are four secondary level of indexes which are guiding force, influence, communication ability and competitiveness. Under each secondary level of index, there are also refined third level of indexes. Every index is explained in detail and the calculation method is presented. Users can query the explanations and the values of the indexes.

Currently there are many different kinds of TV shows. It will be obviously biased if we use a unified standard to evaluate. Different types of TV shows should be evaluated in different evaluation standards. For example, for news programs, opinion-oriented is particularly important, so the proportion of guiding force should be increased. Users can query the evaluation results for different types of programs.

This system provides the overall query of evaluation system architecture so that users can have a deeply understanding of the evaluation indexes and inter-related effects.

C. Technology Layer

Technology layer describes the technical framework of evaluation system. And it's the core of the program evaluation system. This evaluation system gives a multi-dimensional analysis in the aspects of social, economic and cultural. First of all, the system sets up an independent evaluation system for each platform; then it establishes a cross-platform evaluation system according to the common indexes; in addition, it establishes distinctive evaluation index for special needs.

Social dimension evaluates social value of TV shows. Social value refers to the responsibility of meeting the material or spiritual needs of society or people through themes and contents of the programs which TV shows need to take. Economic dimension evaluates economic contribution of TV shows, which refers to the economic significance for social and humans, and these are what the measurement of benefit economic actors can obtain from products and services. Culture dimension evaluates leading role in culture of TV shows, which refers to the culture significance for "people's cultural nature and needs", expressed as a cultural capital, the cost of a culture and a cultural force [13].

D. Data Layer

Data layer describes the required program evaluation data, including broadcasting and television database, the Internet database, micro blogging database, forums database, IPTV database and other new media databases. Broadcasting and television database contains basic viewing data, such as duration of watching and channels, but also includes cost data. Internet database contains relevant the data of page views, time and so on. New media database contains their viewing data on IPTV, hand held terminals and other new media.

III. EVALUATION INDEX

Broadcasting and television program evaluation system is a sub-system of Chinese media program evaluation index system architecture. We propose a TV shows evaluation system, aiming at establishing a comprehensive scientific program evaluation criterion.

Program evaluation index system concludes three first level of indexes: social, economic and culture.

- Social dimension evaluates social value of TV shows. Social value refers to the responsibility of meeting the material or spiritual needs of social or others through theme and content of the TV shows need to take.
- Economic dimension evaluates economic contribution of TV shows, which refers to the economic significance for society and humans, which the measurement of benefit economic actors can obtain from products and services.
- Culture dimension evaluates leading role in culture of TV shows, which refers to the culture significance for "people's cultural nature and needs", expressed as a

cultural capital, the cost of a culture and a cultural force.

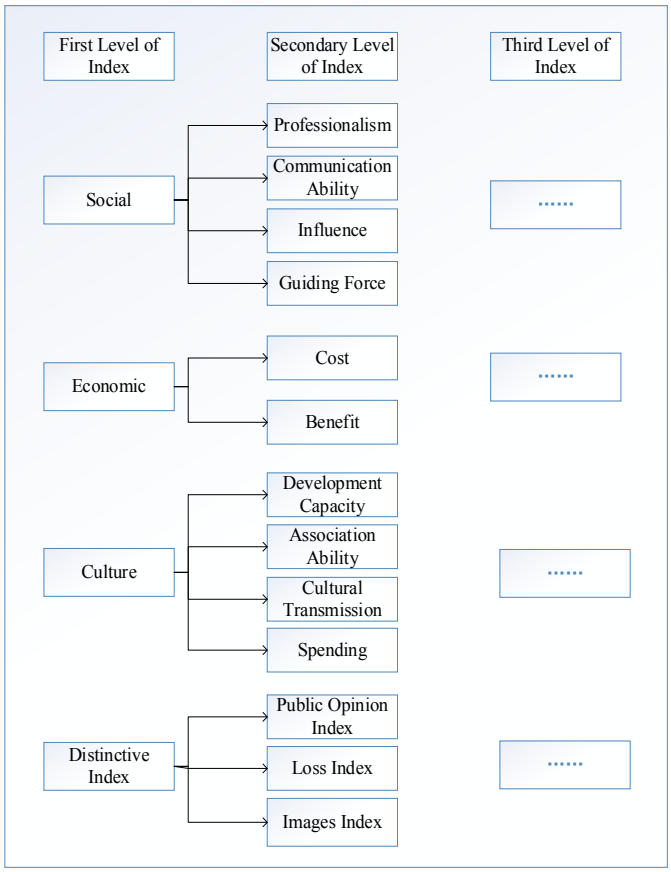


Fig 2. TV shows evaluation index system

The secondary level of indexes in social includes guiding force, influence, communication ability and competitiveness.

- Guiding force shows that whether TV shows play a correct leading role in social life.
- Influence shows the ability that whether TV shows can affect audience thought and action using the way audience willing to accept.
- Communication ability shows the capability to achieve effective dissemination.
- Competitiveness shows the comprehensive ability reflected in the multi-column comparison.

The secondary level of indexes in economy consists of cost and benefits.

- Cost refers to the monetary performance of consuming resource of production of programs.
- Benefit refers to direct benefits obtained after program broadcasting and indirect benefits arising from the program.

The secondary level of indexes in culture includes the development capacity, association and cultural transmission and spending.

- Development capacity refers to TV shows' potential ability of development and growth.
- Association refers to the ability to connect other things with some programs.
- Cultural transmission power refers to TV shows fusion ability of foreign culture, inheritance ability of local cultural and exploration ability of new culture.
- As the product of cultural spending, those programs represent the value and symbolic meaning.

There are also some three-leveled indexes. For example, there are three indexes in the development capacity, which include the completion rate, arrival rate and others in audience rating. According to the Internet, the standard can be divided into CTR (click-through rate), page view and other indexes. First of all, we set up an independent evaluation system for each platform; then we establish a cross-platform evaluation system according to common indicators; in addition, we also establish distinctive evaluation index for special needs.

IV. Weight Assignment of Evaluation Index

To make the evaluation composed of varies indexes reflects the real situation in a more exact way, it is important to give the transformed index values different weight coefficients. And in this paper the weighting methods integrated with subjectivity and objectivity will be discussed to evaluate the indexes.

A. Subjective Weighting Method

The subjective weighting method refers to the professional knowledge, practical experience through the subjective analysis method to determine the weight of evaluation index. In this paper, we use the Delphi method, which means the method of expert opinion to determine the subjective weights. The method is based on the established process, and make comments anonymously, then fill in the questionnaires again and again. And finally get the weight value. The final weight calculation according to the formula (1):

$$w_i = \frac{\sum_{j=1}^{n} E_{ij}}{n} \tag{1}$$

In the formula: w_i is the weight of the index i; E_{ij} for the expert j's score on index i, and n means the total number of experts.

B. Objective Weighting Method

Objective weighting method is directly based on a method of original information of each index after a certain mathematical processing gain weight. This paper uses variation coefficient method to determine the objective weight. Variation coefficient method refers to the method to determine the weight value of evaluation indexes according to the degree of variation in the value of each evaluation index. The basic steps of the variation coefficient method are: with n were evaluated, and the evaluated objects are described by p indexes;

to calculate the mean and variance of each index, and then the coefficient of variation of each index is:

$$v_i = S_i / \overline{x}_i \tag{2}$$

To make v_i be normalized, and we can get the weight of each index:

$$v_i = \frac{v_i}{\sum_{j=1}^{p} v_i}, i = 1, 2, \cdots, p \tag{3}$$

C. Combination Weighting Method

This paper adopts an optimal combination weighting method based on deviation square. Suppose n indexes with L weighting methods on the assignment, consider the following combination weighting:

$$W_c = \theta_1 W_1 + \theta_2 W_2 + \cdots + \theta_l W_l \tag{4}$$

And then,

$$\theta_k \geq 0, k = 1, 2, \cdots, l, \sum_{k=1}^{l} \theta_k^2 = 1 \tag{5}$$

Make partitioned matrix

$$W = (W_1, W_2, \cdots, W_l), \Theta = (\theta_1, \theta_2, \cdots, \theta_l)^T \tag{6}$$

And formula (4) can be expressed as:

$$W_c = W\Theta, \Theta^T\Theta = 1 \tag{7}$$

If the weight of each index of the improper selection, and the value of each difference is very small, it means that this is not conducive to the ranking of the alternatives. So a basic idea of choose a combination weighting coefficient vector W_c is to make each comprehensive evaluation value scatter as far as possible.

The $v_i(W_c)$ represents the deviation square of decision scheme i and other decision schemes. The target function can be constructed as follows:

$$\begin{aligned} J(W_c) &= \sum_{i=1}^{m} v_i(W_c) = \sum_{i=1}^{m}\sum_{i_1=1}^{m}\left[\sum_{j=1}^{n}(b_{ij} - b_{i_1 j})w_{cj}\right]^2 \\ &= \sum_{i=1}^{m}\sum_{i_1=1}^{m}\left[\sum_{j_1=1}^{n}\sum_{j_2=1}^{n}(b_{ij_1} - b_{i_1 j_1})w_{cj_1}(b_{ij_2} - b_{i_1 j_2})w_{cj_2}\right] \\ &= \sum_{j_1=1}^{n}\sum_{j_2=1}^{n}\left[\sum_{i=1}^{m}\sum_{i_1=1}^{m}(b_{ij_1} - b_{i_1 j_1})(b_{ij_2} - b_{i_1 j_1})\right]w_{cj_1}w_{cj_2} \end{aligned} \tag{8}$$

And make matrix B_1 :

$$B_1 = \begin{bmatrix} \sum_{i=1}^{m}\sum_{i_1=1}^{m}(b_{i1} - b_{i_1 1})(b_{i1} - b_{i_1 1}) & \sum_{i=1}^{m}\sum_{i_1=1}^{m}(b_{i1} - b_{i_1 1})(b_{i2} - b_{i_1 2}) & \cdots & \sum_{i=1}^{m}\sum_{i_1=1}^{m}(b_{i1} - b_{i_1 1})(b_{in} - b_{i_1 n}) \\ \sum_{i=1}^{m}\sum_{i_1=1}^{m}(b_{i2} - b_{i_1 2})(b_{i1} - b_{i_1 1}) & \sum_{i=1}^{m}\sum_{i_1=1}^{m}(b_{i2} - b_{i_1 2})(b_{i2} - b_{i_1 2}) & \cdots & \sum_{i=1}^{m}\sum_{i_1=1}^{m}(b_{i2} - b_{i_1 2})(b_{in} - b_{i_1 n}) \\ \sum_{i=1}^{m}\sum_{i_1=1}^{m}(b_{in} - b_{i_1 n})(b_{i1} - b_{i_1 1}) & \sum_{i=1}^{m}\sum_{i_1=1}^{m}(b_{in} - b_{i_1 n})(b_{i2} - b_{i_1 2}) & \cdots & \sum_{i=1}^{m}\sum_{i_1=1}^{m}(b_{in} - b_{i_1 n})(b_{in} - b_{i_1 n}) \end{bmatrix} \tag{9}$$

And then the objective function $J(W_c)$ can be expressed as:

$$J(W_c) = W_c^T B_1 W_c \tag{10}$$

Then based on the M decision scheme for total deviation square and the optimal combination weighting method is the following optimization problems:

$$\max F(\Theta) = \Theta^T W^T B_1 W \Theta \quad s.t. \begin{cases} \Theta^T\Theta = 1 \\ \Theta \geq 0 \end{cases} \tag{11}$$

The optimization problem can be reduced to below the unconstrained optimization problem:

$$\max F_1(\Theta) = \Theta^T W^T B_1 W \Theta / \Theta^T\Theta \tag{12}$$

$\lambda_{\max}$ is set to the maximum eigenvalue of $W^T B_1 W$, Θ^* is the normalized eigenvectors corresponding to the largest eigenvalue, and $F_1(\Theta)$ is the maximum value of $\lambda_{\max}$, and Θ^* is the optimal solution of formula (4). Put Θ^* into formula (4) to obtain the optimal combination weight coefficient vector $W_c^* = W\Theta^*$; normalize W_c^*:

$$w_{cj}^{**} = \frac{w_{cj}^*}{\sum_{j=1}^{n} w_{cj}^*}, j = 1, 2, \cdots, n \tag{13}$$

V. CONCLUSION

The extensive use of data technology and the rapid development of Internet cause the boundaries between different media no longer distinct. In the interaction and integration of television, the Internet and telecommunications networks, the old program evaluation system needs to be improved. This paper presents the completed new media oriented program evaluation architecture; the framework contains the user layer, application layer, technology layer and data layer and is able to provide a wide range of inquiry service including evaluation index, program classification and program evaluation system for each department on broadcast and television industry chain. Program evaluation index system is the core of the comprehensive program evaluation system; the evaluation index system concludes not only the traditional broadcasting and television viewership indicators, but also a comprehensive consideration of new media

platforms such as the Internet and micro blogging, striving to evaluate TV shows more comprehensive, objective and scientific.

REFERENCES

[1] Junjie Ding, Shuting Zhang, Weiningzhang. Preliminary exploration of the TV program impact assessment system under the background of triple play [J]. Modern Communication (Journal of Communication University of China), 2010, 11:99-102.

[2] Hai Yang. Research on comprehensive evaluation system of TV program [J]. News World, 2013, 08:68-70.

[3] Yannan Liu. Longitudinal and horizontal comparison of CCTV's - characteristics, differences and discuss new evaluation system [J]. South China Television Journal, 2011, 04:10-13+2.

[4] Li Zhang. The establishment and development of the City TV program evaluation system [J]. Voice & Screen World, 2013, 12:56-57.

[5] Xiaohu Shang. Exploration and Consideration of Tianjin TV program evaluation system [J]. China Radio & TV Academic Journal, 2013, 11:97-98.

[6] Junchang Zhang, Peng Lu. Radio and television program evaluation system: background, current situation and development trend [J]. China Radio & TV Academic Journal, 2011, 11:11-13.

[7] Limin Gao. Establishment of the television program evaluation system: Practice and Thinking of Shandong TV [J]. TV Research, 2003, 01:55-57.

[8] Weihua Wen. TV program evaluation system: the Anglo-American model and the Chinese practice [J]. China Television, 2011, 11:82-85.

[9] Yannan Liu. Television assessment: public television vs commercial television Britain and the United States and Taiwan's experience and thinking [J]. Journal of China University of Geosciences (Social Sciences Edition), 2011, 02:75-80.

[10] Wen Lei. Discuss the significance of European and American television program evaluation [J]. Journal of Nanchang Junior College, 2006, 03:82-84+87.

[11] Kai Yang. American TV program evaluation system and enlightenment [J]. China Radio & TV Academic Journal, 2005, 02:71-72.

[12] Yan Ni, Shuguang Zhao. Western public television program evaluation: Paradox of ratings [J]. Journal of International Communication, 2004, 02:65-68.

[13] Chun Liu. Animation brand value and management system research based on the evolutionary economics [D].Zhongnan University, 2012.

[14] Yuanxia Wang. Cultural interpretation of TV dating show [D].Anhui University, 2011.

[15] Taewan Kim; Jiwoo Kang; Sanghoon Lee; Bovik, AC. Multimodal Interactive Continuous Scoring of Subjective 3D Video Quality of Experience[C]. Multimedia, IEEE Transactions on (Volume: 16, Issue: 2), 2014:387 - 402.

[16] Yilei Zheng. Audience rating prediction of new TV programs based on GM (1.1) envelopment model[C]. Grey Systems and Intelligent Services, 2009. GSIS 2009, 2009:388 - 391.

A Mahout Based Image Classification Framework for Very Large Dataset

Jun He, Zhi-Yun Xue, Ming-Wei Gao
School of Electronic and Information Engineering
Nanjing University of Information Science and Technology
219 Ningliu Road, Nanjing 210044, China
jhe@nuist.edu.cn, nuistxzy@163.com

Hao Wu
School of Information Science and Engineering
Yunnan University
No. 2 North Green Lake Road, Kunming 650091, China
haowu@ynu.edu.cn

***Abstract*—In this paper, we present a distributed computing framework for image classification towards the current challenge of image *big data* due to enormous streaming image data sources, such as image sharing over online social network and massive video surveillance streams from ubiquitous cameras all over our daily life. The proposed framework consists of four modules aiming at feature extraction, dimension reduction, bag of feature modeling, and supervised learning respectively. This distributed computing framework is implemented on Hadoop with Mahout support. We apply the framework for classifying whether a person is on calling or not in a surveillance video to verify the correctness and scalability.**

Keywords—Map-Reduce; Big Data; Bag of Feature; Image classification

I. INTRODUCTION

There is no need to say how important image classification [1-4] is in computer vision, which is essential for bridging the huge semantic gap between images. Traditional image classification task can be mainly divided into several steps as follows: feature extraction, bag-of-feature model and supervised learning.

OpenCV is an open source library for image and video analysis. It has a wide range of modules that can help us with a lot of computer vision problems. The most useful part is its architecture and memory management. It provides us with a framework in which we can work with images and video conveniently.

With the development of Internet technology, the amount of dataset is streaming, which brings about the challenge of big data. Map-Reduce [5-7] programming model was proposed in 2004 by Google, which is used in processing and generating large data sets. Inspired by the success of Map-Reduce for handling big data, Hadoop [8-11] and Mahout [12-14] are representative open source distributed computing framework that enables data-intensive application in a distributed environment. They support large-scale data applications on large clusters consisting of commodity hardware. The framework is designed to be scalable, which allows the user to add more nodes as the application requires.

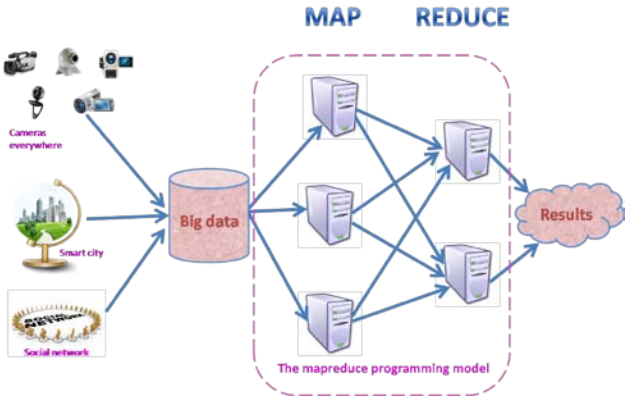


Fig. 1. The Map-Reduce programming model

As the popularity of online social network, image sharing over the Internet, video surveillance and smart city, computer vision has been facing a problem of big data, it is very necessary to design a framework that suits for processing very large image datasets, which combines image classification with big data platform.

The rest of this paper is organized in the following way: section 2 presents the distributed computing framework. Section 3 validates the framework with an application for classifying whether a person is on calling or not. Finally, section 4 concludes the framework and points out our future work.

II. THE DISTRIBUTED COMPUTING FRAMEWORK

We designed a distributed computing framework that can perform image classification for very large dataset. In the training phase, it consists of feature extraction, bag-of-feature model [15] and supervised learning (training). In the prediction phase, put the image into the classification model to predict the class that the image belongs to. The prediction step will be described in detail in section 2.

At the beginning, we optionally perform background subtraction to get the objects, then we choose SIFT algorithm [16,17] to extract feature points for each object, saving these feature points as text file format which can be read by Mahout. Next, construct bag of feature model to get the histogram of

each image that represents the class proportion of the image. Finally, training the classification model with classifier, the classification model can be used for prediction.

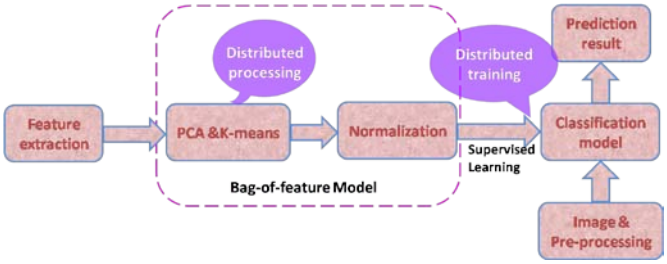


Fig. 2. The general framework structure

Algorithm 1. Distributed image classification framework.

Training phase: Input:**pic**=[$pic_1, pic_2, ..., pic_m$],label. Output: model.

Prediction phase: Input: pic. Output: class.

1: Define two matrices $A_{m\times n}=[f_{ij}]$ and $A'_{m\times n}=[f'_{ij}]$. $A_{m\times n}$ stores all the feature points of pic, f_{ij} represents a feature vector. $A'_{m\times n}$ stores dimension reduced features points, f'_{ij} represents dimension reduced the feature vector. And m is the number of images, n is the number of each features.
Define two vectors $\mathbf{C}=[\mathbf{c_1},\mathbf{c_2},...\mathbf{c_q}]$ and $\mathbf{N}=[\mathrm{n}_1,\mathrm{n}_2,...,\mathrm{n}_q]$, C is the center of each cluster, N is percentage of the image in each cluster.

2: **Training phase:**

2: Sift feature extraction:

3: Dimension reduction: $A'_{m\times n}=pca(A_{m\times n},p)$

4: K-means clustering: $\mathbf{C}$ =kmeans($[A_1,A_2,...,A_m]^T$)

5: **while** $i\le m$ **do**

6: Normalization: $n_i=norm(C,A_i)$, $\sum_{i=1}^{q} n_i=1$

7: **end while**

8: Model training: model = *train_logistic*($[n_1^T,n_2^T,...,n_m^T,label]$)

9: **Prediction phase:**

10: Sift feature extraction: $a=sift(pic)$

11: Project onto subspace of $A'_{m\times n}$: $a'=\mathrm{P}_{A'_{m\times n}}(a)$

12: Prediction: class = *model_predict*(a')

A. *Details of Each Module*

1) Feature extraction with SIFT: The original data set is a video, we first use OpenCV to convert video into a sequence of images and subtract background from these images in order to increase the accuracy, which will be the data set. Then, use SIFT algorithm to extract feature points from the image set, these points are saved as a file named Feature. The procedure was processed on a single machine since it does not have high time complexity.

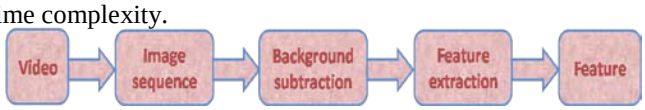


Fig. 3. Feature extraction

2) Bag-of-feature: The bag-of-feature model consists of PCA [13,14] dimension reduction, K-means clustering [13,14] and classification pre-processing that includes normalization and label adding.

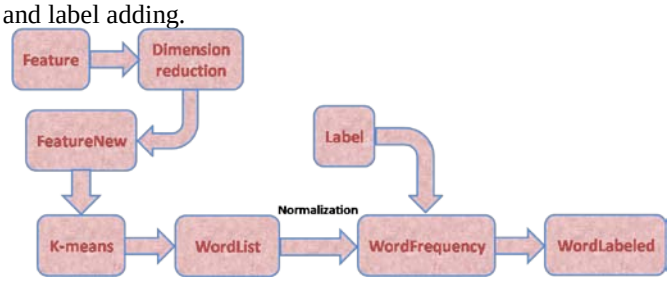


Fig. 4. Bag-of-feature structure

As the feature extracted by the SIFT algorithm has the characteristic of high dimension, in order to accelerate the processing speed and increase the prediction accuracy, we choose to apply PCA to reduce dimension. PCA is a classical tool for dimension reduction, Mahout currently support PCA via Stochastic Singular Value Decomposition, on which can do PCA distributive.

After PCA, we got FeatureNew which will be the input of K-means stage. Mahout contains various implementation of clusters, such as K-means, fuzzy K-means, meanshift and Dirichlet among others. K-means clustering is a widely used partition algorithm. It partitions the samples into clusters by minimizing a measure between the samples and the center of the clusters.

The K-means clustering is simple but it has high time complexity when having the large dataset. Firstly, the algorithm randomly selects k initial objects which represents a cluster center. The rest of the objects will be assigned to the nearest cluster, according to their distances to different centers. Then calculate every center again. This operation is repeated until the criterion function converges. In these circumstances the memory of a single machine can be a restriction. As a consequence, we shall try to use Map-Reduce computing framework to process data distributive. In this paper, using Mahout K-means to get the input FeatureNew clustered with the output Wordlist that stores each center of clusters.

Since the file Wordlist cannot be processed directly by the classifier, it is very necessary to do some pre-processing.

Above all , calculating the Euclidean distance between one feature vector to each clustered center point and then feature vector belongs to the cluster in which the cluster's center point having the smallest Euclidean distance to the feature. Repeat doing calculation until the last feature, count the number of feature vector in each cluster. Finally, doing normalization to get WordFrequency and adding label to WordFrequency, then we get WordLabeled which will be the input of classifier. In this article, the Pre-processing procedure do not high time complexity, a single machine can meet the demand of resource consumption.

3) Training classification model: The classification model was trained for predicting the image class. For the training phase, input for the training algorithm consists of dataset labeled with known target variables. The target variables are, of course, unknown to the model when using new images during real prediction. In prediction, the target variable values are not known, which is why the model is built in the first place.

Fig. 5. Classification model

Classification algorithms differs from the clustering algorithms described previously because the clustering algorithms are able to decide on their own what distinctions appear to be important while classification algorithms learn to mimic examples of correct decisions.

Mahout is a machine learning library that contains different kinds of classifiers [13,14] like Naive Bayes, Logistic regression and Random Forest. Since Mahout currently do not support SVM, we choose to use Logistic regression (SGD) instead for classification.

Image Prediction: The prediction procedure is almost the same as the training procedure. When using new image during the prediction, first of all, subtract the background of this image and use SIFT algorithm to extract feature points. Then, reduce the dimension of these feature points by projecting onto the low dimensional subspace, the reduced feature having the principle component that can almost represent all the original feature points.

In prediction, the target variable values are not known, which is the reason that training the classify model in the training phase. Put the dimension reduced feature points into the classify model which will be used to calculate the column diagram, the one in the column diagram has the highest percentage will be the class this image belongs to.

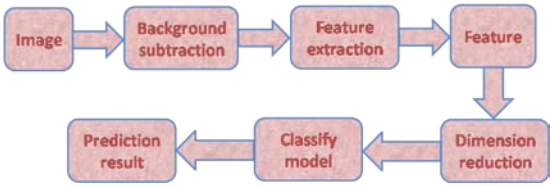


Fig. 6. ImagePrediction procedure

B. *Performance analysis*

We proposed a distributed computing framework for image classification, which enables large-scale image data set to be classified properly. The framework consists of feature extraction, PCA dimension reduction, K-means clustering, Classification Pre-processing, Logistic regression classifier and Image prediction. Among them, we use Mahout to process PCA dimension reduction, K-means clustering, Logistic regression classifier and Image prediction. The others were processed on a single machine.

Mahout provides a machine learning library that runs over Hadoop system. It has a collection of algorithms to solve clustering, classification and prediction problems, which can be used as an inexpensive solution to solve machine learning problems especially the problems with large dataset.

Some parts of this framework such as PCA, K-means and Logistic regression are having a high time complexity, when handling large datasets, single machine cannot meet the demand of resources.

III. FRAMEWORK VALIDATION

A. *The Distributed Platform Parameter*

We've built Hadoop distributed system on Ubuntu with the following characteristics: 2GB memory; 2 Ubuntu compute units (one is master and the other is slave); 20GB of storage; 32-bit platform; high I/O performance. Install mahout on the master that we can use the mahout machine learning library on Hadoop system. Our Ubuntu uses Mahout 0.9 and Hadoop 2.2.0.

B. *The Data Set*

The data set was a sequence of ten thousand images which converted from a video frame, eighty percent of data set were used for training and twenty percent were used for testing. We did the pre-processing that subtracts the background of each image to increase the classification accuracy.

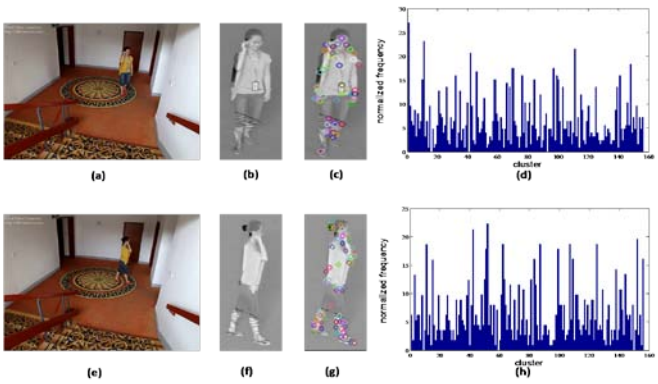


Fig. 7. On calling

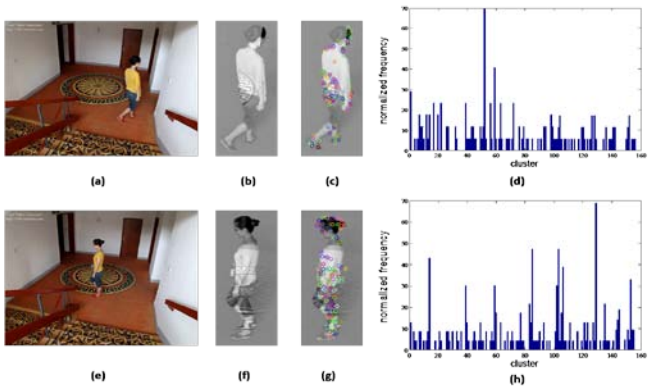


Fig. 8. Not on calling

C. *Image Classification Results*

We use this distributed computing framework to classify the image dataset into 2 categories, one is on calling and the other is not. Above all, using sift algorithm to extract feature from dataset. Then construct bag-of-word model that includes PCA, K-means and normalization to get the WordFrequency. Finally, training to get the classification model input the testing dataset in the model for testing. Figure 9 shows the classification accuracy.

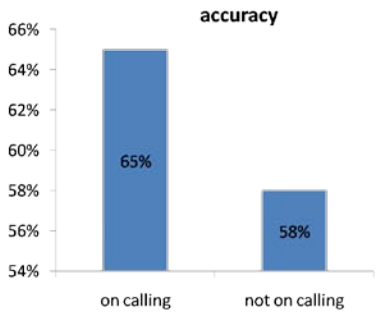


Fig. 9. Image classification accuracy

IV. CONCLUSION AND FUTURE WORK

In this paper, we proposed a distributed computing framework for image classification, compared with the project running on a single machine; it can deal with large-scale data set problems. In the framework, Mahout is used for computing the part which has high time complexity. We conclude that Mahout can be a promising tool for distributed computing. It allows the computing of large datasets, and produces significant gains in performance.

We've run our program using Ubuntu with two nodes on the same machine, strictly speaking, it is not distributed computing meaningfully. In the future, we plan to build platform on Amazon EC2 to meet the challenge of more and more large datasets.

ACKNOWLEDGEMENTS

This work is supported by NSFC (61203273).

REFERENCES

[1] Dollár, Piotr, et al. "Feature mining for image classification." Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on. IEEE, 2007.

[2] Haralick, Robert M., KarthikeyanShanmugam, and Its' HakDinstein. "Textural features for image classification." Systems, Man and Cybernetics, IEEE Transactions on 6 (1973): 610-621.

[3] Höppner, Frank, ed. Fuzzy cluster analysis: methods for classification, data analysis and image recognition. John Wiley & Sons, 1999.

[4] Lin, Yuanqing, et al. "Large-scale image classification: fast feature extraction and svm training." Computer Vision and Pattern Recognition (CVPR), 2011 IEEE

[5] Chu, Cheng, et al. "Map-reduce for machine learning on multicore."Advances in neural information processing systems 19 (2007): 281.

[6] Dean, Jeffrey, and Sanjay Ghemawat. "MapReduce: simplified data processing on large clusters." Communications of the ACM 51.1 (2008): 107-113.

[7] Dean, Jeffrey, and Sanjay Ghemawat. "MapReduce: a flexible data processing tool." Communications of the ACM 53.1 (2010): 72-77.

[8] Apache Hadoop, http://hadoop.apache.org/.

[9] Borthakur, Dhruba. "The hadoop distributed file system: Architecture and design." Hadoop Project Website 11 (2007): 21.

[10] Lam, Chuck. Hadoop in action. Manning Publications Co., 2010.

[11] Shvachko, Konstantin, et al. "The hadoop distributed file system." Mass Storage Systems and Technologies (MSST), 2010 IEEE 26th Symposium on.IEEE, 2010.

[12] Apache Mahout, http://hadoop.apache.org/.

[13] Anil, Robin, Ted Dunning, and Ellen Friedman. Mahout in action.Manning, 2011.

[14] Ingersoll, Grant. "Introducing Apache Mahout." Scalable, commercial-friendly machine learning for building intelligent applications. IBM (2009).

[15] Nowak, Eric, FrédéricJurie, and Bill Triggs. "Sampling strategies for bag-of-features image classification." Computer Vision–ECCV 2006. Springer Berlin Heidelberg, 2006. 490-503.

[16] Ke, Yan, and Rahul Sukthankar. "PCA-SIFT: A more distinctive representation for local image descriptors." Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on. Vol. 2. IEEE, 2004.

[17] Lowe, David G. "Distinctive image features from scale-invariant keypoints."International journal of computer vision 60.2 (2004): 91-110.

An Improved Kademlia Algorithm Based on Qos

Lin Zhu
Computer Center
East China Normal University
Shanghai, China
51121211016@ecnu.cn

Kai Zheng
Computer Center
East China Normal University
Shanghai, China
kzheng@cs.ecnu.edu.cn

***Abstract*—In distributed routing algorithms, Kademlia is the most widely used. Its slow routing table updating, load imbalance and weaknesses in network adapter has become increasingly evident. This paper introduces a kind of Kademlia Algorithm based on Qos. Through simulation experiments, proved that this algorithm is more efficient than the traditional algorithm in appropriate circumstance.**

Keywords—Kademlia; routing table; Qos; DHT

I. INTRODUCTION

In 2002, Petar Maymounkov, from New York University, United States, published an article entitled Kademlia: A Peer-to-Peer Information System Based on the XOR Metric(1). This paper raised awareness of Kademlia. Kademlia is a P2P protocol based on DHT, it is different from other protocols based on DHT, such as Chord (2), Kelips(3), CAN(4) , Pastry(5), because it adopted a unique XOR algorithms to measure the distance between the node, and by this way a new network topology is created. so Kademlia greatly improved the routing efficiency of the node. So far, many P2P systems have been developed based on Kademlia, such as BiTorrent, BitComet, BitSpirit, and eMule.

But, with the continuous growth of the application in P2P systems. The inherent flaws of Kademlia have become more and more evident, such as not supporting fuzzy query(6), routing detours(7) and load imbalance(8). This article presented a storage algorithm especially against load imbalance of Kademlia.

II. LOAD IMBLANCE

When Kademlia was proposed , it assumed that all node had the same processing capacity and space in the overlay network. But in fact, resources are randomly distributed in the overlay network, and the nodes are highly heterogeneous, above all the nodes have different processing capacity and space. Due to the configuration of the node, the calculation capacity is huge difference, this will lead to load imbalance, bringing the performance bottlenecks to the overlay network. The same number of packets, the same request, the high performance nodes can easily cope with, but for the normal or low performance nodes, it will cause network congestion. Once some resources become hot, it means that many users want to download the resource, load imbalance will become more and more evident. This situation has bad influence on the network's Qos. The usual practice is to introduce a concept called super node to distinguish the node in the overlay network, the concept makes those nodes which have strong calculation capacity and big storage capacity become super node in the overlay network, act as a regional server, and the low performance node act as the leaves node. But such a strategy requires a complex algorithm to process the node's launch and expiration, adding additional system overhead. Some paper presented a kind of KAD index information release mechanism based on multiple target ID by having a large user base of eMule for study. This mechanism makes more nodes have high frequency key words index to improve the KAD network file index resource's load balance. In this article, the author proposed a storage arithmetic based on Qos, it means that every time the resource is released, we choose three powerful nodes to act as backup nodes and do the redundant storage, this can improve the network's Qos to some extent, it can avoid sending large requests simultaneously to some node when the node becomes hot.

III. THE STORAGE ALGORITHM

A. Storage Algorithm

Firstly, we calculate the file eigenvalue based on file name or summary by hash function, the file eigenvalue is FileID and the hash function is SHA-1:

$$FileID = SHA\text{-}1(file) \quad (1)$$

Then, because every node has its K-bucket, as shown in Fig.1, we choose the most three powerful nodes by FIND_NODE command, meanwhile the target node's eigenvalue must match the node's top m bits.

Because the number of nodes in different overlay network are different, we must set up mathematical model for m, that is

$$M = f * (160 - logN) \quad (2)$$

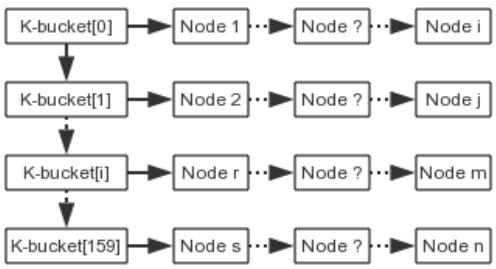


Fig. 1. The K-bucket

Suppose the source node need to store a file, the steps are as follows:

- Check the length of the file, if the length exceeds the limit, the file must be divided into some blocks, every block length does not exceed L and each block is stored separately, case of the file whose length does not exceed L;
- The source node calculates the eigenvalue of file and set up the parameter m, so we can match the node's eigenvalue according to m;
- The source node sends FIND_VALUE request to related K-buckets, the parameters including file eigenvalue, m and the length of file.
- The nodes which receive the request firstly examine whether this node's eigenvalue matches the source node eigenvalue's top m bits, if it is, then examine whether the space this node's space meets the need(space > filesize + c, c is constant). Whether the condition is satisfied or not , the query will be continue in related K buckets.
- If the node can meet the requirements, so this node will send message to source node, the parameters including IP address and the source IP address.
- When the source nodes receive the reply, they will choose the top Q(usually 3) nodes which have high performance as storage target, and send acknowledge information to these nodes. The intermediate nodes which the information passes will store the route, then the target node records the eigenvalue of the file.

B. Search Algorithm

Suppose the source node need to find a file, the steps are as follows:

The source node calculates the eigenvalue and the position in K-buckets according to its eigenvalue;

- The source node sends FIND_VALUE request to related K-buckets, the parameters including file eigenvalue, m and the length of file.
- The nodes which receive the request firstly examine whether the eigenvalue match the value in K-buckets, if not, forward the message to related K-buckets.
- If the target nodes meet the requirements, then this node will send acknowledge information to source node.

C. The K-bucket based on node performance

In terms of load balance, taking highly heterogeneous nodes , the stability of the overlay network into account, we can set up a fixed value fp in the system. We only consider the nodes whose comprehensive performance are greater than fp when the top Q nodes are choosed. Not only the IP address, UDP port and some relevant information are stored in the K-buckets, but the parameter p which means the comprehensive performance is also included, the parameter includes bandwidth represented by B ,computing power represented by C and storage capacity represented by S.

The mathematical model is as follows:

$$p = f(B, C, S) \qquad (3)$$

the higher value of p means the higher performance of the node.

IV. SIMULATION TEST

The Algorithm adopts peersim as simulation platform, Peersim is a network simulator .In the simulation environment, the number of nodes is 500,the file save success rate and overload nodes are the assessment factor between the old and new kademlia algorithm.

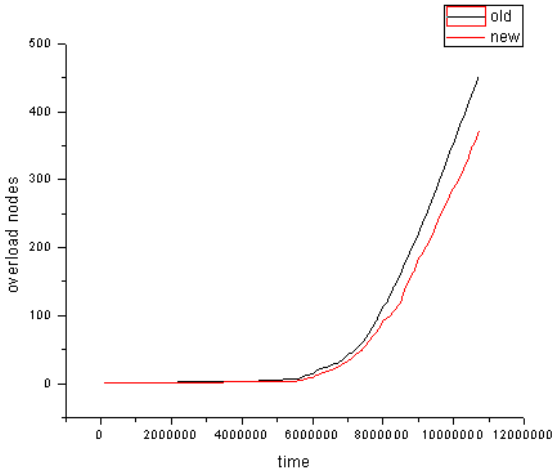


Fig. 2. The number of overload nodes between old and new algorithm

Fig.2 shows the number of overload nodes increases with the time, the random nodes have many information exchange. The improved algorithm reduces the number of overload nodes through resource backup.

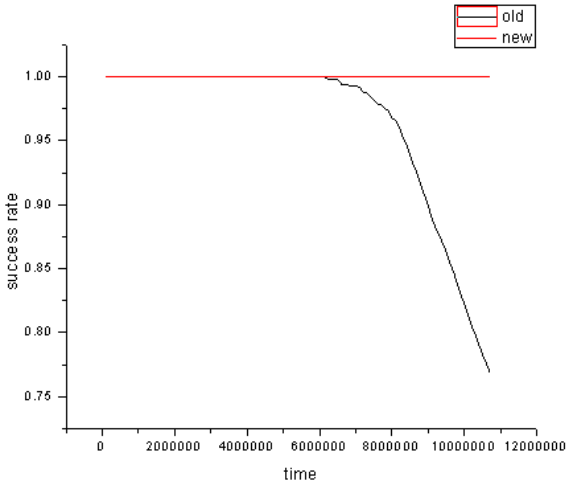


Fig. 3. The file save success rate between old and new algorithm

Fig.3 shows the file save success rate changes with time. In the old algorithm, there is only a resource in the network, when the other nodes want to request it, the rate may be decrease with time, but in the new algorithm there are three backups in

the network, so when other nodes request the resource the rate has little change.

V. CONCLUSION

The simulation test confirms that the improved algorithm has more effective than the old algorithm.

ACKNOWLEDGMENT

This work was supported by National High-tech R&D Program of China (2013AA01A211).

REFERENCES

[1] Maymounkov, P.&D. Mazieres. Kademlia: a peer-to-peer information system based on the XOR metric. Proceedings of the International Workshop on Peer-to-Peer Systems, Cambridge, USA, 2002:53-65.

[2] Stoica, I.&R. Morris.&D. Liben-Nowell.&D.Karger.&M. Frans Kaashoek.&F. Dabek.&H. Balakrishnan. "Chord: A scalable peer-to-peer lookup service for Internet applications," IEEE Trans. Vol. 11, pp.17 2003.

[3] Gupta, I.&K. Birman.&P. Linga, AI Demers&R. van Renesse. "Kelips: Building an efficient and stable p2p DHT through increased memory and background overhead," in Proceedings of the 2nd International Workshop on Peer-to-Peer Systems, IPTPS '03, Berkeley, CA, USA, February 2003.

[4] Ratnasamy, S.&P. Francis.&M. Handley.&R. Karp and Scott Shenker. "A scalable content-addressable network," In Proceedings of the 2001 conference on Applications, technologies, architectures, and protocols for computer communications (SIGCOMM '01). ACM, NEW YORK, NY, USA, 161-172.

[5] Rowstron, A.&P. Druschel. "Pastry: Scalable, distributed object location and routing for large-scale peer-to-peer systems," Proc. 18th IFIP/ACM Int. Conf. Distributed Systems Platforms (Middleware) pp.329-350 2001.

[6] Lan, M.J. Structured P2P algorithm with fuzzy query[J]. Journal of Chongqing University of Posts and Telecommunications, 2013,(10): 680-685.

[7] Xu, Q&S. Lechang.&Z. Haomiao. Improvement of Kademlia Based on Self-organized Clustering[J]. Journal of Chinese Computer Systems, 2010, (8): 1549-1553.

[8] Shi, J.T&Z. Hongli. Study on Load Balancing of KAD Network[J]. Telecommunications Science, 2012, (6): 68-72.

Congestion-aware Data Acquisition for Internet of Things*

Yue Pan, Yue Li, and Junxing Zhang**
School of Computer Science
Inner Mongolia University, Huhhot, China
**Corresponding author (E-mail: junxing@imu.edu.cn)

***Abstract*—In this paper, a data acquisition scheme that is reactive to network congestion is proposed for Internet of Things (IoT). It adjusts data acquisition rates according to the congestion level in a distributed manner, thereby alleviating and avoiding congestion as much as possible. It also makes its best effort to deliver information of high priorities in a timely manner and ensure uninterrupted sensing. Our experiments demonstrate that the scheme offers a much higher delivery rate of the high-priority data compared with the original scheme when the network undergoes different levels of congestion.**

Keywords—Internet of things; data acquisition; congestion control; TCP.

I. INTRODUCTION

Internet of Things (IoT) is a global network infrastructure links physical and virtual objects through the exploitation of data capture and communication capabilities [1]. Many believe it could transform our daily life, so since its inception IoT has attracted attention from sundry communities all over the world.

IoT has evolved from wireless sensor networks, and thus it inherently can be deployed widely and flexibly. As a result, it often experiences changes in network resources either because of alterations in deployment or due to accidents in monitored areas. On the other hand, unlike wireless sensor networks, IoT is designed to be an integrated part of Future Internet [2], so things in IoT have to share or even compete for network resources with other nodes during transmission. Both resource alterations and competitions can lead to variations in the available network resources along transmission paths, which can easily trigger network congestion. Given that many IoT applications require continuous sensing or monitoring, network congestion will affect their performance or prevent them from working properly. In addition, large amounts of data acquired during congestion often become useless when they cannot be delivered in time. Consequently, the energy spent on acquiring these data is also wasted, which is unacceptable under many energy-constrained circumstances. These situations call for an intelligent data acquisition scheme that can deal with network congestion.

The requirement for the intelligent data acquisition scheme is not only raised by the design of IoT but also driven by its continuous development. First, with the improvement of the sensor technology, sensors become more and more powerful and their sensed information get increasingly diverse. Secondly, the duplicate deployment of IoT in the same area for different purposes tends to happen owing to its growing popularity. To avoid the waste of energy and resources, a better deployment alternative is to reuse the same network infrastructure and replace regular sensors with special ones that are capable of sensing multiple kinds of information. Thus, with the growth of IoT, the subnets consist of those special sensors are likely to prevail. In both above situations, the enhancement of sensors in IoT can rapidly boost the variety and amount of data in the network. To alleviate the burden on IoT, we need an intelligent data acquisition scheme that can constrain the account of various acquired data within the transmission capability of IoT.

To meet these requirements, we propose a data acquisition scheme that leverages congestion information to guide data acquisition in IoT. Because congestion detection has been well studied in congestion control mechanisms [3], [7], our scheme makes use of these existing mechanisms to extract congestion information. Specifically, we have designed an algorithm to observe the real-time running state of the TCP protocol. As a reliable transport protocol, TCP has to be sensitive to network congestion. Whenever TCP detects a change in network conditions, it immediately transitions into the corresponding state, such as timeout, retransmission, fast retransmission, and fast recovery. Our scheme infers there is serious, mild, or no congestion according to these running states and then utilizes the information to adjust data acquisition accordingly.

Our data acquisition scheme works with the IoT subnets that are composed of composite sensors. Composite Sensors (or CS for short) are those special sensors that are capable of sensing multiple kinds of information. Some kinds of sensed information tend to generate high-rate data flows during the continuous perception, while others may introduce flows of moderate or low rates. This type of IoT configuration offers more accurate and detailed surveillance than the one employs only simple sensors. IoT applications may not need all kinds of information to work correctly; rather, they offer best services with all the gathered information and provide services with lower qualities when some kinds of information become unavailable. Our scheme prioritizes different kinds of sensed data based on their importance to applications and their consumption of network resources. More importantly, the advanced CS node is equipped with sufficient computational power to run TCP.

Our scheme also includes a Phased Exponential Backoff Sleep (PEBS) algorithm that runs on every composite sensor in

the network. When the sensors detect congestion, they will stop acquiring information that has lower priorities under the discovered congestion level. Accordingly when the network condition clears up, they will resume the acquisition. The algorithm ensures the sensing of lower-priority information will always be turned off earlier and resumed later than the sensing of higher-priority information, as it handlers classes of different priorities in phases. In consequence, applications will have a good chance to retain timely and uninterrupted delivery of information that is critical to them even under congestion. Moreover, given that PEBS is a distributed algorithm, it could cause the sensors to react synchronously to the same congestion. The synchronized actions may further lead to repeated traffic oscillation [4]. To avoid the problem, PEBS has incorporated the exponential backoff [5] mechanism into it. Relying on the uncertainty of the randomly selected backoff time, the sensors will respond after different lags. The lags also help to make sure the congestion is not too short to respond. In view of many sensors work in the energy- constrained environment, when the hardware permits, PEBS actually powers off or on the circuits that sensing the lower-priority information instead of just pausing or restarting the sensing when the time comes.

Most congestion control schemes for IoT or WSN [8, 9] deal with congestion at the time of data transmission, while our scheme adapts to congestion during data acquisition. Because we take actions at an earlier stage, we can greatly reduce the amount of data dropped actively by the congestion control mechanism or passively by the data forwarding mechanism, and thus save the energy spent on acquiring them.

II. System Design

We have designed a system to perform adaptive data acquisition on composite sensors in IoT. The system can adjust what information to acquire and how fast to acquire based on congestion levels of the network. As Fig. 1 shows, the system architecture includes three layers in accordance with the middleware architecture of IoT [6]. We will introduce details of each layer in the following subsections.

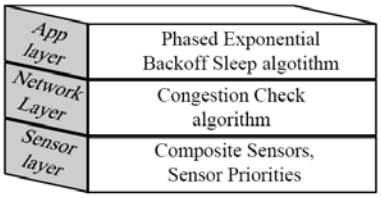


Fig. 1. System Architecture

A. Sensor layer

Fig. 1 reveals the sensor layer in our system mainly includes composite sensors and priorities of the various sensed information. Composite sensors can acquire a variety of information. An example of such sensors is given in Fig. 2. In the example, the composite sensor comprises four simple sensors and one embedded gateway. The simple sensors acquire various data: the video camera captures moving and still pictures, the glass break detector monitors the window, the magnetometer sensor scouts out the door, and the infrared beam detector looks over any place it might be installed. The gateway controls sensors' data acquisition and runs TCP to transmit all the captured data on their behalves. Please note the gateway connects to the simple sensors wirelessly, while it connects IoT in a wired manner. This design avoids the problem of unsatisfactory TCP performance on wireless links.

In order to adaptively adjust data acquisition based on the network condition, we need to prioritize the information the sensors can acquire. The priorities are determined based on two factors: the importance of the sensed information to IoT applications, and their requirements for network resources to transmit. For instance, the surveillance application has the priority values 2, 3, 4, 5, and 6 assigned to the acquired HD video, regular video, still images, infrared beam readings, and magnetometer values correspondingly. Larger priority values represent higher priorities. Here the highest priority HP=6 and the lowest priority LP=2.

This layer also supports powering on and off individual components of composite sensors for adaptive sensing and energy conservation purposes.

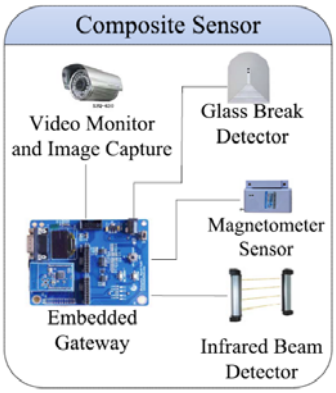


Fig. 2. An Example of the Composite Sensor (CS)

B. Network layer

The network layer of our system extracts the real-time running states of the TCP protocol, analyzes these states to infer how congested the IoT subnets are, and then inform the upper layer the congestion levels of the subnets. As disclosed in Fig. 1, we have designed Congestion Check (CC) algorithm in the network layer. The CC algorithm runs in the kernel space of the operating system to get access to the TCP protocol stack.

The pseudo code of the CC algorithm is given in Algorithm 1 below. It first calls a function named check_tcp_state() to get the current running state of the TCP protocol and saves the return value to the variable tcp_state. Then, according to the value of this variable it infers the latest congestion level of the network as perceived by TCP. If TCP is in the state of overtime retransmission, it will consider the network is heavily congested. If instead TCP is in the state of fast retransmission, it will think there is only mild congestion. When acknowledgements arrive within the expected time frame, they are counted as normal acknowledgements. Whenever a normal acknowledgement is received, the counter *na_count* is incremented. When the counter is larger than 3, the algorithm will reckon the network does not have congestion right now.

The counter is reset to zero each time after the algorithm has inferred a congestion state to prepare for its next use.

```
1  while true do
2      tcp_state=check_tcp_state();
3      if tcp_state==TIME_OUT then
4          con_state=CON_HEAVY;
5          na_count=0;
6      end
7      else if tcp_state==FAST_RETRANSMIT then
8          con_state=CON_LIGHT;
9          na_count=0;
10     end
11     else if tcp_state==NORMAL_ACK then
12         na_count=na_count+1;
13         if na_count> 3 then
14             con_state=NET_WELL;
15             na_count=0;
16         end
17     end
18 end
```

Algorithm 1: Congestion Check (CC) Algorithm

C. Aplication layer

In this layer, we have designed Phased Exponential Backoff Sleep (PEBS) algorithm as indicated in Fig. 1. PEBS makes its best effort to ensure the critical data will be delivered timely and uninterruptedly, since it stops sensing the low-priority data partially or completely when the network is congested. Based on the network congestion states and the information priorities, it gradually turns off and later turns on the sensing of different kinds of information. PEBS is also energy conservative because when it switches the data sensing off and on it actually hibernates and wakes up the simple sensors that perform the sensing in the composite sensor.

Further, PEBS also avoids the synchronized reaction of composite sensors by performing the exponential backoff. The previous studies on the congestion control of TCP [4] suggest that "users' control actions are highly synchronized by the network congestion signaling and that providing users with only a binary network state can lead to repeated traffic oscillation." Given that PEBS is a distributed algorithm and its adaptive data acquisition is also guided by the congestion signaling, it may also bring about traffic oscillation if it takes no measure. The existing measures taken by TCP cannot be adopted by PEBS. Because Tail Drop blindly discards the data near the end of the queue and RED drops the incoming data based on statistical probabilities, neither of them can avoid the loss of the critical data. Therefore, PEBS has borrowed the binary exponential backoff mechanism in the CSMA/CA protocol to solve this problem. In the end, PEBS has not only prevented the problem using the backoff mechanism but also ensured the phased reaction using the backoff time period.

The pseudo code of the PEBS algorithm is shown in the following Algorithm 2. It first calls the function *get_cc_state()* to obtain the current congestion state, *con_state*, inferred by the CC algorithm. The function also addresses the issues of process synchronization and kernel/user space communication. Then, PEBS determines whether or not to hibernate the simple sensors based on the value of con_state. There are three possible values: CON_HEAVY indicates the network congestion is severe, CON_LIGHT means the congestion is mild, and NET_WELL suggests there is no congestion discovered. In the first case, PEBS calls the function *hibernate_abc_sensors()* to power off all but the critical sensors in order to react promptly to the terrible network condition. In the second case, PEBS performs the phased exponential backoff. First, it exponentially backs off to avoid the synchronized reaction and the backoff time is *sleep_time = random (0, $2^{s\_priority}$)*. After the backoff, it calls *get_cc_state()* again to obtain the current congestion state. If the state is still CON_LIGHT, it powers off the simple sensor that has the priority value *s_priority*. Next, it continues to back off and the backoff time is $2^{s\_priority} - sleep\_time$. The second backoff makes sure higher-priority sensors will not take precedence over lower-priority sensors in hibernation. In the third case, because the state value NET_WELL is returned by the CC algorithm after the delay of receiving three normal acknowledgements, there is no need to back off further. PEBS directly calls the *wakeup_sensor()* function to wake up the simple sensor with the priority value *s_priority -1*.

```
1  while true do
2      con_state=get_cc_state();
3      switch con_state do
4          case CON_HEAVY
5              sleep_all_sensors();
6          endsw
7          case CON_LIGHT
8              sleep_time=random(0,2^s_priority);
9              sleep(sleep_time);
10             con_state=get_cc_state();
11             if con_state==CON_LIGHT
12             then
13                 sleep_sensor(s_priority);
14                 sleep(2^s_priority − sleep_time);
15                 s_priority=min((s_priority+1),HP);
16             end
17         endsw
18         otherwise /*NET_WELL*/
19             s_priority=max((s_priority-1),LP);
20             wake_up_sensor(s_priority);
21         endsw
22     endsw
23 end
```

Algorithm 2: Phased Exponential Backoff Sleep (PEBS) Algorithm

III. Performance Evaluation

In this section we evaluate our data acquisition system using OPNET, one of the most popular software adopted by both academia and industry for network modeling and simulation. Fig. 3 illustrates the network topology used in our experiments. Each CS node denotes a composite sensor. There are ten such sensors connected to a remote server via switches. There are also two PCs transmitting non-IoT traffics in the network. The IoT application on the sensors can sense five kinds of information: HD video, regular video, still images, infrared beam readings, and magnetometer values, and their corresponding priorities are 2, 3, 4, 5, and 6. The five kinds of information are sent using two TCP connections and three UDP flows depending on the requirements for the transmission reliability.

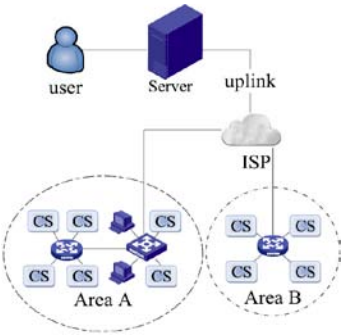


Fig. 3. Experimental Network Topology

We make use of three testing scenarios in the experiments. In the No Congestion scenario, the IoT application has enough network resources to consume. In the Under Congestion scenario, congestion happens 50 seconds after the experiment starts when the two PCs start a video conference with the remote server. In both above scenarios, the IoT application acquires all kinds of information without constraint. In the third Our Scheme scenario, the proposed scheme controls the data acquisition of the IoT application, and congestion happens as in the Under Congestion scenario.

Fig 4. shows the cumulative high-priority IoT data received by the remote server over time. The data is transported via TCP connections for high relaiblity. The slope of the black dotted line shows when there is no congestion the server receives the data at a constant rate. The red dashed line, on the other hand, exhibites when congestion happens the increase of the cumulative data becomes very slow, suggestion the IoT application may have stopped working given there is almost no more supply of the high-priority data. The solid blue line illumnates the performance of our scheme under congestion. After detecting congestion and backing off initially, the schem helps the IoT application regain the fast delivery of the high-priority data since 180s.

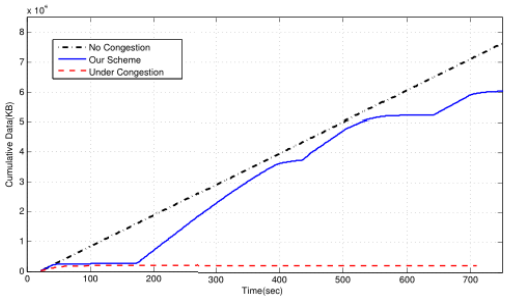


Fig. 4. Cumulative high-priority data received by the server

During 200-380s, the scheme wakes up more simple sensors when it detects more network resources might be available, until it awakes too many of them and brings about a new congesion. This congesion forces the scheme to hibernate some sensors at 400s, which slows down the data acquistion for a while. This pattern of fluctuatations then rolls on in cycles until the end of the experiment.

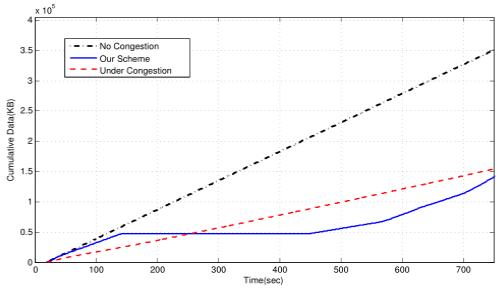


Fig. 5. Cumulative low-priority data received by the server

Fig 5. shows the amount of low-priority IoT data received cumulatively by the server over time. The data is sent via UDP flows since the application can tolerate certain loss. When we compare the slopes of the dotted black line and the dashed red line, we notice the rate of delivery considerablely decreases when the network is congested. The blue solid line indicates our scheme discovers the congestion at 140s. Because it puts the sensors that acquiring the low-priority data to sleep, the slope of the solid line flattens out quickly. The deprived network resources are left to the high-priority data to make sure their normal delivery.

IV. Conclusion

In this paper, we have designed and implemented a data acquisition scheme for IoT that is not only reactive to network congestion but also energy conservative. Our experiments show the scheme has the capability to deliver the high-priority data in a much faster fashion than the previous scheme. We believe this type of data acquistion is essential to the development of IoT and its advantages will become more prominent over time.

Acknowledgment

This work is supported partially by China NSFC-61261019, Inner Mongolia Autonomous Region NSF-113113, and SPH-IMU.

References

[1] Casagras, “Casagras IOT Definition.” http://cordis.europa.eu/search/index.cfm?fuseaction=news.document&N_RCN=30283, 2011.

[2] A. de Saint-Exupay, “Internet of things-strategic research road map.” Cluster of Europe Research Projects on IOT, 2009.

[3] D. Wischik, et al., “Design, Implementation and Evaluation of Congestion Control for Multipath TCP.” *NSDI*. Vol. 11. 2011.

[4] L. Zhang and D. Clark, “Oscillating behavior of network traffic: A case study simulation.” Internetworking: research and experience, 1990.

[5] I. C. S. L. M. S. Committee et al., “Wireless lan medium access control (mac) and physical layer (phy) specifications.” 1997.

[6] L. Atzori, A. Iera, and G. Morabito, “The internet of things: A survey.” Computer Networks, vol. 54, no. 15, pp. 2787–2805, 2010.

[7] H. Wu, et al. “ICTCP: incast congestion control for TCP in data-center networks.” IEEE Transactions on Networking , 2013, 21(2): 345-358.

[8] LEE G, NA S H O, HUH E. Modeling for Congestion Prediction in Wireless Sensor Network Using Traffic Demands Analysis[J]. Mathematical Methods for Information Science and Economics, 2012, 1(1): 206-211.

[9] Huang J, Du D, Duan Q, et al. Modeling and analysis on congestion control in the Internet of Things[C]//Communications (ICC), 2014 IEEE International Conference on. IEEE, 2014: 434-439.

An approach to preprocess data in the diagnosis of Alzheimer`s Disease

Bhagya Shree S R
Research Scholar
PET Research Center
Mandya, India
srbhagyashree@yahoo.co.in

Dr.H.S.Sheshadri
Prof. Department of E & C
PES College of Engineering
Mandya, India
hssheshadri@gmail.com

***Abstract*- The number of people surviving in older age is more. This is mainly due to the developments that have taken place in the field of medicine. Theseold people are prone to many age related diseases. There are numerous neuro degenerative brain related diseases. Dementia is one among them. The people affected by Dementia will have lapse of memory. Alzheimer's disease is one of the types of dementia. Diagnosis of the disease is a time consuming task. To reduce the time needed for diagnosis the medical practitioners use system based approach. To help the practitioners researchers have developed various tools and techniques.**

In this paper the authors focus on classifications of subjects as diseased or not. Before doing classification the data has to be preprocessed. Preprocessing of data is done by applying techniques such as preparation of data, selection of attributes, balancing data, model evaluation and feature selection etc.The authors have collected the data of 466 subjects. The preprocessing techniques are applied on the data set. The subjects are classified using Naïve bayes and J48. The accuracy of the classifications are compared and Naïve bayes is found better.

Keywords-Neuro psychological tests, SMOTE, Wrapping method,Naïve bayes, J48.

I. INTRODUCTION

All over the world there are 44million people suffering from dementia [1].Dementia means loss of memory. This Disease is classified into various types and some of them are Alzheimer's disease, Parkinson's disease, Front temporal lobar degeneration, vascular dementia etc., [2]. Most of the demented patients are grouped under Alzheimer's disease. There are around 38million people suffering from Alzheimer's disease.According to the special report done by Alzheimer`s association done in2013, in 2013, an estimated 5.2 million Americans of all ageshave Alzheimer's disease. This includes anestimated 5 million people of age 65 and olderand approximately 200,000 individuals under age65 who have younger-onset Alzheimer's. One in nine people age 65 and older (11 percent)has Alzheimer's disease. About one-third of people age 85 and older(32 percent) have Alzheimer's disease. Of those with Alzheimer's disease, an estimated4 percent are under age 65, 13 percent are 65 to 74,44 percent are 75 to 84, and 38 percent are 85or older [3]. These facts indicate the need of early diagnosis. There are various risk factors that contribute to the development of the disease. They are Age, Genetics, Smoking, alcohol Intake, Cholesterol, Down Syndrome etc. [4]. The symptoms of the Alzheimer`s diseases are poor decision making, poor judgment, misplacing things, impairment of movements, problem with verbal communication, abnormal moods, complete loss of memory. The diagnosis of AD is done at three different stages namely consulting the General Physician, Undergoing neuro psychological tests and taking MRI scans.Alzheimer's disease and otherdementias are caused by damage to neurons that cannotbe reversed with current treatments [4]. Diagnosis of the disease at the early stage will help the patients to have quality life for the rest of their life. The authors have focused on diagnosis of the disease for neuro psychological test. For diagnosis of the disease machine learning approach is used. In this paper authors focus on various preprocessing techniques that have to be applied to the dataset.The classification techniques Naïve bayes and J48 are applied on the dataset and the results are compared.

II. LITERATURE SURVEY

There are various neuro psychological tests like MMSE, BDIMC, COG, BOMC, MOCA, AD8 and GP CoG etc. Each of these tests has its own advantage and disadvantage and moreover, the tests are meant for a community of people. Of the all MMSE is very popular. But even that has a disadvantage. The disadvantage of MMSE is it is insensitive to early changes of dementia. This indicates the need of a screening test which may be used to the subjects irrespective of gender, religion, culture and education. To overcome this problem 10/66 research group founded by Alzheimer` Association has studied the subjects of various age groups in different developing countries. Theses researchers have designed a battery and they have set normative scores. The paper focuses on diagnosis of AD using 10/66 battery by knowledge discovery from data [5]. This 10/66 battery is preferred compared to the most popular MMSE battery as it is applicable to anyone irrespective of gender, religion, culture and education [6].

The knowledge Discovery process is a procedure that comprises of Data Cleaning, Data integration, Data selection, Data Transformation, Data mining, Pattern evaluation, Knowledge presentations [7].Data mining finds its application in the field of biomedical engineering. Researchers have used data mining for the diagnosis of various diseases.

Abhishek Taneja in his paper has discussed about using data mining for the prediction of heart disease

[8].Tarigoppula V.S sriram et.al has used classification algorithms to detect Parkinson's disease [9].

Breetha S and Kavinila R have discussed about using hierarchical clustering in the diagnosis of cancer and classification of cancer [10].Rashedur M. Rahman and FarhanaAfroz have tested the various classification techniques using various tools likeWEKA, Mat lab, Tanagra for the data sets of diabetes patients [11].

Various techniques are used for discovering the knowledge namely, Association, Sequential pattern, Classification, Decision trees, Neural networks, Visualization, Clustering, Collaborative filtering, Data transformation and cleaning, Deviation and fraud detection, Estimation and forecasting, Bayesian and dependency networks, OLAP anddimensional analysis, Statistical analysis, Text analysis, Web mining etc.

Jyothi Sony has used supervised machine learningnamely Naïve Bayes, K-NN, Decision List algorithm to analyze the datasets of heart disease patients [12].

Tina R. Patil and Mrs. S. S. Sherekar in their paper have done the performance analysis of Naive Bayes and J48 Classification Algorithm for Data Classification [13].

Jehad Ali et.al in their paper compared the classification results of Random Forest and the J48 for classifying twenty versatile datasets. They concluded that random forest gave better results for same number of attributes with large datasets while J48 is handy and it suits only for those with small datasets [14].

Plamena Andreeva and group have tested theparameters of data sets of three different diseases namelybreast cancer, Diabetes Pima and IRIS and published ascholarly article in Google scholar. They have analyzed thedata using various types of classification by using differenttools namely See5, Wiz Why and WEKA. From the results theauthors suggest that WEKA is better in terms of usage, consistency etc. The authors also say that, of the all, WEKApredicts the majority of the data [15].

In this paper the authors focus on application of various pre-processing techniques on the data set. The authors have applied classification techniques and compared the result. WEKA tool is used for implementation.

III. Problem Definition

Data set consist of 466records of subjects aging from 50 to 80 years. Real world data can be incomplete, noisy or it may be lacking in terms of attributes of interest.

The main objectives are:

- To apply the various preprocessing techniques.
- Choosing the appropriate technique.
- Applying classification techniques naïve bayes and J48
- Comparing the results of naïve bayes and J48

IV. Architecture

Data preprocessing is a very important step in knowledge discovery process, as decisions are based on the quality of data. Detection of data anomalies, rectifying the errors and reducing the data to be analyzed will lead to huge pay offs for decision making [16].

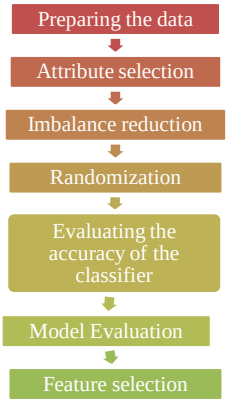


Fig. 1. Flow diagram of various preprocessing techniques.

a) Preparing the data:

The data is often present in the form of spread sheet.HoweverWEKA native data storage format isARFF. The data will be converted from spread sheet to CSV format. Having done this the CSV file is converted to ARFF file. Thus the data has to be converted from spread sheet format to ARFF format[17].

b) Attribute selection:

All the attributes that are present in the file may not be useful. Hence the attributes which are not required are removed using "remove" command [18].

c) Imbalance reduction

The data set consists of positive and negative instances. One of these may be less in number compared to other.This imbalance may lead to under performance of classification methods and experience over fitting.This problem can be overcome by resampling the dataset by applying the Synthetic Minority Oversampling Technique (SMOTE).

c) Randomization

After the application of SMOTE, the number of negative instance will accumulate at the end of the ARFF file.If 10 fold cross validation is applied, to this data set, the data set will be divided into 10 folds. In that case the last fold will have only negative instances. To overcome this problem, unsupervised filter namely 'randomize" is applied. After application of this technique the data set will have same number of records but they will be randomly distributed throughout the ARFF data file.

d) Evaluating the accuracy of the classifier

To obtain a reliable estimate of classifieraccuracy, hold out, random sub sampling, cross validation and boot strap are commonly used techniques. In hold out method the given data are randomly partitioned into two independent sets,Test set and training set. Typically training set will have more instances than test set.

e) Model Evaluation

The data set can be evaluated from,

1. Training set: In this case, the result of each model can be saved and can be visualized.
2. Cross validation: In case of 10 fold cross validation, WEKA develops 10 models, when it displays the result, it uses the average performance of those 10 models. It deletes theremaining models.

From the observations the authors conclude that the model saved with cross validation and with the training set are same.

f) Feature selection

Feature selection is the process of selecting a subset of relevant features for use in model construction.Basically there are two methods,

i. Wrapper method: wrapper method will create all possible subsets from the data set. Then the classification algorithm is used to induce classifiers from the feature in each subset. To find a subset, evaluator will use one of search techniques such as random search, first search, depth search etc.,

ii. Filter method: Filter method uses an evaluator and a ranker to rank all features in a particular dataset. It arranges the attributes according to the rank. By omitting the feature with lowest ranking one at a time, the dominant features can be identified. In this paper authors use wrapper method to select the features of interest.

V. Results and Discussions

The figures below show the results after applying various preprocessing techniques.

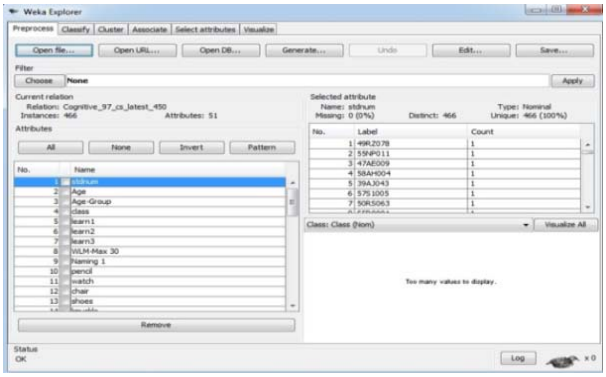


Fig. 2. The CSV file loaded to WEKA

The CSV file is loaded into WEKA. As it can be seen there are 466 instances and 51 attributes.

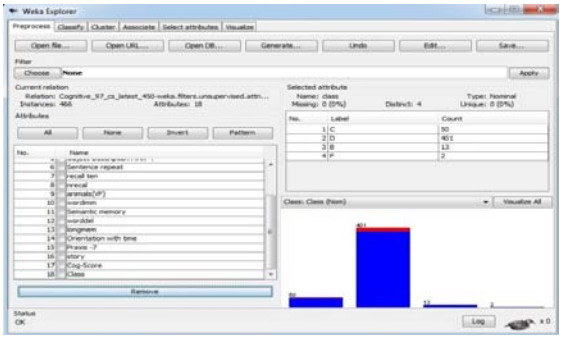


Fig. 3. Attribute selection

The unwanted attributes are selected and removed and the numbers of attributes are reduced to 18.

To reduce the imbalance SMOTE filter is applied. Fig 4 depicts the result after the application of SMOTE.

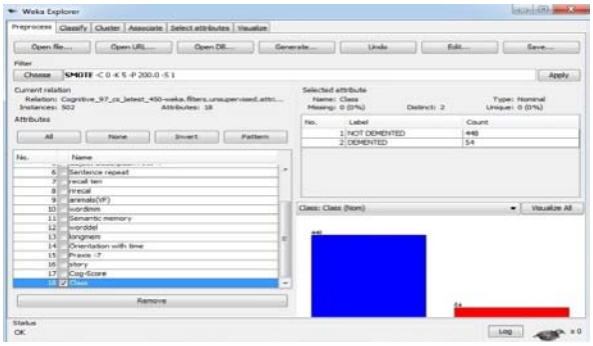


Fig. 4. Imbalance reduction

After applying SMOTE the numbers of positive instances areincreased from 18 to 54, which are accumulated at the end of ARFF file.Fig 5 shows the ARFF file having large number of positive instances accumulated at the end of the file.

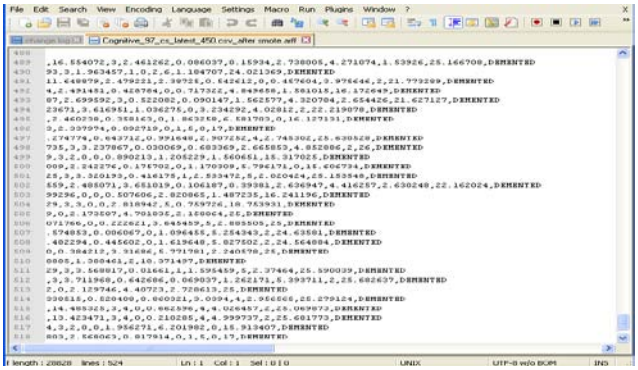


Fig. 5. ARFF file having large number of positive instances.

Fig 6 is the visualization of ARFF data file after randomization.

Classification	*Classification Accuracy*	*Precession*	*Recall*	*Time taken to build the model*
Naïve Bayes	100%	1	1	0.02s
J48	96.5665%	0.998	0.967	0.05s

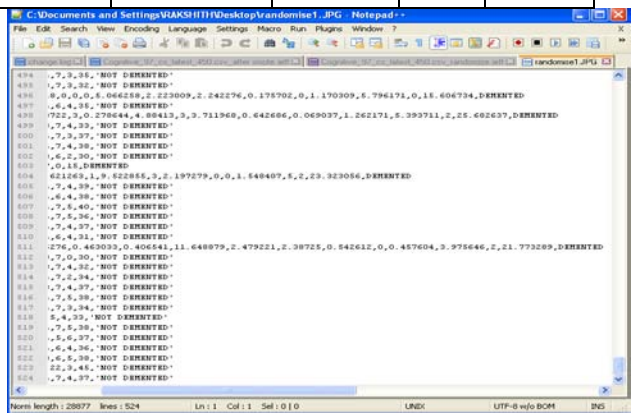


Fig. 6. Randomization result

Wrapper feature selection method, in this method a filter called “best find” is used to select the attribute of interest.

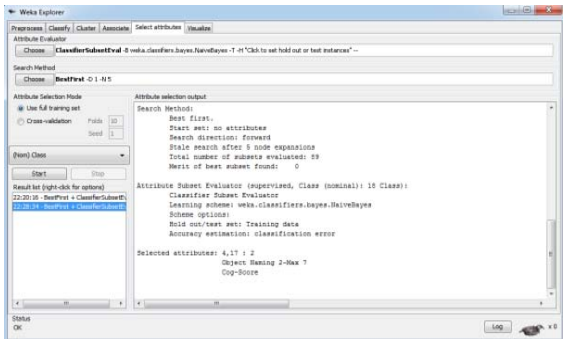


Fig. 7. Selection of attributes which are of less interest.

Fig 7 shows the classifier output which is showing the feature of interest. After selecting the feature, the attributes or features which are of less interest are removed.

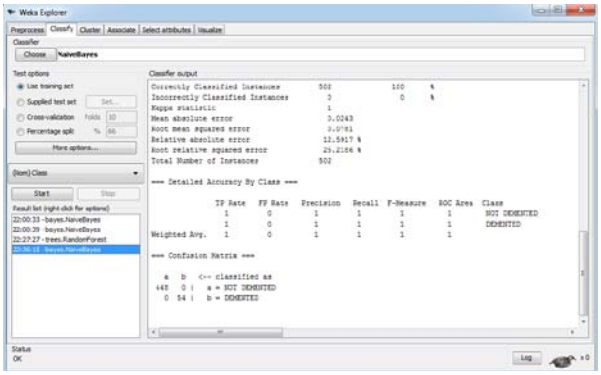


Fig 8: Accuracy is monitored after removing the attributes which are of less interest.

Fig 8 depicts the classifier output with greater accuracy after removing the feature of less interest.

The classification techniques, Naïve bayes and J48 are applied on the data set. The parameters like Classification Accuracy, Precession, recall and time taken to build the model are considered. The summary of the data sets are shown in Table 1.

Table 1: Summary of data sets

The confusion matrices of Naïve bayes and J48 are shown below

=== Confusion Matrix J48===
a b <-- classified as

a	b
433	15
1	17

a = NOT DEMENTED
b = DEMENTED

=== Confusion Matrix Naïve bayes===
a b <-- classified as

a	b
488	0
0	18

a = NOT DEMENTED
b = DEMENTED

VI. Conclusion and Feature Work

As the real word data tends to be incomplete, noisy and inconsistent, the data has to be preprocessed. Various preprocessing methods have been discussed. In preparation of data the missing values can be filled and noisy data can be smoothened. As the dataset is of primarytype there is no

missing of data. The number of attributes is reduced from 51 to 18. The number of instances isbalanced using SMOTE filter.The data which is stored in ARFF format is randomized by using randomize filter. In order to obtain a reliable estimate of classifier accuracy, hold out technique is used. Wrapper feature selection technique is used to select the appropriate feature. By removing the features of less interest accuracy can be improved. The classification techniques are applied and the results are compared. Naïve Bayes performed better than J48.

Future work includes designing an embedded system to facilitate the diagnosis.

ACKNOWLEDGMENT

The authors are thankful to Dr. Murali Krishna, Earlier Scientist Research Fellow, Wellcome DBT Allianz, CSI Holdsworth Memorial Mission Hospital, Mysore, Dr. L Basavaraj, Principal, ATME, Mysore and to the research colleagues who supported with the data in respect of the Alzheimer's disease.

REFERENCES

[1] http://www.capitalfm.co.ke/lifestyle/2013/12/06/44-million- now-suffer-from-dementia-worldwide/

[2] [Viswanathan A, Rocca WA, Tzourio C. Vascular risk factors and dementia: How to move forward? Neurology 72:Pp368–74,2009;.

[3] Thies w, bleiler l, 2013Alzheimer`s facts and figures,"Alzheimer`s dement (journal Of Alzheimer's association), Elsevier Inc. Mar-2013.

[4] Michael saling, Henry Brodaty, Dr. Mark Yates, Dr. Sam Scherer, Professor Kaarin Anstey, "Early Diagnosis of Dementia",2007.

[5] Bhagya shree S. R, Dr. H. S. Sheshadri "An Approach in the Diagnosis of Alzheimer Disease - A Survey"International Journal of Engineering Trends and Technology (IJETT) – Volume 7 Number 1- Jan 2014 ISSN: 2231

[6] Ana Luisa Sosa1 et.al ,\Population normative data for the 10/66 Dementia Research Group cognitive test battery from LatinAmerica, India and China: across-sectional survey," Access NIH public, PubMed central, BMC Neurology,Vol.9,pp 1-11, Aug2009.

[7] Jiawei Han, Micheline Kamber, JianPei,\Data Mining: Concepts and Techniques ," Elsevier , Third edition, 2012.

[8] Abhishek Taneja ,\Heart Disease Prediction System Using Data Mining Techniques," Oriental Journal of Computer Science & technology, Vol. 6,Issue 4, pp 457-466, December 2013 .

[9] Tarigoppula V.S Sriram et.al "Intelligent Parkinson Disease Prediction Using Machine Learning Algorithms" International Journal of Engineering and Innovative Technology (IJEIT) Volume 2, Issue 1,PP 44-52, September 2010.

[10] Breetha S, Kavinila " Hierarchical clustering for cancer discovery using Range check and delta check" International Journal of Scientific and Research Publications, Volume 3, Issue 4, April 2013.

[11] Rashedur M. Rahman et.al "Comparison of Various Classification Techniques Using Different Data Mining Tools for Diabetes Diagnosis" Journal of Software Engineering and Applications, 6, PP 85-97,2013.

[12] Jyoti Soni, UjmaAnsari,Dipesh Sharma and Sunita Soni "Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction", International Journal of Computer Applications (0975 – 8887) Volume 17– No.8, March 2011.

[13] Tina R. Patil, Mrs. S. S. Sherekar "Performance Analysis of Naive Bayes and J48 Classification Algorithm for Data Classification", International Journal of Computer Science and Applications Vol. 6, No.2, Apr 2013.

[14] Jehad Ali, Rehanullah Khan, Nasir Ahmad, Imran Maqsood "Random Forests and Decision Trees", IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 5, No 3, September 2012.

[15] Plamena Andreeva1, Maya Dimitrova1, Petia Radeva2 "Data mining learning models and algorithms for medical applications" http://scholar.google.co.in/scholar /data mining.

[16] Data Mining: Concepts and Techniques by Jiawei Han, Micheline Kamber, JianPei published by Elsevier, Third edition, 2012

[17] Data mining: Practical machine learning tools and techniques by Ian H Witten and Eibe Frank published by Elsevier, second addition 2008.

[18] Insight into data mining theory and practice by K P Soman, ShyamDiwakar and V.Ajay published by PHI learning private limited 2012.

A Multiple-path TCP Congestion Control Algorithm Based on Subflow Correlation Matrix *

Lan Kou
School of Communication and Information Engineering
Chongqing University of Posts and Telecommunications
Chongqing, China

Jianxing Liu
School of Communication and Information Engineering
Chongqing University of Posts and Telecommunications
Chongqing, China

Min Hu
School of Communication and Information Engineering
Chongqing University of Posts and Telecommunications
Chongqing, China

***Abstract*—Due to the multipath transport control protocol(MPTCP) flow can not utilize network resources sufficiently over the unshared bottleneck link, this paper proposes a congestion control algorithm based on the subflow correlation matrix (established by packet loss and RTT information) which is used to detect the status of bottleneck link among subflows by correlation coefficient. Then it performs different congestion control strategies over shared and unshared bottleneck links by adopting fairness definition based on bottleneck link. The results show that the proposed subflow correlation matrix can be used to detect shared bottleneck subflows more accurately and improve overall throughput of MPTCP connection.**

Keywords—multiple path; congestion control; subflow correlationmatrix; throughput

I. INTRODUCTION

Multipath transport control protocol (MPTCP) is one of the most potential Multipath transmission Protocols established by IETF organization MPTCP WG (multipath transport control protocol workgroup)which is responsible for the related standards of MPTCP [1]. MPTCP allows a single data stream to be split across multiple paths, so it can increase network throughput effectively, improve network reliability and enhance the robustness of the network. However, how to control congestion availably when transmit data is increasingly becoming an important issue.

There is fairness[2] problem in MPTCP connection network with traditional TCP [3] congestion control algorithm, while the existing MPTCP congestion control strategy failed to achieve an effective balance between fairness and improving through put [4], which results in MPTCP connection failed to make full use of network resources of the unshared bottleneck link. This article puts forward a MPTCP congestion control algorithm based on the correlation matrix of the subflows (MPTCP Congestion Control Based on Subflow Correlation Matrix, BSCM) through studying the difference of fairness between traditional TCP flows and the MPTCP connections in the shared bottleneck link and unshared bottleneck link.

II. RELATED WORK

At the beginning of the multipath TCP setting up, the reference[5] proposed the Uncoupled TCP algorithm which runs regular TCP congestion control algorithm independently on each subflow. It increases or reduces congestion window on the basis of each flow transmission performance correspondingly. But there is a serious problem that MPTCP connection occupies too much bandwidth with the algorithm to traditional TCP flow in shared bottleneck link which fails to achieve the aim of fairness with other regular TCP flows[6].

In order to solve fairness and congestion equilibrium problems caused by Uncoupled TCP algorithm, the reference [7] proposed the Coupled MPTCP congestion control algorithm by coupling congestion window. The biggest difference between this algorithm and Uncoupled TCP algorithm is that the window of each subflow changes is determined by the total congestion window size. Thus the algorithm can effectively achieve balance congestion, maximize the use of existing network resources [8]. But there is also serious "jitter" so that pose a threat to the stability of the network[9].

For the jitter problem of Coupled MPTCP algorithm, RFC6356 proposed a congestion control algorithm which called "Link increases algorithm" (LIA) [8]. The algorithm ensures that the throughput of network in multipath environment can reach the value when it runs regular TCP with the best single-path. Meanwhile it takes fairness into account among users [9]. However, the algorithm adopts fairness definition based on network, so that the throughput of multipath flow is equal to the best subflow of the multipath with regular TCP no matter the multipath subflows share the same bottleneck link or not. So it cannot reflect the advantages of multipath flow and meet the throughput increase target[10, 11].

III. MPTCP BSCM CONGESTION CONTROL ALGORITHM

As proposed in this article, BSCM congestion control algorithm is based on subflow correlation matrix by establishing the correlation matrix between subflows and then judging the status of correlation to determine whether the subflows pass through the same shared bottleneck link. The BSCM algorithm runs appropriate congestion algorithm based on correlation matrix to improve the network through put under ensuring the fairness of network. This section will elaborate the correlation matrix and congestion control algorithm design.

A. Subflow Correlation Matrix

Each subflow has its own congestion window in the multipath transport streams, the data transmission between the subflows are mutually independent. When subflow i of a MPTCP connection occurs packet loss, while another subflow jalso occurs packet loss before 1 / 2RTT and after 1 / 2RTT, the subflow i is considered ithas certain correlation with subflow j. In another word, the two subflows may share the same bottleneck link [12]. We can establish a correlation matrix between the subflow i and j with parameter φ(si, sj) indicated the degree of correlation between subflow i and j.φ is initialized with 0, which indicates subflows are not sharing the same bottleneck link with each other. Assumed there are 4subflows, the initial subflow correlation matrix can be expressed as shown in Table 1.

TABLE I. INITIALVALUEOFφ(si, sj)WHEN MPTCP CONNECTION SET UP

Subflow	0	1	2	3
0	*	0	0	0
1	0	*	0	0
2	0	0	*	0
3	0	0	0	*

When subflow 1 occurs packet loss, mark congestion flag on subflow 1. Detect whether other subflows occurs congestion simultaneously (observation time before 1 / 2RTT and after 1 / 2RTT when packet loss on subflow 1, observe whether there is packet loss or latency time growing too largely). The flag of packet loss: receive 3 duplicated ACKs or retransmission timeout; the flag of delay growing: smoothed RTT is larger than threshold RTT.

The method of calculating of smooth delay as equation 1:

$$R_{s,r}(i) = (1-\theta) * R_{s,r}(i-1) + \theta * R_r(i) \quad (1)$$

where, $\theta = {}^{1}/_{8}$,$\mathrm{R_r(i)}$ is the latest sampling value of RTT.

Threshold RTT is calculated by the max and min RTT recorded in recent interval of congestion. We can calculate as equation 2:

$$R_{th,r}(k+1) = R_{max,r}(k) - \delta * (R_{max,r}(k) - R_{min,r}(k)) \quad (2)$$

where,$0 < \delta < 1$, and we also find the algorithm can achieve the best performance when $\delta = 0.3$.

If detect there is congestion flag on subflow2, then modify $\varphi(s1, s2)$ with 0.5. We can consider the degree of correlation between subflow 1 and 2 is 50%. In this case the possibility of subflow 1 and 2 sharing the same bottleneck reaches to 50%. Update Table 1 accordingly. At the moment, we cannot think subflow 1 and 2 share the same bottleneck link, we need further observation.

If there is packet loss again on subflow 1, trigger to check the congestion flag on subflow 2. If the congestion flag is true of subflow 2, we will update $\varphi(s1, s2)$ with 0.8. When $\varphi(s1, s2)$>=0.8, we can think the possibility of sharing the same bottleneck link between subflow i and j is more than 80%. So we can think they share the same bottleneck link.

If there is packet loss again on subflow 1, however, there is no congestion on subflow 2 among observation time, if $\varphi(s1, s2)$ >0,reset $\varphi(s1, s2)$ =0, then we no longer think subflow 1 and 2 share the same bottleneck link.

B. BSCM Algorithm Design

As mentioned in the previous section of the subflow correlation matrix that can be quantitatively analyzed to obtain the correlation of two subflows. When the subflow correlation coefficient $\varphi(s1, s2)$=0.8, then it is considered subflow 1 and 2 sharing the same bottleneck link. Based on this result, in this paper, for the different subflows we have designed different congestion control strategy. In order to ensure the network resources can be utilized fully at the unshared bottleneck link to maximize the throughput of MPTCP flow, this algorithm requires subflow to run traditional TCP congestion control algorithm on the unshared bottleneck link. However, in order to ensure the fairness between MPTCP subflow and single TCP flow in the shared bottleneck link, in other words, MPTCP connection cannot get more network resource than the best path with traditional TCP. And the algorithm requires the subflow r to run congestion control algorithm as below.

- Each ACK on subflow r, increase the congestion window w_r :${}^{\alpha_r}/_{w_{total}}$.
- For each loss on subflow r, decrease congestion windoww_r: $w_r/2$.
- In the slow start phase, each new ACK, increase window $\mathrm{w_r}$: θ_r.

where, rule ① and ② are limited to use in congestion avoidance phase;③ is adopted in the slow start phase on shared bottleneck links. It will run the same strategy as traditional TCP in other phase. w_r denotes the congestion window size of subflow r. w_{total} denotes the total congestion window size of subflow r and other subflows shared the bottleneck link with subflow r. θ_r denotes the increment value of congestion window size when subflow r receives a new ACK in RTT in the slow start phase.

A. BSCM Algorithm Analysis

In order to make full use of network resource, we need to assign weight for each subflow dynamically according to the congestion status of each subflow; the optimal path has greater weight is allocated more traffic. Define the weight of each subflow d_i, $0<d_i<1$, $\sum_1^n d_i = 1$.Thus the subflow can obtain throughput T_i as equation 3:

$$T_i = d_i * T_{total} \le d_i * T_{TCP} \quad (3)$$

Set the increment of each subflow congestion window based on its weight of each subflow. The weight is calculated by the ratio of difference between the max congestion window and current congestion window as shown in equation 4.

$$d_1 : d_2 : \ldots : d_r : \ldots : d_n = (w_{max1} - w_1) : (w_{max2} - w_2) : \ldots : (w_{maxr} - w_r) : \ldots : (w_{maxn} - w_n) \quad (4)$$

We can get the value of weight d_r as equation 5.

$$d_r = \frac{w_{maxr} - w_r}{\sum_1^n w_{maxi} - w_{total}} \quad (5)$$

Next, we analyze the fairness between traditional TCP stream and MPTCP subflow from the perspective of throughput model proposed in reference [8]:

1) Subflow r passes through unshared bottleneck link

It will perform as traditional TCP stream when it transmit data with other application in network. The throughput model as equation 6:

$$T = \frac{s}{R\sqrt{\frac{2p}{3\alpha}} + t(3\sqrt{\frac{3p}{8\alpha}})p(1+32p^2)} \quad (6)$$

where R denotes RTT, t denotes retransmission time, s denotes packet size, p denotes rate of packet loss, α denotes increasing factor of congestion window.

Since α_r is 1 at the moment, subflow r can obtain throughput $T = \frac{s}{R\sqrt{\frac{2p}{3}} + t\left(3\sqrt{\frac{3p}{8}}\right)p(1+32p^2)}$.

The regular TCP stream can obtain throughput $T_{TCP} = \frac{s}{R\sqrt{\frac{2p}{3}} + t(3\sqrt{\frac{3p}{8}})p(1+32p^2)}$. So, subflow r and regular TCP stream can keep fairness in the unshared bottleneck link. Meanwhile, MPTCP can make full use of network resource to improve its own throughput.

2) Subflow m and another subflow r are sharing the same bottleneck link

In order to ensure fairness between MPTCP subflow and regular TCP flow we need to keep the total throughput of the subflow m and r not greater than the best path with regular TCP as shown in equation 7.

$$\sum_1^{m+1} T_r \le \max T_{TCP} \quad (7)$$

So,

$$\sum_1^{m+1} \frac{s}{R\sqrt{\frac{2p}{3\alpha}} + t(3\sqrt{\frac{3p}{8\alpha}})p(1+32p^2)} \le max \frac{s}{R\sqrt{\frac{2p}{3}} + t\left(3\sqrt{\frac{3p}{8}}\right)p(1+32p^2)} \quad (8)$$

We get max throughput:

$$\sum_1^{m+1} \frac{s}{R\sqrt{\frac{2p}{3\alpha}} + t(3\sqrt{\frac{3p}{8\alpha}})p(1+32p^2)} = max \frac{s}{R\sqrt{\frac{2p}{3}} + t\left(3\sqrt{\frac{3p}{8}}\right)p(1+32p^2)} \quad (9)$$

Combine equation 3 with 4, we can get:

$$\alpha_r = {d_r}^2 \quad (10)$$

Thus, we need keep $\alpha_r = \left(\frac{w_{maxr} - w_r}{\sum_1^n w_{maxi} - w_{total}}\right)^2$ to ensure the fairness among data streams and make full use of network resource of shared bottleneck link. Meanwhile, we need keep $\theta_r = d_r$ to ensure friendless between MPTCP subflow r and regular TCP flow.

IV. SIMULATION

In this paper, we verify BSCM algorithm with NS3 platform to check whether it can detect accurately subflows share the same bottleneck link or not in shared bottleneck link and unshared bottleneck link scenarios as shown if Fig.1 and Fig.3.And to verify whether MPTCP flow can keep friendly with regular TCP and improve the overall throughput at the same time.

A. Simulate on Unshared Bottleneck Link

The unshared bottleneck link simulation scenario is shown in Fig.1. Where s1 and s3 is regular TCP stream, s2 is MPTCP connection contains subflow sf1 and sf2. The RTT of all the flows are 60ms.

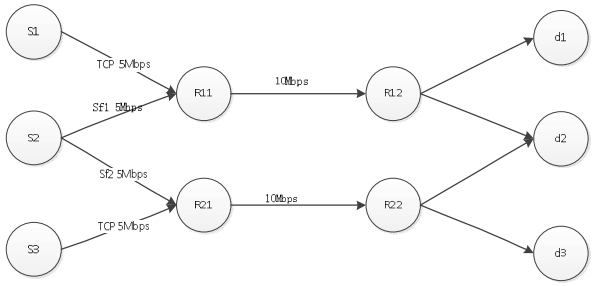
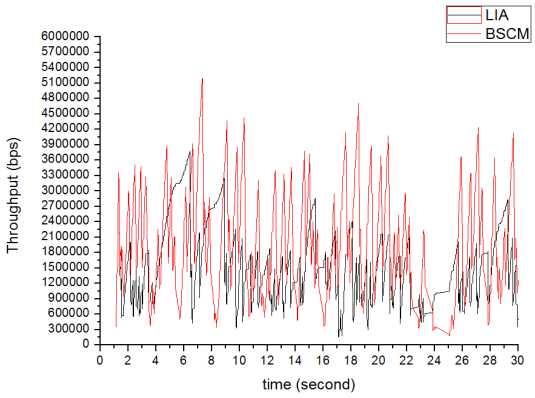
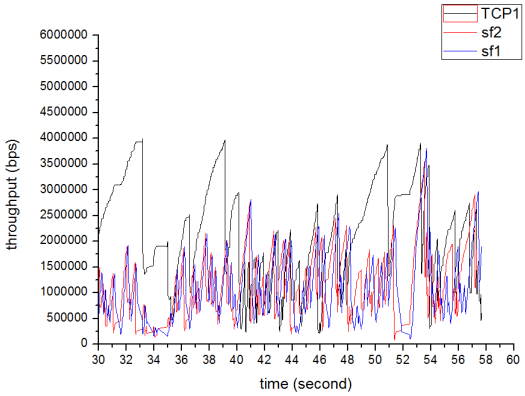


Fig. 1. Unshared bottleneck link simulation topology

In this scenario, the subflow sf1 and sf2 don't share the same bottleneck link. If throughput of sf1 is almost equal to regular TCP stream s1, which means the MPTCP subflow can keep fairness with regular TCP at bottleneck link R11-R12 and BSCM can improve utilization of network resource. The result of simulation with BSCM is shown in Fig.2.



a. MPTCPtotal throughput with LIA and BSCM



b. Throughput ofsf1 with BSCM andregularTCPstream

Fig. 2. Simulation result of unshared bottleneck link scenario

The throughput of MPTCP connection with LIA and BSCM as shown in Fig.2 (a), we can get from the simulation

result: the total throughput of MPTCP connection is 2.7Mbps with BSCM at unshared bottleneck link scenario and 1.5Mbps with LIA, so$T_{BSCM} > T_{LIA}$. That because BSCM is based on bottleneck link fairness definition. The subflows of MPTCP connection don't share same bottleneck link and it will allocate more data than subflow with LIA. So MPTCP connection with BSCM can improve utilization of network resource.

We can get from the simulation result Fig.2 (b), subflow sf1 with BSCM can obtain the throughput almost equal to regular TCP stream, $(T_{SF1})_{BSCM} \approx T_{TCP} = 1.5\,Mbps$ .That meets MPTCP goal about keeping fair with regular TCP. Hence, BSCM can detect accurately whether the subflow share bottleneck with other subflows and improve the throughput under ensure fairness.

B. Simulate on Shared Bottleneck Link

The shared bottleneck link simulation scenario is shown in Fig.3. Where s1 and s3 is regular TCP stream, s2 is MPTCP connection contains subflow sf1, sf2 and sf3. The RTT of all the flows are 60ms and the bandwidth as shown in Fig.3.

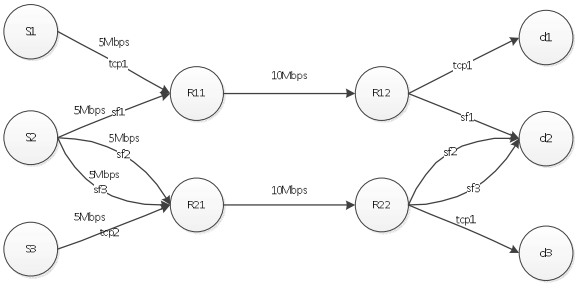
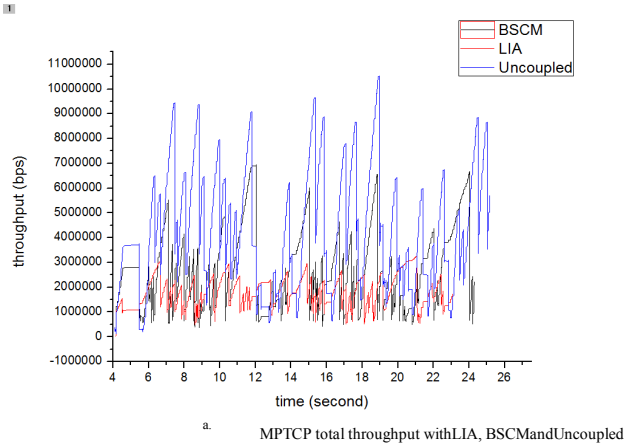
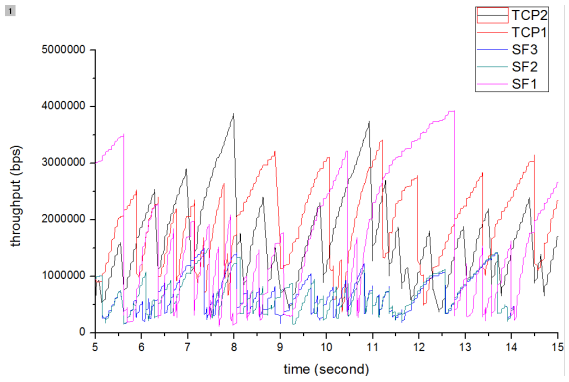


Fig. 3. Shared bottleneck link simulation topology

In this scenario, the subflow sf1 doesn't share the same bottleneck link with other subflows and subflow sf2 and sf3 share the same bottleneck link. If throughput of sf1 is almost equal to regular TCP stream s1 and the total of throughput of subflow sf2 and sf3 is almost equal to s3, which means the MPTCP connection with BSCM can not only keep fairness with regular TCP at bottleneck link R11-R12 and R21-R22 but also improve utilization of network resource. The result of simulation with BSCM is shown in Fig.4.



a. MPTCP total throughput withLIA, BSCMandUncoupled



b. Throughputof sf1 with BSCM and regular TCP stream

Fig. 4. Simulation result of shared bottleneck link scenarios

The throughput of MPTCP connection with LIA, BSCM and Uncoupled as shown in Fig.4 (a), we can get from the simulation result: the total throughput of MPTCP connection is 3.0 Mbps with BSCM at unshared bottleneck link scenario, 2.0 Mbps with LIA and 4.0 Mbps with Uncoupled, so$T_{Uncoupled} > T_{BSCM} > T_{LIA}$. BSCM can detect the subflow sf1 doesn't share the same bottleneck link with other subflows and run congestion control algorithm as regular TCP to utilize the network resource.

We can get from the simulation result Fig.4 (b), the total throughput of subflow sf2 and sf3 with BSCM is almost equal to regular TCP stream s2, $(T_{SF2} + T_{SF3})_{BSCM} \approx T_{TCP}$. That meets MPTCP goal also in this scenario.

V. CONCLUSION

In this paper, we proposed BSCM congestion control algorithm from fairness perspective based on bottleneck link.Simulation results prove that the algorithm can make better use of network resources and achieve to balance congestion than traditional MPTCP congestion control algorithm. Due to limitations, we just tested the algorithmwith relative simple scenarios. Next step we will build a more complex network environment to test the performance of the algorithm for further research and analysis.

ACKNOWLEDGMENT

This work was financially supported by the Foundation and Frontier Research Project of Chongqing Municipal Science and Technology Commission (GrantNo.cstc2014jcyjA40039) and Scientific and Technological Research Program of Chongqing Municipal Education Commission (Grant No.KJ1400402).

REFERENCES

[1] Ford A, Raiciu C, Handley M, et al. Architectural guidelines for multipath TCP development[J]. Internet Engineering Task Force, 2011, RFC 6182:3-7

[2] Honda M, Nishida Y, Eggert L, et al. Multipath congestion control for shared bottleneck[C]//Proc. PFLDNeT workshop. 2009.

[3] Raiciu C, Paasch C, Barre S, et al. How Hard Can It Be? Designing and Implementing a Deployable Multipath TCP[C]//NSDI. 2012, 12: 29-29.

[4] Postel.J. Transmission Control Protocol[J].Internet Engineering Task Force, 1981, RFC 793:10-15.

[5] Wischik D, Raiciu C, Greenhalgh A, et al. Design, implementation and evaluation of congestion control for multipath TCP[C]//Proceedings of

the 8th USENIX conference on Networked systems design and implementation . Berkeley, CA, USA: Usenix NSDI, 2011:8-8.

[6] Chen Y C, Lim Y, Gibbens R J, et al. A Measurement-based Study of Multipath TCP Performance over Wireless Networks[C]//Proceedings of the 2013 conference on Internet measurement conference. ACM, 2013: 455-468.

[7] Raiciu C, Handley M, Wischik D. Coupled congestion control for multipath transport protocols[J]. Internet Engineering Task Force, 2011, RFC 6356: 10-16.

[8] Raiciu C, Pluntke C, Barre S, et al. Data center networking with multipath TCP[C]//Proceedings of the 9th ACM SIGCOMM Workshop on Hot Topics in Networks. ACM, 2010: 10.

[9] Khalili R, Gast N, Popovic M, et al. MPTCP is not pareto-optimal: performance issues and a possible solution[J]. IEEE/ACM Transactions on Networking (TON), 2013, 21(5): 1651-1665.

[10] Raiciu C, Barre S, Pluntke C, et al. Improving datacenter performance and robustness with multipath tcp[C]//ACM SIGCOMM Computer Communication Review. ACM, 2011, 41(4): 266-277.

[11] Zhou D, Song W, Shi M. Goodput improvement for multipath TCP by congestion window adaptation in multi-radio devices[C]//Consumer Communications and Networking Conference (CCNC), 2013 IEEE. IEEE, 2013: 508-514.

[12] Hassayoun S, Iyengar J, Ros D. Dynamic window coupling for multipath congestion control[C]//Network Protocols (ICNP), 2011 19th IEEE International Conference on. IEEE, 2011: 341-352.

Analysis of Information Transmission in GEO+IGSO+MEO Constellation

Yi Liu, Bin Wu, Bo Wang
Beijing Institute of Tracking and Telecommunications Technology
Beijing, China
liuyibittt@163.com

***Abstract*—Satellite constellations with more and more powerful functions are now in busy construction, which makes its status higher and higher in aerospace filed. GEO+IGSO+MEO constellation can provide regional and global navigation. Different demands of various information transmission via Inter-satellite links (ISL) are analyzed based on research hotspots.**

Keywords—Satellite Constellation, GEO+IGSO+MEO, Inter-satellite Links, Demand Analysis

I. INTRODUCTION

Nowadays with aerospace field's making big progress, the capability of a single satellite and satellite platform are increasingly strong [1]. Meanwhile, the connect of each satellite is getting more and more close, which means one single satellite cannot satisfy diverse aerospace mission. Multi-satellites' working together is the main trend of satellite applied technology developing [2]. When several satellites stay in a certain fixed time-space relationship and a stable geometric construction, we call it satellite constellation [3]. Satellite constellation can not only expand the cover area but also improve constellation functions, substantially enhance information acquisition quality, and make the task model more flexible through satellite association and cooperation [4]. Before talking about the constellations, we must know the concept of GEO、MEO and IGSO[5].

GEO, Geosynchronous Earth Orbit, is an orbit of 36000km. Satellites in this orbit have the same period of the earth.

MEO, Medium Earth Orbit, is an satellite orbit between 2000km and 36000km. Satellites of GPS and GLONASS are in this height.

IGSO, Inclined Geosynchronous Satellite Orbit, is an orbit in the height of 36000km. It's a kind of synchronous orbit.

With satellite constellation's rapidly developing, the development of satellite constellation has stepped into a new age [6]. In the meantime, this association can also make the functions of constellation more and more powerful [7]. The structure of satellite constellation is a description of satellites' space distribution, orbit types, and the association. The types of constellation structure in a certain extent will influence link status and routing calculation [8].

The history of satellite constellation covers single satellite age, several satellites age, and constellation age[9], which can be illustrated in Fig.1.

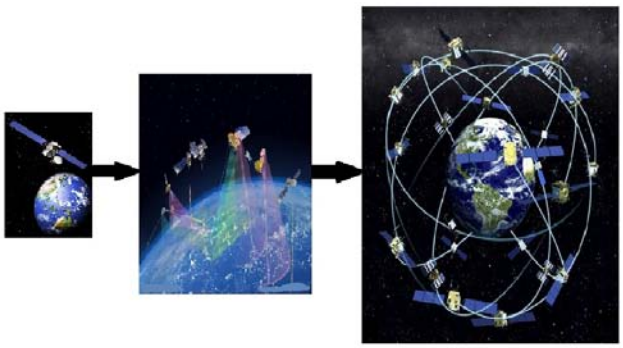


Fig. 1. The Development of Satellite Constellation

When focused on engineering application, constellation system owns an incomparable advantage to single satellite. It is improving rapidly in satellite communication, satellite detection, satellite navigation area and so on. In the design of satellite constellation, mainly applied constellation includes Walker constellation, compound Walker constellation,

elliptical orbit + equatorial orbit constellation, sun-synchronous orbit constellation, heterogeneous constellation, IGSO+GEO constellation, GEO+IGSO+MEO constellation and other constellations. GEO+IGSO+MEO constellation is based on IGSO+GEO regional navigation system by adding into MEO satellites. BeiDou Navigation system consists of GEO、MEO and IGSO satellites, to ensure the global navigation and regional navigation enhancement.

The research of GEO+IGSO+MEO constellation mainly concentrates on the design of ground coverage, constellation structure stability and phased deployment. Researches on inside constellation information transmission are rare. This article analyses the characteristics of GEO+IGSO+MEO and information types, which is the foundation of the constraint and purpose of different information transmission in constellation [10].

II. MODELING OF GEO+IGSO+MEO CONSTELLATION

This satellite constellation is proposed based on GEO+IGSO and GEO+MEO constellations, the basement of which are GEO and IGSO, while global operation is increased with the MEO satellite. In order to realize the global navigation with less satellites, we can arrange the satellites on the pre-computing position, which can be a smooth transition from regional navigation system to the global one. It is illustrated in Fig.2.

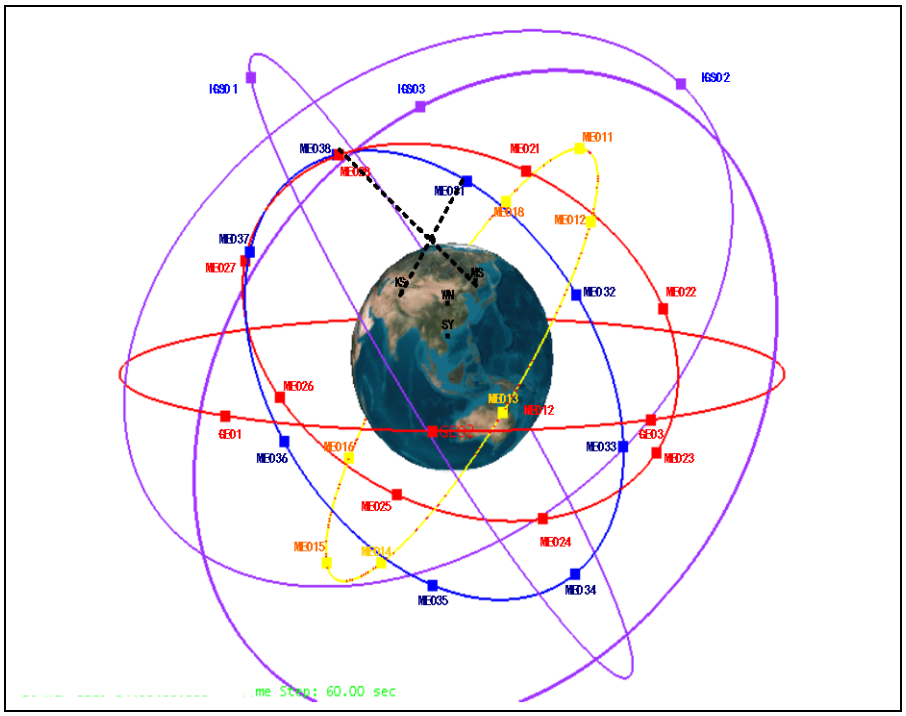


Fig. 2. Fig.2 GEO+IGSO+MEO Constellation

In satellite constellation design of GEO, IGSO and MEO, the ration of coverage on the ground, GDOP value for the specific region, availability, continuity, stability, expansibility, etc, are mainly considered to have more than 5 satellites to cover the same area under the elevation angle constraint condition. In order to get a better geometric characteristic, GDOP<5 in the general service area (such as China and the Asia Pacific region) are required while in the hot service areas (such as a densely populated area), GDOP<4 are required. At the same time, when the constellation systems provide services normally, the service ability of constellation systems does not significantly reduced with one or a few satellites failure, and the quality of the key regional service must be guaranteed. Constellation construction is a long process, which requires to take short-term and long-term benefit and cost into consideration. Meanwhile, the combination with other satellite systems asks for a real high constellation expansibility.

To satisfy future constellation with the ability of autonomous navigation, ISL is needed. Information including telemetry, telecontrol, topology information and others will transmitted through ISL.

III. RESULTS AND ANALYSIS

In order to realize autonomous navigation, precise orbit determination and other purposes, satellites should have capability of inter-satellite communication, inter-satellite ranging and orbit data processing. Inter-satellite link (ISL) is mainly to meet the measurement and communication, to achieve the goal of inter-satellite data transmission. power precision measurement function, and other special extension requires the function etc.

The information discussed in this paper mainly includes remote control, telemetry information, navigation information, satellite ephemeris, time synchronization information, the global integrity information, long term ephemeris and satellite constellation network topology information. Different information has different priority, at the same time, when

different information with different transmission time has different requirements.

Telecontrol is a method of controlling satellite attitude, orbit and other states by microwave or laser. It can be divided into conventional telecontrol and emergency telecontrol. Conventional telecontrol means a remote control when satellites function well, which makes up a real high proportion. When satellites failure or particular control of certain satellites occurs, emergency telecontrol takes place of former control. It occupies small part of the whole operation period. Conventional telecontrol instruction, which is top pouring periodically, has the same priority to other information from ground to satellites. Emergency telecontrol, belonging to sudden information in emergency period, needs real time accurately transmitted to the target satellites. Telemetry is a to measure and transmit inner satellite parameters and other space physics parameters. The information transmitted by telemetry contains different parameters inner satellites, which leads to requirement of stable downloading. Constellation topology, satellite message, routing and navigation information including ionosphere model parameters and satellite clock error, belong to period pouring information. Besides what mentioned above, integrity information is to guarantee autonomous navigation. Other load information is also transmitted in constellation.

When discuss the requirement of different transmission, researchers have to focus on different characteristics. Telecontrol mainly concentrates on transmitting accuracy. Telemetry is to keep stability of transmission. Integrity information needs to be transmitted to target satellites in shortest time. Various data in various missions have different priority. For example, navigation and integrity information are top ones in autonomous navigation while telecontrol and telemetry top the priority in TT&C mission. As satellite failure occurs in constellation, emergency operation should be paid most attention. The main demands of several different information have been illustrated in Fig.3.

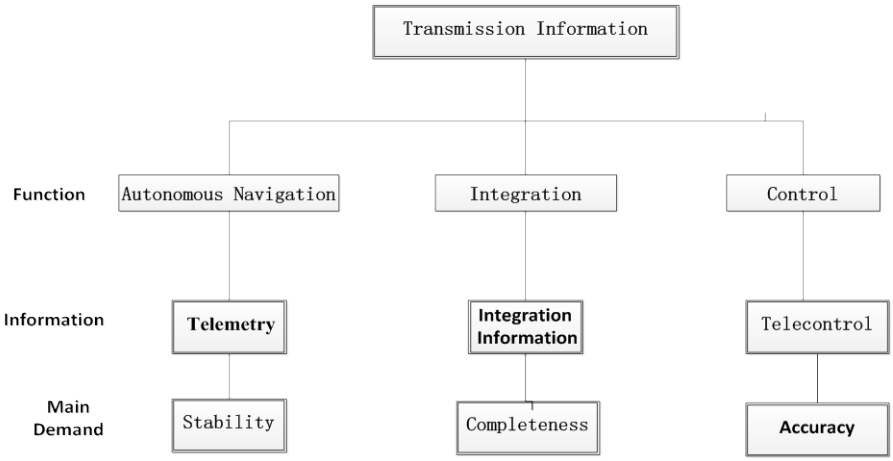


Fig. 3. The Main demand of Several Information

IV. CONCLUSION

Satellite constellation construction is a main trend in aerospace field. Despite of the diversity of satellite constellation, this article concentrates on GEO+IGSO+MEO constellation, and analyses its characteristics. Information transmission among constellation has been discussed. To identify the quantity of different data in different operation is still need to be studied in further research.

REFERENCES

[1] Akjidiz I. F, Jeong S. Satellite ATM networks: A survey [J]. IEEE Communication Magazine,1997, 35(7): 29-40.

[2] Wang K, Zhao Z.W. and Yao L. An agile reconfigurable key distribution scheme in space information network[R]. Second IEEE Conference on Industrial Electronics and Applications.

[3] Allison F. Zuniga, David B. McKissock, and Charles Bard, Integrated Design Analysis for the Constellation Architecture[R], AIAA SPACE 2007 Conference & Exposition. Long Beach,California. p:1-15.

[4] Wertz JW,Collins, JT,Dawson S, etal. Autonomous Constellation Maintenance [A]. Satellite Constellations[C]. Toulouse, France, 199811-11.

[5] LIANG Jun,ZHANG Ji-wei, XIAO Nan. Research and Simulation on an Autonomous Routing Algorithm for GEO-LEO Satellite Networks. International Conference on Intelligent Computation Technology and Automation, ICICTA 2011,Shenzhen, China, 28-29 March, 2011: 657-660.

[6] Information on http://www.beidou.gov.cn

[7] A. Lindgren, A. Doria, and O. Schelen. Probabilistic routing in intermittently connected networks. SIGMOBILE Mobile Computing and Communication Review, 7(3), 2003.

[8] Long Fei, Sun Fu-chun, Wu Feng-ge. A QoS routing based on heuristic algorithm for Double-Layered Satellite Networks. 2008 IEEE Congress on Evolutionary Computation, CEC 2008: 1866-1872.

[9] X. Zhang, G. Neglia, J. Kurose, and D. Towsley. Performance modeling of epidemic routing. Technical Report CMPSCI 05-44, Umass, 2005.

[10] Wood L, Eddy WM, Holliday P. A bundle of problems. In: Proc. of the 2009 IEEE Aerospace Conf. Big Sky: IEEE Press, 2009. 1−17.

A Modified Ant Colony Algorithm to Solve the Shortest Path Problem

Yabo Yuan, Yi Liu, Bin Wu
Beijing Institute of Tracking and Telecommunication Technology
Beijing, China
Email: yaboyuan@gmail.com

***Abstract*—To solve the problem that the ant colony algorithm is easy to fall into local optimal solutions in solving the shortest path problem, improvements on the classical ant colony algorithm are provided in three aspects. Firstly, direction guiding is utilized in the initial pheromone concentration to speed up the initial convergence; secondly, the idea of pheromone redistribution is added to the pheromone partial renewal process in order to prevent the optimal path pheromone concentration from being over-damped by the path pheromone decay process; finally, a dynamic factor is invited to the global renewal process to adaptively update the pheromone concentration on the optimal path, in which way the global searching ability is improved. The results of the simulation experiment show that this modified algorithm can greatly increase the probability of finding the optimal path while guaranteeing the convergence speed.**

***Index Terms*—ant colony algorithm; shortest path; direction guiding; pheromone**

I. Introduction

The shortest path problem, finding the path with minimum distance, time or cost from a source to a destination, is one of the most fundamental problems in network theory. It arises in a wide variety of scientific and engineering problem including geographical information systems, graph algorithms, network optimization and robotics. Classic methods to solve the shortest path problem include the Dijkstra algorithm, the A^* algorithm, *et al.*, but their efficiency is poor when the calculation is large and the real-time requirement is high. Recently, the appearance of heuristic algorithm, such as simulated annealing algorithm (SA), genetic algorithm (GA), artificial neural network algorithm (ANN), the particle swarm optimization algorithm (PSO) and the ant colony system algorithm (ACS), bring light to the settlement of large scale optimization problem [1].

In 1991, Dorigo *et al.* introduced the ant colony system algorithm [2,3], and invited it to traveling salesman problem (TSP), job-shop scheduling problem (JSP), assignment problem [4,5]. It shows the characteristics of robustness, parallelizability, positive feedback and so on. However, there are also shortages such as slow convergence and stagnation phenomenon. In order to solve those problems, a series of algorithm were invited. In [6], Thomas *et al.* presented a MAX-MIN Ant System(MMAS), which limited the pheromone trail within a certain range. In [7], Jun *et al.* used piecewise function to set algorithm parameters. In [8], Zakzouk *et al.* invited a rewards and punishment strategy to control the pheromone trail. In [9], Hufa *et al.* optimized the initial pheromone trail to improve convergence speed.

In this paper, we added heuristic direction information to decide the initial pheromone trail and designed a dynamic factor to adaptively adjust the renewal of pheromone on the optimal solution.

II. The Shortest Path Problem

In graph theory, given a source node S in a weighted directed graph G, with N nodes and h arcs, the single-source shortest path problem (SSSP) from S is the problem of finding the minimum weight paths from S to the destination node E. In this section, we use classic ACS algorithm to solve the shortest path problem.

ACS is meta-heutistic search algorithm which was inspired by the behavior of ant system. Real ants are capable of finding shortest path from a food source to the nest. While working, ants deposit on the ground a substance called pheromone, and follow, in probability, pheromone previously deposited by other ants. This pheromone trail can be observed by other ants and motivates them to follow the path, *i.e.* a randomly moving ant will follow the pheromone trail with high probability. That is the way how the trail is reinforced and more and more ants follow that trail [2,3,10].

In the ACS algorithm, m ants are initially positioned on a N nodes graph and $d_{ij}(i,j=1,2,...,N)$ denotes the distance between node i and node j, $\tau_{ij}(t)$ denotes the pheromone trail on the arc between node i and node j at the *tth* iteration. ACS works as follows [4,5]:

1) Initialize pheromone trail: Initially, the pheromone trail on each arcs are set to be the same, $\tau_{ij}(0)=c(c\neq 0)$.

2) The state transition rule: Ant $k(k=1,2,...,m)$ at node i use rule as below to choose the next node j.

$$j=\begin{cases} arg\ \max\limits_{z\in allowed_k^i} \tau_{iz}(t)\eta_{iz}^{\beta}(t), & q\leq q_0 \\ s, & q>q_0 \end{cases} \tag{1}$$

where $allowed_k^i$ includes notes that are connected with i and have not been visited by ant k, which ensure ants to visit each node at most once. $\tau_{iz}(t)$ is the amount of pheromone trail on edge $<i,z>$ and $\eta_{iz}(t)$ is a heuristic function which is the inverse of the distance between node i and z. β is a parameter which weighs the relative importance of pheromone, q is a value chosen randomly with uniform probability in [0,

1], $q_0(q_0 \in [0,1])$ is a parameter, and s is a random variable selected node according to the distribution given by Equ.(2) which gives the probability with which an ant at node i choose node s to move to.

$$P_{is}^k(t) = \begin{cases} \frac{\tau_{is}(t)\eta_{is}^{\beta}(t)}{\Sigma_{z \in allowed_k^i} \tau_{iz}(t)\eta_{iz}^{\beta}(t)}, & s \in allowed_k^i \\ 0, & s \notin allowed_k^i \end{cases} \quad (2)$$

3) The local updating rule: After ant k finished one tour, use Equ.(3) to update pheromone locally.

$$\tau_{ij}(t+1) = (1-\rho)\tau_{ij}(t) + \rho \triangle \tau_{ij}(t) \quad (3)$$

where

$$\triangle \tau_{ij}(t) = \Sigma_{k=1}^{m} \triangle \tau_{ij}^k(t)$$

ρ is the pheromone decay parameter, $\triangle \tau_{ij}^k(t)$ indicates pheromone gain that ant k bring in after this tour, where $\triangle \tau_{ij}^k(t)$ equals zero if ant k does not visit arc $< i,j >$ or it equals Q/L_k if otherwise (Q is the total pheromone an ant carries at one tour, which is a constant and L_k is the path length of this tour).

4) The global updating rule: After all ants have completed their tours, use the rule in Equ.(4) to update the pheromone of shortest path whose length is denoted as $L_{localmin}$.

$$\tau_{ij}(t+1) = \tau_{ij}(t) + \mu \triangle \tau_{ij} \quad (4)$$

where

$$\triangle \tau_{ij} = \begin{cases} \frac{1}{L_{localmin}}, & < i,j > \in localmin \\ 0, & otherwise \end{cases}$$

III. Improved ACS For Shortest Path Problem

Classic ACS algorithm suffer from shortages such as slow convergence and stagnation phenomenon which are mainly because the pheromone cannot reflect path information accurately []. In this paper, we solve this problem by improving the updating rules.

A. Pheromone Initialization

In classic ACS algorithm, the pheromone trail on each arcs are set to be the same, which will lead ants to search blindly at the beginning and result in a huge amount of biased results [11,12]. In this paper, we added heuristic direction information to decide the initial pheromone trail.

$$\tau_{ij}(0) = Q/(d_{Sj} + d_{jE}) \quad (5)$$

where d_{Sj}, d_{jE} denotes the distance from node j to the start node S and the destination node E separately.

The shortest path between start node S and destination node E is the line connection between them. According to Equ.(5), when node j is close to line SE, $\tau_{ij}(0)$ gets bigger, which lead ants to choose it more frequently. In this way, the initial pheromone carries heuristic direction information, which helps the algorithm to convergent more quickly.

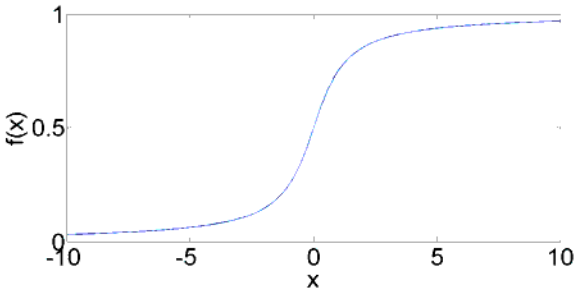


Fig. 1. Function curve of dynamic factor σ.

B. The Local Updating Rule

The local updating is aimed at adjusting the pheromone according to ant tour results and maximizing pheromone of the shortest path so that more ants will follow that trail. However, the local updating rule in Equ.(3) is easy to get biased because the former pheromone decays every time after an ant completes its tour, which makes the positive feedback too weak. Thus, we modify the rule by only update arcs that are visited by the ants. The local updating rule is rewritten as bellow:

$$\tau_{ij}^{k+1}(t) = \begin{cases} (1-\rho)\tau_{ij}^k(t) + \rho \triangle \tau_{ij}^{k+1}(t), \\ \qquad\qquad < i,j > \in path^k \\ \tau_{ij}^k(t), \qquad otherwise \end{cases} \quad (6)$$

where $k = 1, 2, ..., m-1$, $\triangle \tau_{ij}^{k+1}(t) = Q/L_{k+1}$, L_{k+1} is the path length visited by ant $k+1$.

C. The Global Updating Rule

Classic ACS algorithm only updates the optimal results at the global global updating. In this paper, we design a dynamic factor to adaptively adjust the renewal of pheromone on the optimal solution.

$$\tau_{ij}(t+1) = \tau_{ij}(t+1) + \mu\sigma \triangle \tau_{ij} \quad (7)$$

where

$$\sigma = \frac{1}{\pi} arctan(\gamma \frac{L_{localmin} - L_{min}}{\overline{L} - L_{min}}) + \frac{1}{2}$$

γ is the parameter that control the shape of the tangent function. $\overline{L}$ is the average length of all the local optimal results, $L_{localmin}$ is the optimal result of this iteration and L_{min} is the global optimal results.

We set the dynamic factor $\sigma(\sigma \in [0,1])$ to be an arctan function $f(x) = \frac{1}{\pi} arctan\alpha x + \frac{1}{2}$, where $x = (L_{localmin} - L_{min})/(\overline{L} - L_{min})$. As we can see in Fig.1, when $L_{localmin}$ is large, σ will be close to 0, otherwise, it will be close to 1. In this way, the global updating rule can adaptively adjust the renewal of pheromone on the optimal solution.

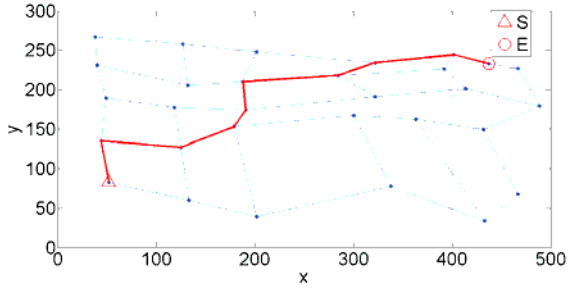


Fig. 2. 30 nodes network.

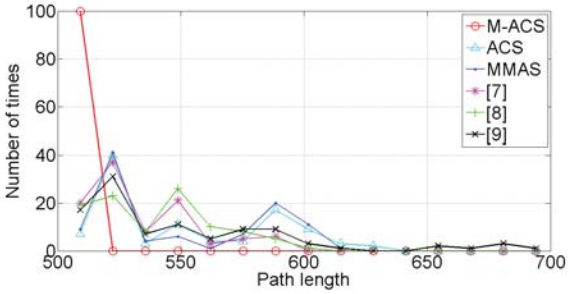


Fig. 3. 30 nodes network.

IV. Numerical Results

The performance of the improved ant colony optimization is tested on 30 nodes network, shown in Fig.2. And the optimal solution is marked with length 509.1250.

We give simulation to the modified ACS (M-ACS) algorithm invited in this paper and compare it with classic ACS algorithm, MMAS algorithm and algorithms provided in [7,8,9]. The simulation parameters are given in Table 1.

TABLE I
Simulation Parameters

m, Number of ants	15
β, Weigh parameter	1
ρ, Pheromone decay parameter	0.2
μ,	2
γ,	1
q_0,	0.5
Q,	100
n_{max}, Maximum iterations	200
$\tau_{ij}(0)$,	1
τ_{max},	0.9
τ_{min},	0.01
M, Number of simulations	100

Simulation results are shown in Fig.3. It can be seen that in 100 times of simulations, the results of the reference algorithms are mainly concentrated on 530, but the results of our algorithm are almost all located at the optimal result (509.1250).

Table 2 shows the statistical results of our simulation. According to Table 2, the percentages of reference algorithms

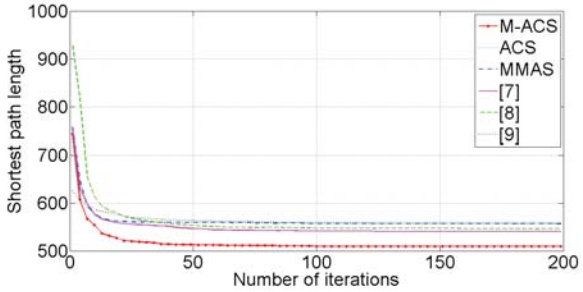


Fig. 4. 30 nodes network.

of getting the optimal results are under 15% and the average results are also poor. However, the M-ACS algorithm can get the shortest path at a percentage of 99% at the cost of small increase in average iteration.

TABLE II
Statistical Results

	Optimal Results	Average Results	Percentage	Iteration
M-ACS	509.1250	509.2266	99%	38.76
ACS	509.1250	565.2422	4%	9.66
MMAS	509.1250	565.5554	9%	16.20
[7]	509.1250	540.5069	11%	30.21
[8]	509.1250	545.9714	10%	37.90
[9]	509.1250	556.6545	5%	29.78

Fig.4 gives the astringency analysis. It can be seen that all reference algorithms stagnate and end at locally optimal solutions while the M-ACS algorithm convergents rapidly to the global optimal solution.

V. Conclusion

To solve the problem that the ant colony algorithm is easy to fall into local optimal solutions in solving the shortest path problem, this paper provided improvements on the classical ant colony algorithm in three aspects. Firstly, direction guiding is utilized in the initial pheromone concentration to speed up the initial convergence; secondly, the idea of pheromone redistribution is added to the pheromone partial renewal process in order to prevent the optimal path pheromone concentration from being over-damped by the path pheromone decay process; finally, a dynamic factor is invited to the global renewal process to adaptively update the pheromone concentration on the optimal path, in which way the global searching ability is improved. The results of the simulation experiment show that this modified algorithm can greatly increase the probability of finding the optimal path while guaranteeing the convergence speed.

References

[1] Tang J T, Wang T, Wang J. Shortest path approximate algorithm for complex network analysis[J]. Ruanjian Xuebao/Journal of Software, 2011, 22(10): 2279-2290.

[2] Colorni A, Dorigo M, Maniezzo V. Distributed optimization by ant colonies[C]//Proceedings of the first European con-ference on artificial life. 1991, 142: 134-142.
[3] Dorigo M, Gambardella L M. Ant colony system: a coopera-tive learning approach to the traveling salesman problem[J]. Evolutionary Computation, IEEE Transactions on, 1997, 1(1): 53-66.
[4] GAMBARDELLA L M, DORIGO M. Ant-Qa rein-forcement learning approach to the traveling sales man prob-lem[J]Proceedings of the Twelfth International Conference on Machine LearningML-951995: 252-260.
[5] DORIGO M, Gambardella. Solving symmetric and asym-metric TSPs by colonies[J]. Proceedings of the IEEE Con-ference on Evolutionary Computation1996: 622-627.
[6] Stutzle, Hoes. MAX-MIN Ant System[J]. Future Genera-tion Computer System Journal200016(8)889-914.
[7] Wang Y, Xue G, Long S. An improved ant colony algorithm for vehicle shortest path problem[C]//Conference Anthology, IEEE. IEEE, 2013: 1-3.
[8] Bullnheimer B, Hartl R F, Strauss C. An improved ant System algorithm for thevehicle Routing Problem[J]. Annals of operations research, 1999, 89: 319-328.
[9] Hu-fa W U, Xue-jun L I, Yu-long Z. Improved Ant Algorithm to Solve Shortest Path Problem[J]. Computer Simulation, 2012, 8: 053.
[10] A A A, Zaher H M, El-Deen R A Z. An ant colony optimization approach for solving shortest path problem with fuzzy constraints[C]//Informatics and Systems (INFOS), 2010 The 7th International Conference on. IEEE, 2010: 1-8.
[11] Shah S, Kothari R, Chandra S. Debugging ants: How ants find the shortest route[C]//Information, Communications and Signal Processing (ICICS) 2011 8th International Confer-ence on. IEEE, 2011: 1-5.
[12] Bell J E, McMullen P R. Ant colony optimization techniques for the vehicle routing problem[J]. Advanced Engineering Informatics, 2004, 18(1): 41-48.

Security Transmission Routing Protocol for MIMO-VANET

Liu Feng[1, 2] , Yang Xiu-Ping[3], Wang Jie[1]

[1] School of Software
Central South University
Changsha, China

[2] Department of Information Science and Engineering
University of Jinan
Jinan, China

[3] Shandong Jianzhu University
Jinan, China

***Abstract*—The security routing protocol for Multi-Input-Multi-Output Vehicular Ad-hoc Network (MIMO-VANET) was presented in this paper, to prevent the eavesdropping in VANET. When building the routing table, this protocol stores the current channel states and restricts the VANET nodes, excepted source/ destination nodes, from obtaining the unrelated routing information. During the information transmission process, the presorted channel states in network layer are transparently transmitted to the physical layer, and decomposed as total channel state. The decomposed results, which are only known by the source/destination nodes from routing information, can be utilized to equalize the channel and to demodulate the information. As a result, the other nodes cannot effectively equalize the channel, which causes little information can be eavesdropped. The theory analysis and simulation illustrate that , when the destination nodes obtain the good signal quality , the eavesdropper suffers poor bit error rate.**

Keywords—Routing protocol; Multi -antenna; Vehicular Ad-hoc Network (VANET); Security transmission

I. INTRODUCTION

Vehicle ad hoc network is self-organized communication network which is carrying on vehicles, and it can provide the services of communication between vehicle, traffic information sharing, traffic safety guarantee and other aspects. It is also one of the most effective solutions of future traffic communication system. Because the information will pass a number of vehicles (node) in the transmission process of VANET, each node on the route receives complete information, resulting in serious hidden danger of physical layer.

In order to improve the security of network information transmission, we can consider from two aspects of route choosing and physical layer transmission.

1)The existing VANET routing protocols can be divided into two categories, network topology-based and geographical position-based, the ones of typical routing protocols based on network topology are AODV(Ad-hoc On Demand Distance Vector Routing) and DSR(Dynamic Source Routing in Ad-hoc Wireless Networks).And one of typical routing protocols based on geographical position is GSR(Geographic Source Routing).In the aspect of VANET routing protocol, the existing research is focused on improving the routing discovery, the efficiency of establishing process and the reliability, and the research on secure routing is relatively weak.

2)The research of VANET physical layer is focused on designing high-performance modulation in order to overcome the influence of channel noise and multiuser interference, and to improve the reliability of information transmission in the physical channel. The key points in discussing content include updating algorithm of channel estimation, information modulation method and so on. In addition, MIMO (Multi-input-Multi-output) technology can provide directional propagation to effectively reduce multiuser interference and to improve the safety of physics layer. Thereby MIMO becomes an important candidate for the physical layer of VANET.

In conclusion, the information security transmission problem of in the existing MIMO-VANET is relatively weak in two aspects of routing protocol and physical layer transmission mode. The efficiency maximization of multi protocol layer can be achieved by transmission of important parameters. Therefore, this paper is aimed at securing information transmission by designing VANET transmission protocol. The idea is to build a restrictive routing discovery process on condition that the nodes are known its own location information. So that each node can only get the necessary routing node location and channel information for transmission information and cannot get the illegally receiving information. Combined with more powerful tropic MIMO transmission technology, by the transparent transmission channel state information in the process of establishing the routing table, the VANET transmission scheme for the secure transmission of information will be eventually achieved.

II. THE PROBLEM MODEL

The VANET model discussed in this paper is shown as figure 1. The system contains several mobile nodes (cars), and each of the vehicles is loaded with several send-receive antennas. Assuming the send-receive antennas as N. The vehicle can be informed of its real-time geographical location, and can only communicate with other vehicles within the scope of communication. In the figure, there are the vehicle B and C within the range of vehicle A and vehicle A, C, D and E within the range of vehicle B, and other nodes' communication range is shown in figure 1. Assuming that wireless channel does not change in routing survival time. Channel state matrix between

the vehicle A and B is set for H_{AB} . The element of corresponding matrix H_{AB} from sending antenna i to receiving antenna j is set for $H_{AB}(i,j)$. Assuming that the sending signal is X_k of vehicle $k(k \in \{A,B,C,D,E,F,G\})$, and the receiving signal Y_{ik} is from vehicle i to vehicle k, and therefore $Y_{ik} = H_{ik}X_k$. To vehicle B, if the expected receiving signal is Y_{AB}, it's defined that the Signal-to-interference Ratio is

$$\eta_{AB} = Y_{AB} / \sum_{i \neq A} Y_{iB} \quad (1)$$

The traditional VANET routing protocol based on geographical location in the process of establishing and updating is relatively independent from the physical layer transmission process. Therefore we cannot take advantage of its geographical location information to improve the reliability of information transmission in physical layer. For example as Figure 1, the issues discussed in this paper is that how does vehicle A choose routing by protocol designing in order to transmit information from A to G more safely and to minimize multi-user interference $e\eta_{AB}$ caused at the same time.

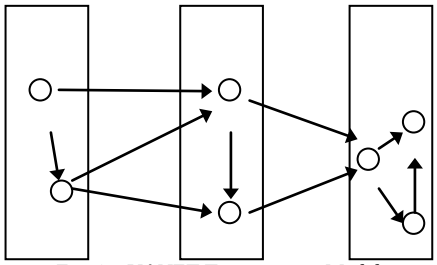


Fig. 1. VANET Transmission Model

III. SECURITY PROTOCOL DESIGN

A. Protocol Architecture

Cross-layer security protocol of MIMO-VANET is aimed at solving two problems: 1) Establishing reliable and efficient transmission route; 2) Securing information transmission in physical layer.

MIMO technology can reduce multi-user interference effectively, and provide natural protection to establish reliable route for VANET. Besides, using nodes' positional information, VANET can find the best route by traditional shortest path algorithm. On the other hand the quality of physical channels between nodes must be considered when constructing the route. Therefore, channel information of physical layer must be transmitted through to the network layer when establishing routing in order to facilitate the routing of the channel condition into consideration.

Physical layer information security transmission is a key problem of VANET, the use of MIMO technology can effectively restrict the range of information transmission, and can also improve safety to a certain degree. But on the route where information passes, all nodes will receive transmitted information perfectly and stably, therefore, the physical layer is still facing some threats in security. In fact, if the condition of which information be restricted when transmitting through each wireless channel is known to originating node and destination node, and other nodes cannot use complete channel condition to achieve channel equalization, we can effectively prevent the threat caused by the nodes on the route. To achieve the ideas above-mentioned, node-corresponding channel state information need to be transparently transmitted from protocol senior layer to physical layer.

Multi-antenna Technology in the physical layer can be used to reduce multi-user interference of nodes, and protocol complexity of physical layer channel transmission can be reduced by the information transparent transmission. Principle of its structure is divided into two parts of process: the routing table update and route choosing.

B. Protocol Procedure

The routing table update process: in this stage, the source node set target to the destination node to flood nodes all around, so that a range of node location information and channel state information can be obtained, so as to establish the local routing sorrow. The steps are as follows: 1) the source node broadcast first to its neighbor nodes with RREQ packets, which contains destination node ID, broadcast serial number, the source node geographical coordinates (GPS coordinate), source node ID, the hop count information and the routing node geographic coordinate channel state of which has already passed. In broadcasting, broadcast information is repetitive transmitting from each antenna of the source node according to time division in different time slots in sequence.

Each antenna at the transmitting time slot, sends synchronous pilot frequency information first, and then RREQ packet information. When receiving the RREQ packet, neighbor node extracts source node location information and source node ID from it and establishes a reverse route to previous one, on the other hand, achieves channel estimation according to the pilot information transmitted from each antenna and stores channel state matrix.

Then the neighbor node query its storage routing table, if effective routes to the destination node are found, the route reply packet RREP will be transmitted according to the shortest route from current node to source node, including the geographical position, the source node to the destination node, the source node ID required for routing node location, channel state matrix and hop count.

If effective route to the destination node is not be found, RREQ packet will be transferred to the neighbor node. At the same time, the physical layer transmit RREQ information from different antennas in turn.

After the routing table is established completely, the relative routing of geographic location information is stored, and the channel state information of each routing node is also contained.

Assume that at initial time all nodes do not have geographic location information of nearby nodes, node A need to build routing table reaching to node G. First, node A broadcast RREQ packet information all around, and node C which is closer to node A receives the information and begin to broadcast, meanwhile node B also receives the message and begin to broadcast. After receiving the broadcast message from node B, node A and node C query broadcasting number and discard the message. D and E receive message and broadcast RREQ messages in turn all around until the message is

transmitted to node G. After received broadcast message from Node E and node F, node G stores channel state matrix $H_{E,G}$ $H_{F,G}$ first and then sends a RREP packet to E, F. After receiving RREP, E, F store geographical position and channel state information of node G.

For an example as specific storage process of node E, due to node E receiving RREQ message sent from B and after broadcasting to the node G and node F, node E will receive RREP message sent from node F and node G. In this process E stores route B→E→G and E→F→G, and associated node location and channel state information, then E will send the RREP information to B. The RREQ message broadcasted from node D is later than node B, so node E no longer broadcast channel information after receiving message from node B, and transmit subsequent node and channel information stored in it as RREP mode to node D. Other node's routing table establishment process is similar to node E, therefore we can find out that routing information stored in node A to node G is sequentially reduced, and the information of back-end node of which is irrelevant to current node is reduced too. And thus, receiving data illegally is interfered in the information transmission, and physical layer transmission security is improved.

The process of information transmission: MIMO-VANET routing consists of two parts: routing establishment and settings of physical layer transmitting parameters. Due to node A has grasped all routes reaching to destination node G, node A transmit channel condition routing A→B→E→G to node G through routing A→C→D→E→G, then node A will transmit the information waiting to be sent to node G along optimal routing A→B→E→G. In order to guarantee the physical layer transmission reliability, the parameters of multi-antenna transmitting needs to be set. If the established node on the route is P_1、P_2、...、P_L, the channel state matrix is$P_{1,2}$、$P_{2,3}$、...、$P_{L-1,L}$. When the information is transmitting between two adjacent nodes, protocol senior layer will transmit channel state through to the physical layer. Then calculate the total channel state matrix of the whole route

$$H = \prod_{i=1}^{L} H_{i,i+1} \tag{2}$$

By formula (2), using multi-antenna beam forming method on the whole routing, first, process channel Singular value decomposition $H = U \sum V$ before sending.

In advance, the sending node is correspond to the route storage U and the receiving node storage V. Sending node use U to send message after pre-equalization, and the received information of the second node P_2 on the route is

$$\begin{aligned} Y_2 &= H_{1,2}V^{-1}X + N_1 \\ Y_3 &= H_{2,3}\left(H_{1,2}V^{-1}X + N_1\right) + N_2 \end{aligned} \tag{3}$$

N_j is the channel additive noise, V^{-1} is pre-equilibrium matrix. And the like, the receiving information of M node $(2 \le m \le L-1)$ is

$$\overline{Y}_L = \prod_{i=1}^{L} H_{i,i+1}\ V^{-1}X + \sum_{k-1}^{m}\left(\prod_{j=k+1}^{m-2} H_{j,j+1}\,N_k\right) \tag{4}$$

Due to the addition of source node, the M node on the path is unknown to pre- coding matrix V, and to channel matrix before node P_m. Therefore, it's very difficult to restore information X in the intermediate node routing.

The signal of destination node using V to equilibrium is

$$\begin{aligned} \overline{Y}_L &= U^{-1}\prod_{i=1}^{L} H_{i,i+1}\ V^{-1}X + \\ &U^{-1}\sum_{k=1}^{L}\left(\prod_{j=k+1}^{L-2} H_{j,j+1}\,N_k\right) \\ &= \sum X + U^{-1}\sum_{k=1}^{L}\left(\prod_{j=k+1}^{L-2} H_{j,j+1}\,N_k\right) \end{aligned} \tag{5}$$

Destination node is stored in the equilibrium matrix U^{-1}. Through exhaustion for the storage equilibrium matrix and drawing a check mechanism (such as in the physical to join the CRC check digit), we get the right diagonal matrix $\sum$.

IV. Protocol Performance Analysis

A. Comparation with GSR - 802.11 n

Using multiple antenna beam forming method, we do the singular value decomposition before sending, $H_{ij} = U_{ij}\sum_{ij} V_{ij}$, $i,j \in 1,2,\ldots,L$. $U_{i,j}, V_{i,j}$ is orthogonal matrix. $\sum_{ij}$ is diagonal matrix. The sending node will store $U_{i,j}$ in the routing table in advance the next-hop node corresponding to the position. The receiving node store $V_{i,j}$. After preliminary equilibrium, the sending node sends information. The receiving node receives information is

$$\begin{aligned} Y_{802.11n.j} &= H_{i,j}\cdot V_{i,j}^{-1}\cdot X + N_j \\ &= H_{i,j}\cdot X + N_j \end{aligned} \tag{6}$$

N_j is channel additive noise, $V_{i,j}^{-1}$ is preliminary equilibrium matrix. After equilibrium, the signal at the receiving end is

$$\begin{aligned} Y_{802.11n.j} &= U_{i.j}^{-1}\cdot Y_j \\ &= U_{i,j}^{-1}\cdot H_{i,j}\cdot X + U_{i,j}^{-1}\cdot N_i \\ &= \sum_{i,j}\cdot X + U_{I,J}^{-1}\cdot N_j \end{aligned} \tag{7}$$

And the like the receiving information of node m$(2 \le m \le L-1)$ is

$$\begin{aligned} Y_{802.11n,m} &= \prod_{k=1}^{m-1} U_{i,i+1}^{-1}\,H_{i,i+1}V_{i,i+1}^{-1}X + \\ &\sum_{k=1}^{m-1}\prod_{j=k+1}^{m-1} U_{j,j+1}^{-1}V_{k,k+1}^{-1}U_{k,k+1}^{-1}\,N_k \\ &= \prod_{i=1}^{m-1}\sum_{i,i+1} X + \sum_{k=1}^{m1}\prod_{j=k+1}^{m-1}\sum_{i,j+1} U_{k,k+1}^{-1}N_k \end{aligned} \tag{8}$$

From formula (8), we can tell that when VANET uses the physical layer transmission scheme of 802.11n, any node on the route can still successfully received information X through $\prod_{i=1}^{L-1}\sum_{i,i+1}$. Contrast to formula (4), the formula proposed in this paper is better.

For the safety of analysis formula, we simulate the protocol proposed in this paper. The physical layer information

modulation mode is QPSK, and the antennas number of the node is N=8. The process of information transmission is from A to G node. The elements of channel state matrix $H_{i,j}$ obey the Rayleigh fading, and then, transmit information as the way shown by formula (5) and as the routing table shown in this paper. And compares with 802.lln physical layer transmission method shown in formula (8).

B. Error Performance Simulation

In order to further observe the performance of the transmission mode, we get the statistical information of transmission error rate. The simulation parameters set the antenna number is eight, information for QPSK modulation way, information transmission process from point A to G nodes, channel state matrix elements $H_{i,j}$ is Rayleigh fading. Simulation of transport process is according with the formula (5). We set noise power N_k from 0 to 2mW. After singular value decomposition, we observe the second, three, four singular value corresponding to the error of parallel channel.

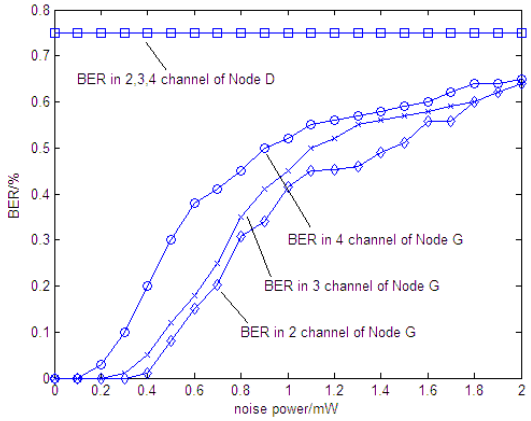


Fig. 2. The BER curve of node G and D

First of all, according to the routing discovery process to establish A to G routing A and B to E, G, and then transmission of information is carry on. The simulation results as shown in figure 2, 3, figure 2 is receiving signals error rate of node D and G node. Node D is in the reception area of B, so it can receive information from the B, Because there is no storage of complete channel information on routing A and B to E, G, so it cannot demodulate information correctly. The figure 2 show that the 2, 3, 4 road information error rate close to 75% in node D, so judgment is the four constellation points as the probability of a same.

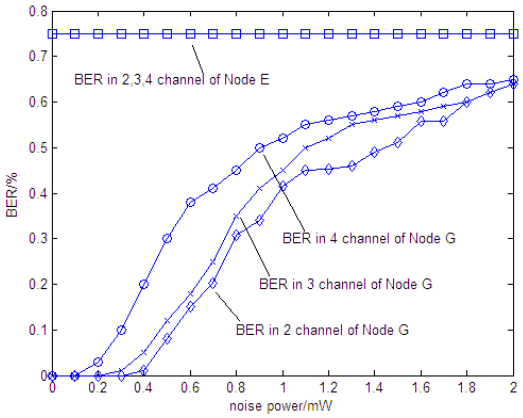


Fig. 3. The BER curve of node G and E

Figure 3 is node G and E received signal bit error rate curve. From the diagram, Although node E in the transmission route, because it only stores a part of the routing channel state information, so it cannot do the whole routing channel equalization. Therefore, shown in figure 3 when C nodes maintain normal reception, E ber curve remain at around 75%

V. CONCLUSION

This paper proposes a information security transmission protocol in MIMO - VANET.

The protocol is divided into two parts: 1) The first is using a routing method based on geographical location information in the process of route setup, and join the limitation of node learn some other related routing information, and corresponding position in the routing table to store the physical through to the network layer of the channel state information; 2) The second is the total channel equalization between the source and destination node at the time of information transmission, by using channel state of the network layer through to the physical layer. Due to do not have the balance of total channel information, so other nodes cannot receiving transport information smoothly, finally we realizes the information security in the MIMO - VANET. Because of the complexity of the wireless channel, the next step work is discussed the influence of channel estimation accuracy for security.

REFERENCES

[1] Task Group p, IEEE P802.11p: Wireless Access in Vehicular Environments (WAVE). draft standard ed. IEEE Computer Society, Los Alamitos (2006)

[2] Wang, S.Y., Lin, C.C., et al.: On Multi-hop Forwarding over WBSS-based IEEE 802.11(p)/1609 Networks. In: IEEE 20th International Symposium on Personal, Indoor and Mobile Radio Communications, Tokyo, September 13-16 (2009), E-ISSBN 978-1-4244-5123-4 I.S. Jacobs and C.P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, G.T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271-350.

[3] Farah, A.E., Bertrand, D.: A Light Architecture for Opportunistic Vehicle-to-Infrastructure Communications. In: Proceedings of the 8th ACM International Workshop on Mobility Management and Wireless Access, MobiWac 2010, Bodrum, Turkey, October 17-18 (2010)

[4] Yi, W., Akram, A., et al.: IEEE 802.11p Performance Evaluation and Protocol Enhancement. In: Proceedings of the IEEE International Conference on Vehicular Electronics and Safety, Columbus, OH, USA, September 22-24 (2008)

[5] Todd, M., Tammy, M., et al.: Measuring the Performance of IEEE 802.11p Using ns-2 Simulator for Vehicular Networks. In: IEEE International Conference on Electro/Information Technology (EIT), Ames IA (2008)

[6] Stephan, E.: Performance Evaluation of the IEEE 802.11p WAVE Communication Standard. In: 66th IEEE International Conference on Vehicular Technology, VTC (2007)

[7] Kang, L.C., Huang, L.C.: Development of Telematics Communication System with WAVE DSRC. In: Proceedings of the 2009 IEEE International Conference on Systems,Man, and Cybernetics, San Antonio, TX, USA (October 2009)

[8] European ITS Communication Architecture: Overall Framework, Proof of Concept and Implementation, Draft Version 2.0, COMeSafety Specific Support Action (October 2008)

[9] IEEE Standard for Wireless Access in Vehicular Environments (WAVE)-Multi-channel Operation, IEEE Std. 1609.4-2010, Aug. 2010.

[10] N. Suganuma and T. Uozumi.: Precise position estimation of autonomous vehicle based on map-matching, in Proc. IEEE IV Symp., 2011,pp. 296–301.

Cognitive Theory Applied to Radar System

Chenghong Zhou, Weiping Qian
Beijing Institute of Tracking and Telecommunications Technology
Beijing, China
somezbx@163.com, qianweiping@bittt.cn

***Abstract*—Cognition, a basic concept in psychology, is the set of a human's inner mental activities and processes related to knowledge. Actually it is possible to introduce cognition into engineering systems to construct intelligent systems with human cognitive functions. Cognitive radar is such a successful case that improves the working performance in complex environment. The article starts with the concept of cognition and cognitive radar, subsequently followed by the basic elements including intelligent information processing, information feedback. The fundamentals how cognition increases the perceptual ability of radar are discussed in detail.**

Keywords—cognition; cognitive radar; perception action cycle; intelligent information processing; information feedbac.

I. INTRODUCTION

Cognitive science is a developing field which focuses on the concept of cognition from different perspectives, such as psychology, artificial intelligence, neuroscience and so on. Cognition originates from psychology, aiming at analyzing a human's inner mental activities and processes related to knowledge, including attention, memory, language, thinking, etc. Cognition contains different contents with various characteristics, for example conscious attention and unconscious perception, concrete or abstract, as well as intuitive knowledge and conceptual model.

It is innovative to introduce cognition into engineering systems and proves to be feasible and successful[1]. For past fifteen years, cognitive principle draws much attention and plays an increasingly important role in the fields of signal processing, telecommunication, control and so on. The application of cognition principle to engineering and improves the performance of the corresponding systems and promotes the foundation of a new integrative field, cognitive dynamic system, focusing on the radar[2], radio[3, 4] and control[5] system with human cognition functions. Specifically, once a dynamic system is capable of all the human cognition elements, including the perception-action cycle, memory, attention as well as intelligence, it is cognitive. Thank to the cognitive function, cognitive dynamic system[6, 7] usually shows good environmental interactivity, such that it perceives the environment in an intelligent way and subsequently reacts to that adaptively.

The development of cognitive radar is only in the beginning stages[8]. In 2006, the concept of cognitive radar, an intelligent radar system that emulates the way the visual brain perceives the environment and responds to the dynamic environment by transmitting adaptive waveform, was put forward by Canadian professor Simon Haykin from McMaster University[9]. He emphasized importance and function of the feedback information from receiver and the help of the knowledge association system.

At the same time, The Defense Advanced Research Projects Agency (DARPA)[10] and the Air Force Research Laboratory (AFRL) led to carry out research on cognitive radar, from fundamental theory, radar system, performance optimization, system framework and embedded computation construction, etc., subsequently designed and constructed a ground moving target indicator (GMTI), which worked against the interference of complex electromagnetic environment, such as dense target background, large discrete scatterer, non-stationary clutter, inhomomorphic clutter, and electronic countermeasure[11]. In the first cognitive radar academic monograph, Cognitive Radar: The knowledge-aided fully adaptive approach, Joseph R. Guerci compares the differences between classical modern radar and cognitive radar. The adaptability of traditional adaptive radar are often confined to the receiver, while the cognitive radar are appropriate for both transmitter and receiver. Compared with the classical radar, cognitive radar introduces many advanced modes, structures, functions, such as multiple degrees of freedom adaptive transmission, knowledge-aided system and so on.

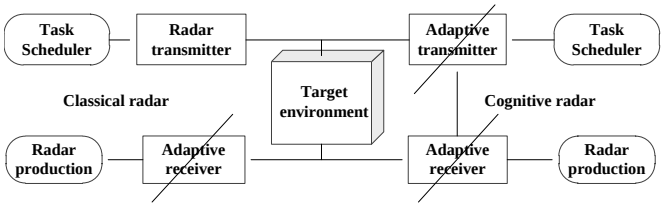


Fig. 1. Basic structures of classical radar and cognitive radar

II. PERCEPTION-ACTION CYCLES

Perception action cycle [12] is the basic structure of cognitive system. Professor Principe from University of Florida mapped the visual brain cognition to engineering[6]. It is the same structure that connects human brain and engineering systems. Fig.2-Fig.4 illustrate the perception action cycles of visual brain system, cognitive dynamic system and cognitive radar.

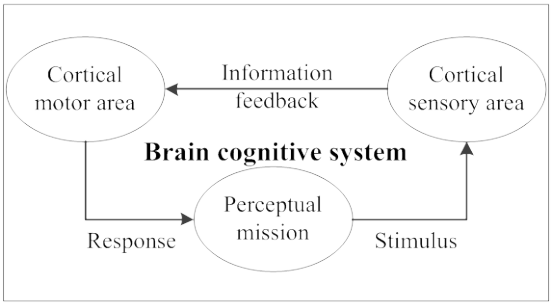


Fig. 2. Perception-action cycle of brain cognitive system

The perception-action cycle of brain cognitive system is discussed by Fuster in Cortex and Mind: Unifying Cognition. [13]

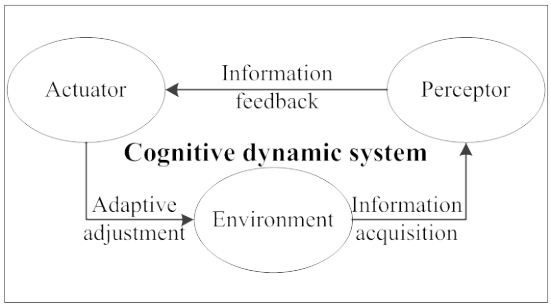


Fig. 3. Perception-action cycle of cognitive dynamic system

The feedback cycle of cognitive dynamic system has the same structure with perception-action cycle.

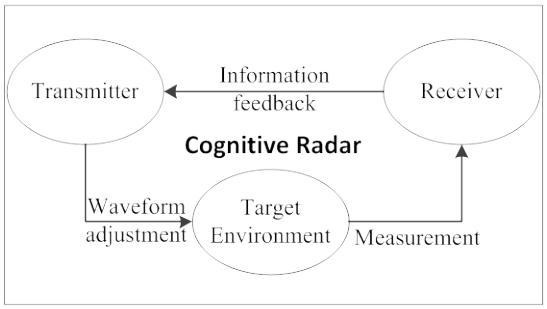


Fig. 4. Perception-action cycle of cognitive radar system as an example cognitive dynamic system

Cognitive radar transmits adaptive waveform according to the feedback information from receiver. Clearly, there is a common structure in all these cognitive systems which is similar to the perception-action cycle of human cognitive system. Consequently, the perception-action cycle acts as a necessary foundation of information feedback during the process of environment sensing.

III. Fundamentals of Cognitive Radar

Cognitive radar can be defined as an intelligent, dynamic, closed-loop system that perceives surrounding environment with the help of priori knowledge and by the means of interaction with environment. Cognitive radar subsequently adjust transmitter and receiver to accommodate the time-varying environment effectively, reliably and steadily.

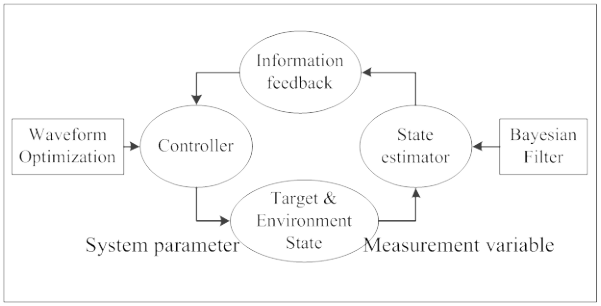


Fig. 5. Information flow of cognitive radar

Cognitive radar differs from classical radar from three aspects, including intelligent information processing, information feedback and knowledge-aided system. Intelligent information processing improves the perception performance by interaction of radar with environment. The information feedback from receiver to transmitter is the foundation of construction. Knowledge-aided system is made of database of environment and targets, as well as the storage of radar echo data, aiming at providing and accumulating more useful priori knowledge[10, 14].

Bayesian state estimator is fundamental for intelligent information processing. State-space model and filter algorithm are the basic parts that have significant effects on the estimation of target and environment.

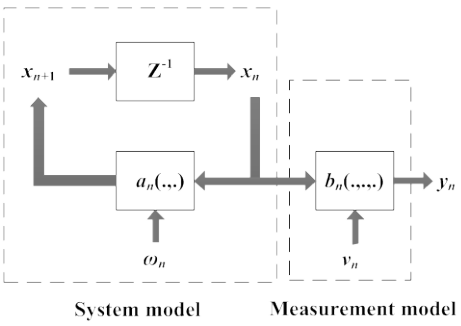


Fig. 6. State-Space model

The state-space model are as follows:

$$\begin{cases} x_{n+1} = a(x_n) + \omega_n \\ y_n = b(x_n) + v_n \end{cases} \tag{1}$$

where x_n is the system state at time n, y_n is the measurement variable, $\boldsymbol{\omega}_n$ and $\mathbf{v}_n$ are system noise and measurement noise that are usually supposed as additive with covariance P and Q.

The prediction of state and covariance at time n from history data:

$$\begin{cases} \hat{x}_{n|n-1} = E\left[a_n\left(x_{n-1}\right)\right] \\ \quad = \int a_n\left(x_{n-1}\right) N\left(x_{n-1}; P_{n-1|n-1}\right) dx_{n-1} \\ P_{n|n-1} = E\left[\left(x_n - \hat{x}_{n|n-1}\right)^2\right] \\ \quad = -\hat{x}^2_{n|n-1} + \int a\left(x_{n-1}\right)^2 N\left(x_{n-1}; P_{n-1|n-1}\right) dx_{n-1} + Q_{n-1} \end{cases} \tag{2}$$

The estimation of state and covariance is shown as

$$\begin{cases} x_{n|n} = \hat{\hat{x}}_{n|n-1} + W_n\left(y_n - y_n\right) \\ P_{n|n} = P_{n|n-1} - W_n P_{yy} W_n^T \end{cases} \tag{3}$$

where W_n is Kalman gain. The accuracy of estimation depends on the state-space model and the calculation of integrations such as $\hat{x}_{n|n-1}$, $P_{n|n-1}$,and W_n. On one hand, many models are developed to indicate and descript the evolution of radar target and environment. On the other hand, much effort are made to improve the performance of filter, including extended Kalman filter by linearizing $\boldsymbol{a}_n(.,.)$ and $\boldsymbol{b}_n(.,.)$, unscented Kalman filter[15] and Cubature Kalman filter[16] by sampling the distribution, and particle filter by Monte carlo sampling in a random way[17]. Compared with the common filter methods, it is proved that CKF works better because of the excellent performance for nonlinear model.

Classical radar employs fixed waveform to perceive target and environment, while cognitive radar doesn't. Waveform can be adjusted in real-time according to the present state. Consequently, radar waveform selection [18] is a dynamic process that dynamic optimization and dynamic programming [19, 20] are feasible to be applied to waveform optimization.

For a tracking radar[2], suppose a LFM pulse waveform is used with chirp rate b and pulse duration λ, it is shown in theory that the measurement noise v depends on b and λ[1], as a consequence covariance $P_{n|n}$ is dependent of b_{n-1} and λ_{n-1} at time n-1, the optimal parameters can be found by minimizing $P_{n|n}(b_{n-1}, \lambda_{n-1})$ in some manner, where $\theta \square \{b, \lambda\}$.

Once y_{n-1} is received, the optimal waveform for next time can be obtained by optimization algorithm, then the waveform is adjusted to the optimal, and next cycle starts.

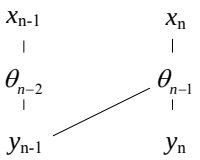


Fig. 7. Information flow for dynamic waveform optimization

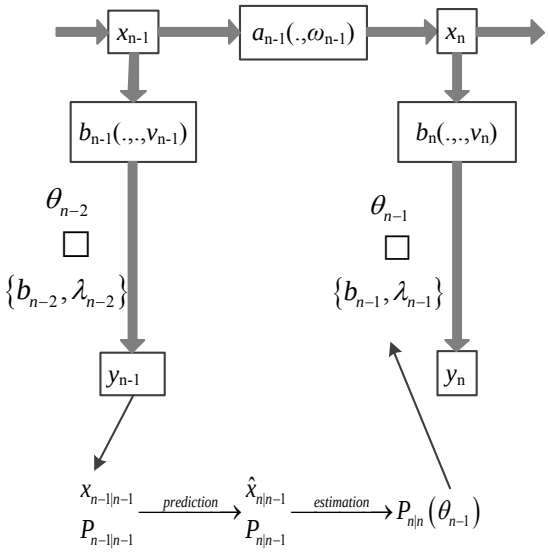


Fig. 8. Optimization of waveform parameter for LFM tracking radar

In fact, the waveform can be optimized in all dimension, such as frequency domain, time domain, directional diagram, polarization, and so on. The process is similar to LFM parameter optimization above.

IV. SUMMARY

In this article, the model of cognitive radar is discussed including the basic theory and the elementary framework. The realization of cognitive radar is dependent of the mature intelligent system and reliable hardware configuration. The most important of the soft aspects contain the waveform optimization in multiple dimension and the knowledge-aided system with database of target and environment as well as expert rules. The goal of scientists is to unify all the nature laws in an universal manner. Cognitive radar is such a goal for radar engineers to develop in order to be applied to any scene.

ACKNOWLEDGMENT

The authors thank all the colleagues for their useful discussion and meaningful help.

REFERENCES

[1] S. Haykin, Y. Xue, and P. Setoodeh, "Cognitive radar: Step toward bridging the gap between neuroscience and engineering," Proceedings of the IEEE, vol. 100, pp. 3102-3130, 2012.

[2] S. Haykin, A. Zia, I. Arasaratnam, and Y. Xue, "Cognitive tracking radar," in Radar Conference, 2010 IEEE, 2010, pp. 1467-1470.

[3] J. Mitola, "Cognitive Radio---An Integrated Agent Architecture for Software Defined Radio," 2000.

[4] J. Mitola and G. Q. Maguire Jr, "Cognitive radio: making software radios more personal," Personal Communications, IEEE, vol. 6, pp. 13-18, 1999.

[5] S. Haykin, M. Fatemi, P. Setoodeh, and Y. Xue, "Cognitive control," Proceedings of the IEEE, vol. 100, pp. 3156-3169, 2012.

[6] S. Haykin, Cognitive dynamic systems: Perception-Action cycle, radar and radio: Cambridge University Press, 2012.

[7] S. Haykin, "Cognitive Dynamic Systems: Radar, Control, and Radio [Point of View]," Proceedings of the IEEE, vol. 100, pp. 2095-2103, 2012.

[8] M. Wicks, "Cognitive radar: A way forward," in Radar Conference (RADAR), 2011 IEEE, 2011, pp. 012-017.

[9] S. Haykin, "Cognitive radar: a way of the future," Signal Processing Magazine, IEEE, vol. 23, pp. 30-40, 2006.

[10] J. R. Guerci and E. J. Baranoski, "Knowledge-aided adaptive radar at DARPA: an overview," Signal Processing Magazine, IEEE, vol. 23, pp. 41-50, 2006.

[11] J. R. Guerci, "Cognitive radar: a knowledge-aided fully adaptive approach," in Radar Conference, 2010 IEEE, 2010, pp. 1365-1370.

[12] S. Haykin, "Cognitive radar," Knowledge based radar detection, tracking and classification, F. Gini and M. Rangaswamy, Eds. John Wiley and sons, pp. 9-30, 2008.

[13] J. M. Fuster, Cortex and mind: Unifying cognition: Oxford university press, 2003.

[14] M. C. Wicks, M. Rangaswamy, R. Adve, and T. Hale, "Space-time adaptive processing: a knowledge-based perspective for airborne radar," Signal Processing Magazine, IEEE, vol. 23, pp. 51-65, 2006.

[15] E. A. Wan and R. Van Der Merwe, "The unscented Kalman filter for nonlinear estimation," in Adaptive Systems for Signal Processing, Communications, and Control Symposium 2000. AS-SPCC. The IEEE 2000, 2000, pp. 153-158.

[16] I. Arasaratnam and S. Haykin, "Cubature kalman filters," Automatic Control, IEEE Transactions on, vol. 54, pp. 1254-1269, 2009.

[17] G. Kitagawa, "Monte Carlo filter and smoother for non-Gaussian nonlinear state space models," Journal of computational and graphical statistics, vol. 5, pp. 1-25, 1996.

[18] S. Haykin, Y. Xue, and T. N. Davidson, "Optimal waveform design for cognitive radar," in Signals, Systems and Computers, 2008 42nd Asilomar Conference on, 2008, pp. 3-7.

[19] D. P. Bertsekas, D. P. Bertsekas, D. P. Bertsekas, and D. P. Bertsekas, Dynamic programming and optimal control vol. 1: Athena Scientific Belmont, MA, 1995.

[20] R. Bellman, "Dynamic programming and Lagrange multipliers," Proceedings of the National Academy of Sciences of the United States of America, vol. 42, p. 767, 1956.

Researching the Key Technologies of Wireless Sensor Network Node in Power Distribution Room Status Monitoring

Hong Lv
Anhui Jianzhu University
Electronic and Information Engineering College
Hefei, China
1030467496@qq.com

Zhixiang Hua
Anhui Jianzhu University
Electronic and Information Engineering College
Hefei, China
1065517780@qq.com

Xinsheng Xia
Anhui Jianzhu University
Electronic and Information Engineering College
Hefei, China
Xiaxinsheng319@163.com

Yonglin Yu
Anhui Jianzhu University
Electronic and Information Engineering College
Hefei, China
405041850@qq.com

***Abstract*—Power distribution room is the key position of the power supply system, it is very important to monitor power distribution room and real time warning in order to ensure the normal power supply. A design of a detection system based on Zig Bee wireless sensor network, the hardware based on CC2530, software development based on Zig Bee. Designing a novel microstrip antenna because of the harmonic interference of power distribution room and serious reflection of equipment. The antenna has the characteristics of good omnidirectional, small echo loss, high gain, small volume etc. Experiments show that it can real time monitor every node that it can effectively transmit and receive data.**

Keywords—power distribution room; WSN; ZigBee; antenna design

I. INTRODUCTION

Power distribution room as a transit point for each family, its safety is related to the normal operation of each home, but also on the lives of users, real-time detection power distribution room safe is very important. Although there are specialized in large construction personnel on duty, and the use of fire, wired surveillance technology in large buildings. Resulting in high annual operating costs, and it can not detect the temperature, humidity, smoke, flooding and other parameters of local environment; nor real-time acquisition and display environmental parameters of every corner in power distribution room. In economically underdeveloped areas, there are no staff in power distribution room, so it is particularly important to monitor safe of power distribution room and effective implementation of environmental safety. Currently, the country will also have to monitor environmental safety WSN technology application in the power distribution room. Power distribution room safety monitoring system based on WSN shows unique advantage comparing with traditional safety monitoring system [4-6]. However, the use of common nodes of WSN has limitations in the power distribution room. For example, (a) transmitting and receiving antennas of terminal device node, coordinator nodes and router nodes do not have the character of omni-direction, (b) the power distribution room equipment radiates electromagnetic waves each band, interferes with the communication among the nodes, (c) the equipment scattering of electromagnetic signal in power distribution room is very large, and so weakens the signal strength. To solve the above problems, this paper presents WSN-based environmental monitoring system, and designs a new type of node antenna. By loading the patch and slotted ring method ,to improve the efficiency of the antenna's radiation and electromagnetic wave respectively inhibitory effect to make the antenna transceiver with high efficiency, to prevent other frequencies band.

II. THE OVERRALL DISTRIBUTION ROOM DETECTION SYSTEM CONSTRUCTION

A. Structure of the System Design

The main power distribution room detection system mainly comprising the wireless sensor network and computers wireless sensor networks are mainly terminal nodes, the coordinator node, routing node, shown in Figure 1. Terminal node is in charge of detecting flooding, temperature, humidity, smoke and other parameters of every corner in the room, then the data is transmitted to the routing nodes, and aggregated transmission router to the coordinator, the coordinator of the monitoring center receives data. Sure to grasp the environmental parameters of every corner in the room in real time; and reference to the relevant provisions of the State promulgated, set alarm value, provide information to the duty officer.

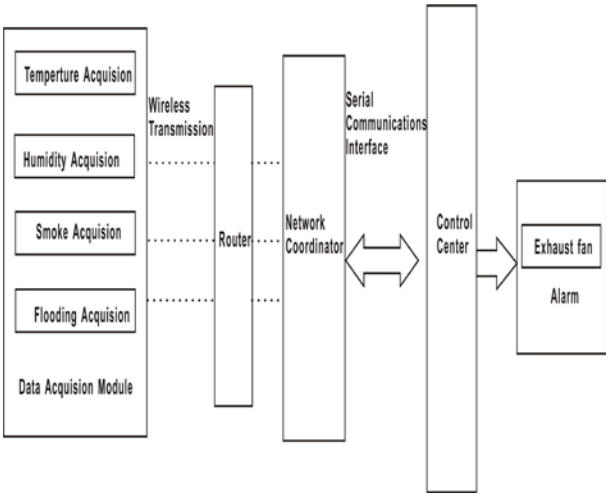


Fig. 1. System structure diagram

B. *System topologies*

Zig Bee network supports star, tree and mesh ,three kinds of network topology ,shown in Figure2, respectively, followed by a star network, tree (cluster) form networks and mesh networks.

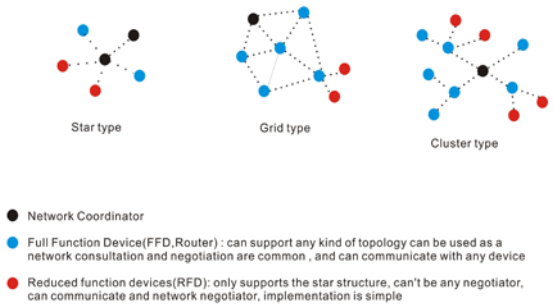


Fig. 2. Topology structure

Star network consists of a PAN coordinator and a large number of end devices, there is only communication with the terminal of the PAN coordinator, communication among terminal devices are required to be forwarded by the PAN coordinator.

Tree network consists of a coordinator and one or more connected in a star configuration, equipment in addition to with their parent or child nodes communicate directly point to point, the other can only be done through the tree routing messaging.

Mesh networks are implemented on the basis of a tree network with a network different from the tree is that it allows the network to all nodes in a direct interconnection with a routing function, implemented by the router routing messages mesh routing table. The advantage of this topology is to reduce the message latency, enhanced reliability, the disadvantage is the need for more storage space overhead [8].

It shows that the network is a mesh network to increase the reliability and reducing latency for communication between the devices in the network structure by comparison, the routing is variable. Routing decisions based on cost of routing device communication , when there is a chain road disconnect yes, you can also choose other routes, because routers can also communicate with each other , increases the coverage and reliability of information transmission network, it can well avoid missing data or missing. This system is designed to detect power distribution room temperature and require a long-term testing and real-time early warning capability, the entire wireless sensor network has multiple nodes, so it uses a reliable and strong, smaller delay mesh network.

III. HARDWARDWARE DESIGN

A. *Detection node hardware design*

Monitoring the overall structure of the nodes as shown in Figure 3. Seen from the figure, the monitoring nodes consists of three parts, the first part of the sensor is mainly responsible for the acquisition of signal data. The second part CC2530 module is mainly responsible for a data processing and control of the microprocessor, memory, and RF modules. The third part is the power supply module, it is responsible for the energy supply of the entire node. Follow Zig Bee protocol monitoring nodes will be collected to get the data and sent to the gateway node via RF module, enabling to monitor environmental parameters of power distribution room.

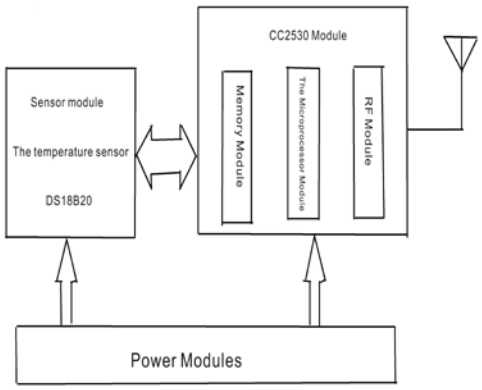


Fig. 3. Overall structure of monitoring nodes

B. *Design of Microstrip Antenna*

Inverted F antenna is a kind of small size, simple structure, easy to match the monopole antenna, it is suitable for short distance wireless communication, radiation of inverted F antenna contains both horizontal polarization component and vertical polarization component, so it is

used in indoor equipment particularly, because indoor wall, decorations, indoor equipment scattering can cause electric field between the horizontal polarization and vertical polarization transformation, the polarization of inverted F antenna can effectively increase the reception .We choose to design a inverted F antenna that it has the center frequency of 2.4GHz and bandwith of 300 MHz.

1) Design of inverted F antenna

The design of the antenna substrate material is FR4 board general, the substrate thickness is 0.8mm, a dielectric constant is 4.4 and the center frequency is 2.4 GHz. Antenna size is calculated as:

$$W = \frac{c}{2 f_r}\left(\frac{\varepsilon_r + 1}{2}\right)^{-1/2} \quad (1)$$

$$\Delta l = 0.415h\left(\frac{\varepsilon_r + 0.3}{\varepsilon_e - 0.258}\right)\frac{\frac{w}{h} + 0.264}{\frac{w}{h} + 0.8} \quad (2)$$

$$\varepsilon_e = \frac{\varepsilon_r + 1}{2} + \frac{\varepsilon_r - 1}{2}\left(1 + \frac{10h}{w}\right)^{-1/2} \quad (3)$$

$$L = \frac{c}{2 f_r \sqrt{\varepsilon_e}} - 2\Delta l \quad (4)$$

where in, L is the length of the patch, w is the width of the patch, ε_e is relative dielectric constant, ε_r is the dielectric constant, Δl is the length of the gap, f_r is the center frequency of the antenna.

Through theoretical analysis and the above (1), (2), (3), (4), deduced inverted-F antenna structure shown in Figure 4, the entire antenna structure can be divided into three parts, namely the shape of an inverted-F antenna , dielectric layer and a ground plane, a ground plate located under the surface of the dielectric layer, a length of 90mm and 50mm. Shaped inverted F antenna located on the surface of the dielectric layer, the resonator length L is 16.2mm, the antenna height is 3.8mm, the distance grounding point and the feed point is 5mm, the width of the microstrip line is 1mm.

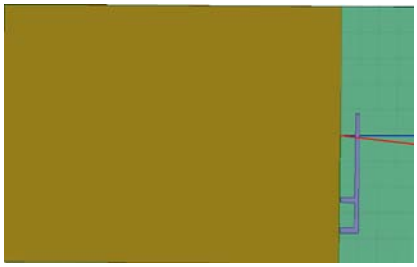


Fig. 4. Inverted F antenna model

2) Inverted-F antenna analysis

Through the simulation analysis, we can see echo loss from figure 5, the resonant frequency of the antenna is 2.4GHz, the center frequency is -24.24 ^ 5, bandwidth of 10 dB is about 440MHz, covering the wireless sensor network nodes working 2.4GHz-2.484GHz frequency range.

Figure 6 shows that H-plane antenna radiation pattern and normalized E-plane radiation pattern normalized in the 2.4GHz frequency can be seen from the figure H-plane radiation approximate omnidirectional and symmetry in the 2.4GHz frequency; shape of E-plane radiation is “8”, which is similar to the radiation field dipole radiation field, indicating that the omnidirectional antenna can send and receive electromagnetic signals.

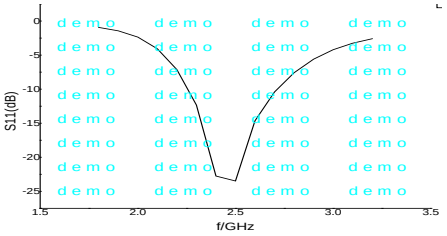


Fig. 5. Inverted-F antenna S11

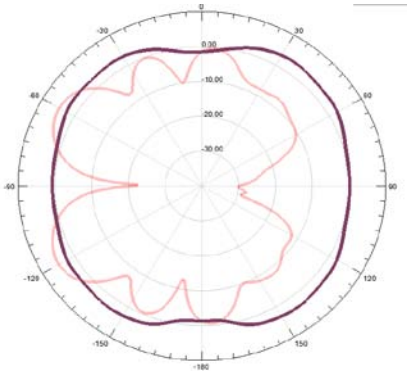


Fig. 6. Inverted-F antenna pattern

Design of omnidirectional inverted F antenna after repeated simulation get preferable inverted F antenna, omni-directional antenna is cross polarization characteristics, so it is suitable for complex equipment environment of transformer substation.

C. Coordinator node hardware design

The main function of coordinator is to use wireless communication module terminal to receive the sensor nodes to collect signals, then through the existing serial interface to upload first machine, realize the storage and display of the collected data, etc. The hardware module includes CC2530 wireless receiver module circuit design, power supply design and serial transmit module. Coordination entire CC2530 chip peripheral circuit design shown in Figure 7. 25 -pin and 26 -pin used to connect the input of the receiving antenna, connect with the CC2530 module through these two key; 15 -pin serial port module as a data receiving terminal Rx and 16 pins as data the sender TX.

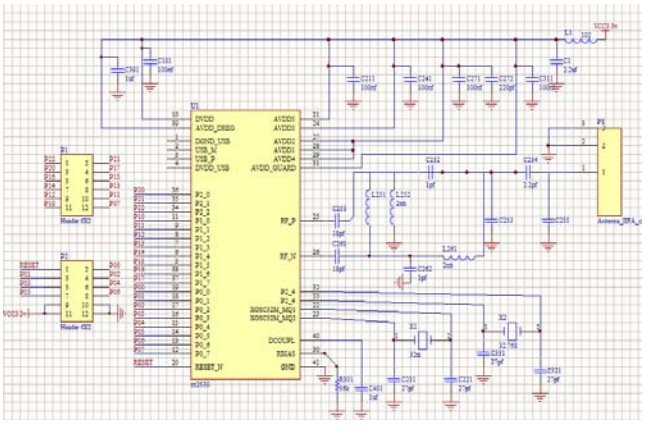


Fig. 7. Protocol of hardware design

IV. SOFTERWARE MODULE DESIGN

A. Software Design software module coordinator

The coordinator role mainly involves network startup and configuration. Once these are completed, the coordinator works like a router. When the coordinator is energized, the system initialization, the coordinator will scan the DEFAULT-CHANLIST specified channel. After the establishment of a network node will have to wait to monitor the network and to join the network node discovery. Along with LED displays to find out whether the network has been established to join the network, the coordinator will automatically assign addresses to join the network, to achieve data transmission. The process is shown in Figure 6.

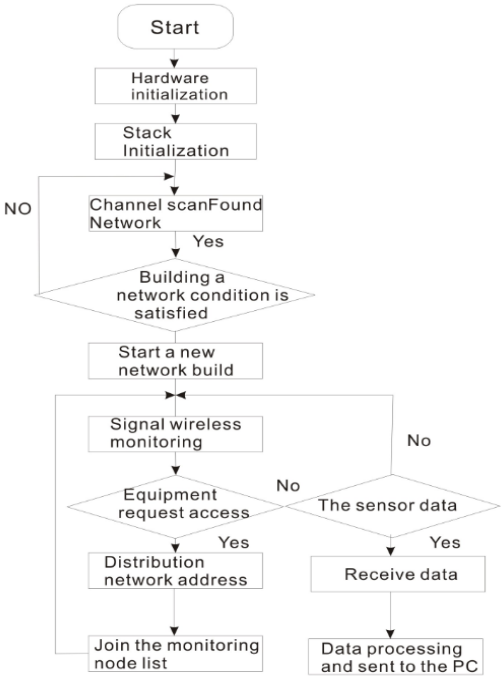


Fig. 8. Ordinator node workflow

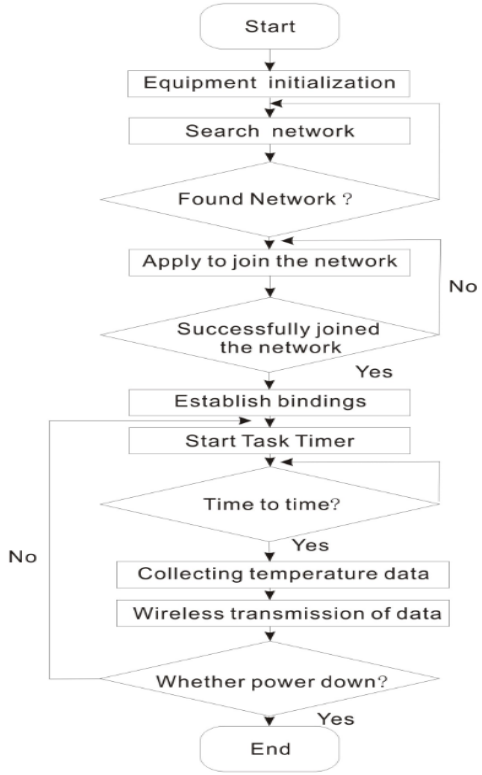


Fig. 9. Terminal nodes working process

B. Signal Acquisition Module Software Design

After the sensor nodes to electricity will be configured to ZigBee terminal node equipment, initialized after power on the node, and then search the network after power on and request to join the network coordinator established. Sensor nodes and coordinate point to establish a binding relationship after node successfully joins to establish network, the LCD on CC2530 module displays the success to join .After binding, sensor nodes transmit the collected temperature information and other relevant data to the coordinator node in the specified time. Its collection of information flow shown in Figure 7.

V. EXPERIMENTAL RESULTS AND ANALYSIS

In order to verify the performance of the system design, to select a power distribution room, the data acquisition nodes are placed at the bottom of the distribution cabinet, and the back of the power distribution cabinet, corners that people can not properly view of tight, testing the effect of design system to collect environment parameter. Nodes in common market replace by design nodes each other to test data. Figure 8 shows designed system in the duty room detected temperature data outside 10 meters below the range of distribution cabinet in the power distribution room; but ordinary node can not display any data.

Fig. 10. Displays data detection

VI. CONCLUSION

Research on the detection system based on WSN distribution room, several gas sensor systems use real-time monitoring of power distribution room, use high reliability mesh network to realize the wireless transmission of data. According to the actual environment of substations, to design a new antenna, the antenna has good characters of high efficiency, omnidirectional in the 2.4 GHz transceiver, to control the equipment of electromagnetic wave scattering; The inhibition technology reduce electromagnetic wave antenna interference on nodes. This system can be used by large construction and some unmanned substations, provide a reliable guarantee for the real-time detection of power distribution room duty status.

ACKNOWLEDGMENT

Foundation: National Natural Science Foundation (61071001), National Natural Science Foundation (61372094)

REFERENCES

[1] Jian Ming, Shen, Hong Liwei. Circuit Design of Panel Points Based on Zigbee[J]. Lecture Notes in Electrical Engineering.2013, 3-9.

[2] He Zhengli, Ahaowang da.Design of power distribution room temperature and humidity monitoring system [J] Computer Measurement & Control, 2008.16 (7): 980-982.

[3] Liujuan, Ju Dengfeng, Wang Yong, Lu Liping, Zhang Qiang. Zou Danping, The design and application of intelligent substation equipment condition monitoring system [J], power construction, 2012, 33 (6): 6-10.

[4] ZHENG K, ZHAO H W, ZHANG X L. Research on monitor system of electrocardiograph based on Zig Bee network[J].Chinese Journal of Scientific Instrument, 2008, 28(9); 1908-1911.

[5] Gao Qingming, He Huan, Shi Ruijie. ZigBee-based wireless sensor network applications in substation monitoring system [J] North China Water Conservancy and Hydropower College 2010, 31 (1): 53-55 .

[6] Wheeler Andrew.Commercial applications of wireless sensor networks using ZigBee [J]. IEEE Communications Magazine, 2007, 45(4):70-77.

[7] Wangli, Li Chunling. Indoor air quality testing and sensor applications [J]. Heilongjiang Science and Technology Information, 2012, 33 (8): 75-7.

LEO-User-Oriented Space Integrated Information Network

ShichaoWang, BinWu, BoWang
Beijing Institute of Tracking and Telecommunication Technology
Beijing, China
wangsc8896@126.com

***Abstract*—A new structure of space integrated information network is proposed in this paper, in which the LEOs are users to the backbone network, rather than parts of it. The structure is named as LEO-user-oriented space network. This paper first introduces traditional concept of space information network, then discusses the new concept of the LEO-user-oriented architecture; next probes into some key techniques related to the space information system, including laser links, routing and access technology. At last feasibility of the walker constellation is proved though simulations.**

Keywords—space information network; LEO user oriented; laser link; routing; access algorithm

I. INTRODUCTION

Chinese satellites on the orbit sum up to more than 200[1], forming a comprehensive space equipment system which includes a vast number of various facilities, examples are navigation satellites, communication satellites, reconnaissance satellites, space station, deep space exploration spacecraft, etc.. But on the whole, this system is a composition of isolated spacecraft instead of an integral whole, for there are limited inter-satellite links (ISL) between them. With the development of space technology, it is corporation and synergy between constellation, rather than increase of the number of satellites, that can meet the application requirements of spacecraft. As a consequence, integrate development of space information acquisition satellites, space information transmission satellites and PNT (Positioning, Navigation and Timing) satellites is an inexorable trend. In this context, the concept of space integrated information network comes into being.

Space integrated information network has become popular research focus in recent years. Traditionally, nodes in space information network include multilayer satellites, airships, spaceflights, aero crafts, and UAVs (Unmanned Aerial Vehicles) in near space. The capabilities that space integrated information network can provide are not only connecting and communicating with spacecraft and grand stations, but also acquiring, processing, and distributing information on orbit. Self-healing and self-organization are also important characteristics that space integrated information network should have. In this network architecture, LEO, MEO and GEO constitute the backbone of the network. But in fact, there are such a large number of LEOs that it makes the backbone network very complicated. On one hand, more than one links are needed by LEOs to form the network, which is a high cost, on the other hand, the difficulty of designing routing algorithm increases. To solve these problems, this article proposes a new concept of space information network named LEO-user-oriented architecture, of which the backbone nodes include only MEOs and GEOs, all the LEOs are users of space information network, rather than parts of it. In this sense, LEOs just need to connect to the backbone network when necessary, which can reduce the cost of satellite links and make the routing algorithm of backbone network easier.

The reminder of this article is organized as follows: Section 2 introduces basic concept, constituents, and characters of traditional space information network. Section 3 proposes the new concept of space network architecture: LEO-user-oriented architecture. Section 4 presents some key techniques of the new space architecture. Section 5 gives simulation result. Section 6 concludes the paper.

II. SPACEE INTEGRATED INFORMATION NETWORK

At present there's no unifying definition about space integrated information network, it is generally accepted that space information network is an integrated information network that connects different types of satellites on different orbits and ground facilities by ISLs and satellite-ground links, on the rule of maximum efficiency. The network has abilities to acquire, store, process, and transmit information intelligently, and abilities of self-operation and self-management [2]. Structure of space integrated information network is showed in Fig.1. The functions of each part are as follows [3]:

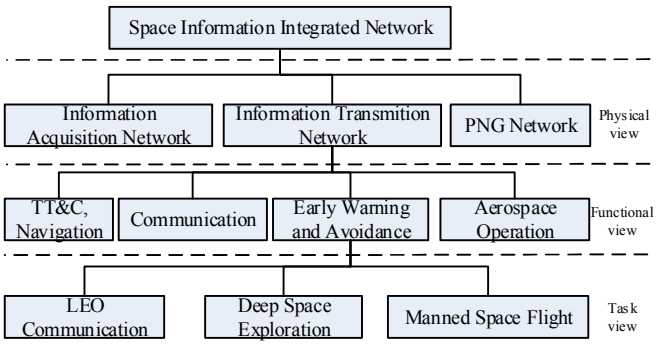


Fig. 1. Different views of space network

a) Space-based Information Acquisition Network is responsible for collecting information, for example weather information and investigation information. This part mainly

consists of early-warning satellites, earth resource satellites, meteorological satellite, space surveillance satellite and reconnaissance satellite.

b) Space-based Information Transmission Network is responsible for transmitting, delaying, and distributing information. A good example is GPS which can transmit nuclear exploration monitoring information without ground station. Another example is tracking and data relaying satellite system (TDRSS) although it is not networked obviously, for there's no ISL between TDRSS satellites.

c) Space-based PNT Network provides position, navigation and time service for stable and movable users. GPS and Chinese Beidou system are of this kind.

In addition, space integrated information network also has the function of TT&C, network management, and information pretreatment. Physical and functional structure views of the network are showed in figure 1.

Space integrated information network is spacecraft-based and various-service-information-integrated, it is a network that fuses different types of networks, and the network has its particular characters like: [4, 5, 6]

a) Open Network: Space integrated information network is open and compatible with multi satellite systems. In order to achieve global information sharing, different networks are interconnected, forming a heterogeneous network.

b) Information processing integration: space information network must process information compositively and transmit information rapidly so as to realize multi-network interconnection.

c) Stereo information exchange: Space integrated information network consists of space-based, grand-based and sea-based facilities. The stereo topology leads to stereo information exchange, which increases transmitting efficiency and communication resource utilization.

d) Dynamic networking mechanism: Space information network has ability to adjust and recombine some nodes dynamically when necessary, and ability of self-origination.

III. LEOs-Oriented Space Information Network

A. Concept

To design the space integrated information network architecture is a difficult work. Some researches proposed a terrestrial-user-oriented architecture, which considers the terrestrial and airborne users as the major component of the users (as shown in Fig. 2). In this architecture, low earth orbit satellites (LEOs) are part of the space information network rather than users to it, and there are ISLs between LEOs. In this article, a new LEO-user-oriented space network architecture is proposed. In this network architecture, LEOs like reconnaissance satellites, meteorological satellites, etc., are users, as well as some satellites located in Medium Earth Orbit (MEO) and geosynchronous orbit (GEO), which is described in Fig.3. There's no ISL between LEOs. LEO accesses to backbone network as a user, and transmits information to destination through the backbone network.

The backbone network refers to the part which is responsible for information transmission with high data rate and high capacity for users. In the physical view of space network as shown in Fig.1, space transmission network can be seen as backbone network. The design of the LEOs-oriented space network architecture is very different from the terrestrial-user-oriented one, which has been optimized to serve only terrestrial and airborne users. Major differences include: (1) different data types (e.g., reconnaissance information, telemetry and telecontrol information) (2) different QoS requirements (e.g., latency, throughput) (3) location of users (e.g., users in LEO, MEO, GEO) (4) status of users (i.e., the relative velocity is high). [7]

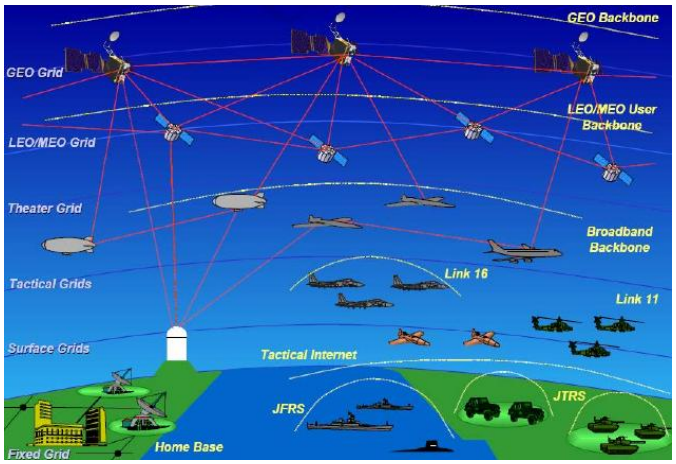


Fig. 2. Terrestrial-user-oriented Space Network Architecture

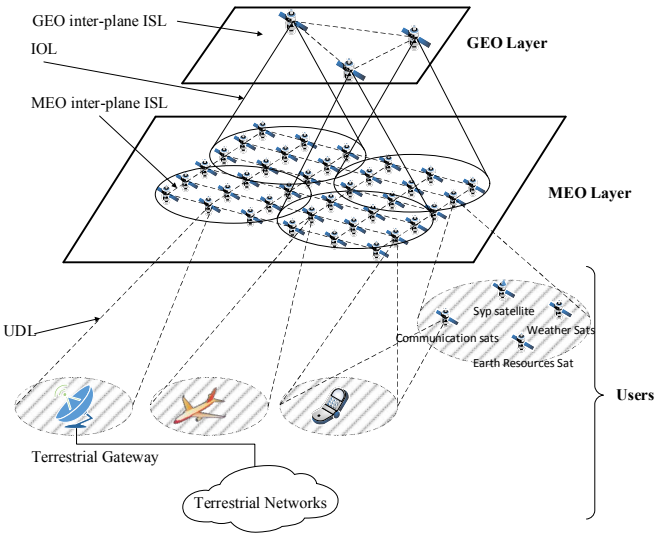


Fig. 3. LEO-user-oriented Space Network Architecture

B. Benifits

The reasons why LEOs are considered as users are:

(1) Simplify backbone network of the space information network by reducing the number of its nodes, which can contribute to simplifying of the design of routing algorithm.

(2) Improve expandability of the space network. Various types and large number of LEOs can be connected to the backbone network.

(3) Enhance flexibility of the space network. In the terrestrial-user-oriented architecture, topology of LEOs and other nodes of the backbone network is relatively stable (which means it changes regularly). How to interconnect LEOs with other nodes must be considered carefully. While in the LEOs-oriented space network architecture, there's no such problems, LEOs just consider how to access to the nodes of backbone network when they need to transmit information.

(4) Reduce the cost. In the terrestrial-user-oriented architecture, LEOs needs to have more than 1 link as a node of backbone network. But LEOs in the LEOs-oriented architecture just need one link to connect to backbone network. LEOs are users and served by the backbone space network, so there's no need to interconnect users, which can reduce the cost.

IV. KEY TECHNOLOGIES ABOUT LEOS-ORIENTED SPACE NETWORK

A. Laser Links

Compared with the traditional satellite microwave communication, the laser link has many advantages for the high frequency (300MHz ~ 300GHz) and the small beam divergence angle (only about 10 micro arc). The advantages include: high data rate ($\geq$ 10Gbits/s), high communication capacity, good confidentiality, excellent anti-interference performance, and the terminal equipment has the advantages of small volume, weight, and power consumption. As a result of these advantages, laser communication can be used to construct the backbone of space information network, forming a space information superhighway [8].

Different from the traditional satellite microwave communication, the characters of small beam divergence angle and long link distance make it difficult to aim at the communication target. This requires the pointing acquisition tracking (PAT) subsystem to achieve aiming at and tracking each other accurately and stably. How to realize high capacity, reliability and data rate of the communication link under condition of vibration of the satellite platform, relative movement between satellites, and various mechanical and electronic noise, has been a complicated research focus.

B. Routing

In the space-based information network, routing implementation is one of the basics for the normal operation of network. Topology of space network changes frequently, which makes it different from terrestrial internet. Currently, the solution of space network's routing is to store the routing table on satellites, which is pre-computed on the ground. When the topology changes, all satellites switch to new corresponding routing table. This solution places restrictions on robustness and flexibility. If one node is broken, the network cannot adopt to the changes flexibly, the whole network would be affected.

Several researches has been carried on about routing mechanism, the existing routing algorithm includes DRA (Datagram Routing Algorithm), PRP (Probabilistic Routing Protocol) [9, 10, 11]. But all the researches focus on LEO or LEO/MEO satellite network, a typical constellation is DLSN (A Double-layered Satellite Network). There's few researches focused on routing algorithm that applied to MEO/GEO satellite network, which is the main part of the space integrated information network.

C. Access Technology

With the new concept of LEOs-oriented space information network, there needs to be a connecting algorithm between LEO users and backbone networks. Different from traditional access technology of the internet, backbone of the space information network has characteristics of limited communication resources, changing topology and complicated geometric relationship. These make access technology of space information network facing with new problems. For example, a reconnaissance satellite needs to transmit a photo to ground station on the other end of earth using backbone network, it can be seen by n nodes of backbone simultaneously, how to choose the best one to access to backbone network, achieving maximum throughput, highest efficiency, and satisfying LEO users QoS requirements, is the key problem that the access algorithm need to resolve.

In fact, access technology has been realized in terrestrial networks namely IP network, cellular network and global communication system (e.g. Iridium). But access mechanism in space network is different. In global cellular network, a user is generally covered by only one base station, which makes it easy to design access algorithm. Even in junction of signals of several base stations, the user just need to choose the strongest one to access to. While in the space network, the node whose signal is strongest may not be the best one owing to many factors (e.g. relative position and velocity relationship, length of time when the user can communicate with backbone nodes, congestion, etc.). Differences of access technology between four networks are shown in Table I.

TABLE I. DIFFERENCES OF ACCESS TECHNOLOGY BETWEEN THE FOUR NETWORKS

	IP Network	**Cellular Network**	**Iridium**	**Space Information Network**
Primary Users	Computer	Mobile Terminal	Mobile Terminal	**Spacecraft located in LEO**
Velocity	0	Low	High	**Highest**
Access	Stable	Involve the selecting algorithm only in junction of signals of different base station	Involve the selecting algorithm only in junction of different beam coverage	**User can be seen by more than one backbone nodes, involve the selecting algorithm all the time**
Link Switch	None	Infrequent	Infrequent	**Frequent**

V. SIMULATION

In order to realize communicating with LEOs anywhere and anytime, the backbone of space information network must have two capabilities: 1) at least one of the backbone nodes is in sight of the domestic ground stations at any time; 2) the backbone network can cover any LEOs at any time. In this section, feasibility of the walker constellation is proved though simulation according to the two requirements above, which is used in Beidou system with 3 planes and 8 satellites in each.

From the perspective of simplification, we use only two planes of the walker constellation. Simulation parameters are: 7 satellites in each plane, semi major axis of the orbit is 24000km, eccentricity is 0, inclination of the orbit is 0, and simulation time is 24 hours. We suppose that sensor type on the satellite is simple conic and cone angle is 15 deg.

Fig. 4, 5, 6 gives percentage coverage for the LEOs whose orbit is no more than 120km. We can conclude that it needs at least 2 planes and 7 satellites in each to acquire global coverage.

Fig 7 gives simulation results about access to the constellation for ground stations, include XiAn, QingDao, KaShi, JiaMuSi, and HaiNan. We conclude that no station can realize reaching the constellation all the time by itself, but a network composed of the 5 stations above can do.

VI. CONCLUSION

In this article, we proposed a LEO-user-oriented space network architecture. In this architecture LEOs are users to the backbone, which is different from the one that LEOs are parts of the backbone. The new architecture simplifies the structure of backbone network and improves the flexibility and extendibility of it. This article also probes into several key techniques of the space integrated network. At last, feasibility of the walker constellation being as the backbone network is proved though simulations.

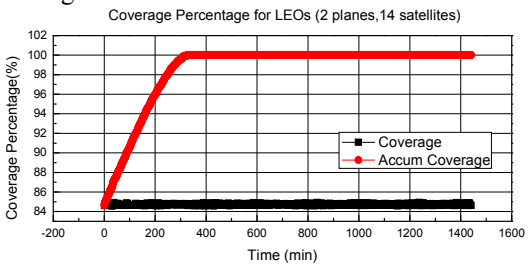


Fig. 4. Coverage Percentage for LEOs (1 planes,7satellites)

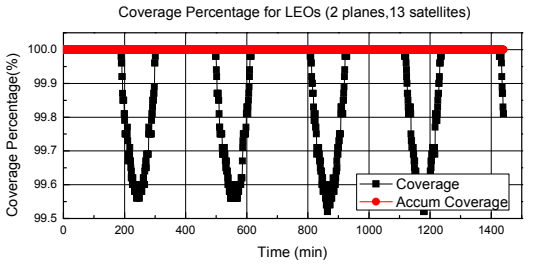


Fig. 5. Coverage Percentage for LEOs (2 planes,13satellites)

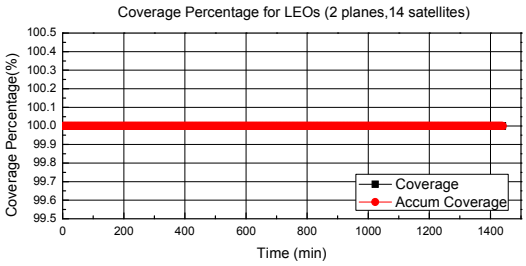


Fig. 6. Coverage Percentage for LEOs (2 planes,14satellites)

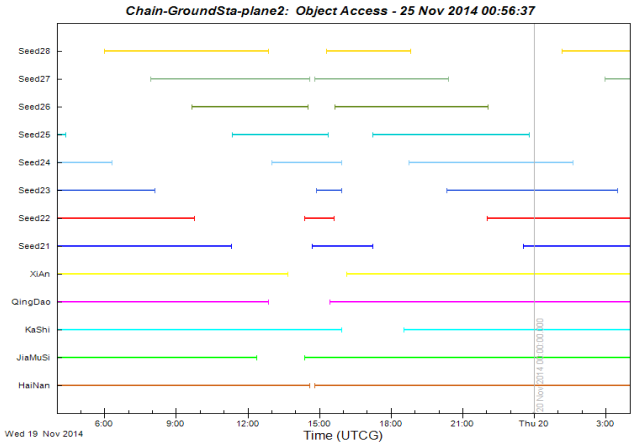


Fig. 7. Access to the backbone

REFERENCES

[1] LaunchHistoryView. [EB/OL]. [2014-11-10]. http://satellitedebris.net/-Database/LaunchHistoryView.php?hname=LaunchHistoryViewDetailEdit0LaunchHistoryView_handler&fk0=PRC

[2] MIN Shiquan, “A survey on space integrated information network”, [J], International Space, 2013.8

[3] MIN Shiquan, “An Idea of China's Space-based Integrated Information Network”[J], Spacecraft Engineering, 2013-05

[4] WEI Bing- chao, TU Hai-yang, “Application and Key Technologies of Space-based Information Transmission and Dissemination System” [J], Computer & Network, 2012, 38(11)

[5] WANG Shi-qiang, HOU Yan, “Requirement Analysis on Air Space Information Transmitted Systems” [J], Ordnance Industry Automation, Vol. 28, No. 12

[6] Guan Qingbo, Feng Shuxing, “Research on Mode of Space-based Intelligence Service” [J], Journal of Academy of Equipment, Vol. 23, No. 6

[7] Chan. V W S, Chan S, Division. M I O T E S.“ Architectures for a space-based information network with shared on-orbit processing”[J]. Massachusetts Institute of Technology, 2005.

[8] Li Xin, “Optimization Research on Link and Communication Performance For Intersatellite Laser Communications” [D], Harbin, Harbin Institute of Technology, 2013, P2-3

[9] Ekici E, Akyildiz I F, Bender M D. A multicast routing algorithm for LEO satellite IP networks[J]. Networking, IEEE/ACM Transactions on, 2002, 10(2):183 - 192.

[10] Yue G, Ekici E, Akyildiz I F. A new multicast routing algorithm in hierarchical satellite networks[C]. //Global Telecommunications Conference, 2002. GLOBECOM '02. IEEE. IEEE, 2002:2925 - 2929.

[11] Akyildiz I F, Ekici E, Bender M D. MLSR: a novel routing algorithm for multilayered satellite IP networks[J]. Networking, IEEE/ACM Transactions on, 2002, 10(3):411 – 424

Research on Outdoor Solar Cell Distributed Monitoring with ZigBee Wireless Sensor Network: Algorithms and Application

Zheng Tao, Chen Yan-Guang
College of Economics and Management
Yan Shan University
Qinhuangdao, Hebei, China
{ztao,cyg}@ysu.edu.cn

Li Meng-zhu, Yang Yi, Zhou Hong-wei, liu Xu-yang, Yao Jin-kun
College of Information Science & Technology
Donghua University
Songjiang, Shanghai, China
yiyang@dhu.edu.cn

***Abstract*—I-V curve and MPP are key parameters for solar cell performance measuring. A low-cost testing system based on Zigbee for outdoor solar cell outdoor measuring is presented in this paper. The testing algorithm combining constant scale and incremental conductance algorithm is discussed. Interpolation method for I-V curve reconstruction and its error source are analyzed. This paper also reports the specific structure of our outdoor solar cell testing system. Using the developed algorithm and test system, we present a detailed numerical investigation and experiments about distributed outdoor solar cell measurement. The results show that this method can provide high-precision solar cell characters measuring with low complexity.**

Keywords- photovoltaic power system; I-V curve measurement; maximum power point; wireless sensor networks; ZigBee

I. INTRODUCTION

The needs for green energy boost the development of solar cell. Solar cells are developed to be more efficient and cheaper [1-3]. But solar cells are easily affected in a harsh outdoor environment while temperature, cloudy and dewing can 'weak' them [4-7]. Given the premise, it shows necessary to monitor the operational efficiency of solar cells in the outdoor environment. The results will not only provide a reference for end system design, but also help improve solar cells production. Most previous solar cell outdoor testing methods are with high complexity, high cost and the number of cell under testing is limited. With the development of embedded system and wireless industrial sensor networks, distributed measuring method can be introduced for solar cell outdoor testing [8, 9]. It simplifies the construction of test sites and reduces the total cost of the system. In this paper, a solar cell measurement system is presented. Meanwhile the testing algorithm and specific system design are discussed. ZigBee provides high flexibility in distributed outdoor measuring. This system could be a competitive candidate for outdoor distributed PV system measuring in the future.

II. TESTING ALGORITHMS

In order to measure the operating characteristic of solar cells, we should collect data for I-V curve under different condition and measure the maximum power point (MPP) which is the product of the maximum cell current and voltage. We can get fill factor, transfer efficiency, shunt resistance, series resistance and other characteristic parameters through the I-V curve and MPP. It is shown that the MPP has direct significance for practical PV use. In actual implementations for the efficiency testing of PV system, we need to get envelop of MPP changing by using maximum power point tracking (MPPT). Under different condition and faced with different solar cells, the system should be applicable and reliable. Since the variable resistor matrix method has good stability, it can be employed for IV curve scanning. The load current of solar cell can be expressed by the following formula [10]:

$$I = I_{ph} - I_0\left\{\exp\left[\frac{q(U+IR_s)}{AkT}\right]-1\right\}-\frac{U_D}{R_{sh}} \tag{1}$$

In the formula above, I_{ph} means photo-generated current which is closely related to environment properties ,sunshine intensity for example. I_0 means reverse saturation current. q is electron charge. U_D is the equivalent voltage of the diode. A is a coefficient which is related to the material characteristic of the PN junction in solar cells. k means boltzmann constant. T means absolute temperature. R_s means series resistance. R_{sh} is bypass resistance. It can be shown in formula (1) that the operating characteristic of solar cells is not only enslaved to environment, but also related to manufacturing process and material. The power on the load of solar cells is:

$$P = UI_{ph} - UI_0\left\{\exp\left[\frac{q(U+IR_s)}{AkT}\right]-1\right\} \tag{2}$$

According to the extremum condition, when the system works at MPP, $\partial P/\partial U = 0$

$$I_0\frac{qU_m}{AkT}\exp\left(\frac{qU_m}{AkT}\right) = I_{ph} - I_0\left\{\exp\left(\frac{qU_m}{AkT}\right)-1\right\} \tag{3}$$

It can be shown from (1) to (3) that we can get the corresponding characteristic parameters of solar cells by through the I-V curve and MPP. However, the I-V curve and MPP show random variation in different outdoor conditions. So it is important to scale solar cell working capability in different nature environment by field testing.

Solar cell field testing is easily affected by environment and weather condition. For example, the sun radiation may be affected by cloudy and sunlight incident angle changes in different time of a day. The long-term variation in outdoor environment can be described by the product of the random function and the curve of sun radiation, which can be expressed as follows in a certain period of time: $S=S0\times R(\theta,\zeta)$. θ means incident angle and ζ means sun radiation absorbed by atmosphere. If the data collecting of solar cells takes little time, the randomness caused by environmental variation will lead little impact to the I-V curve scanning with resistor matrix. But this variation can't be neglected in long-term field testing. Solar cell field testing needs to calculate the long-period working efficiency parameters of solar cells in a certain area.

III. SOLAR CELL MEASURING

In order to achieve the large-amount and long-period monitoring about solar cells, we should focus on rapid-convergent data collection and multi-node data access. The parameters of outdoor environment show variation, as a result, the algorithm of rapid-convergent data collection makes contributions to the reduction of errors caused by environment change. If multi-node data access can be achieved, we can make it possible to collect a large amount of data so that the result tends to be accurate. The environment factors should be measured too. A threshold triggering method is employed for data collection. Also, we should measure them. If they change a lot, we have to delete the monitoring data. The judge should base on a suitable triggering threshold. Besides, MPP may deviate a lot and show series of discontinuity because of weather conditions and long-time intervals while measuring, so convergence algorithm should be reset at each interval.

In our algorithm, scaled constant and resistance matrix scanning are important for I-V curve measuring. Since under the same irradiation intensity, MPP is not sensitive to environmental change and related to scaled open circuit voltage. We can estimate the voltage at MPP when we have got the open circuit voltage. It can be shown in formula (3) that the incremental admittance of solar cell is equal to the negative admittance beside MPP. According to the method, a negative scanning step in I-V curve is adopted for the starting point. Then gradually increase the value of resistance to get the best value of MPP. We can set the step value depend on the permutation and combination of each component of resistance matrix to achieve the successive approximation of scanning interval, which can not only avoid disturbance oscillation, but also avoid complex fuzzy algorithm and improve scanning speed. In our system, we scan the I-V curve by resistance matrix with MOFET, determine MPP point, and then optimize the result with spline interpolation. This method has many benefits, including avoiding baseline shifting caused by aging in outdoor environment, more accurate measuring of open circuit voltage, wider range of accurate I-V curve scanning, more accurate resistance test and fewer computational steps to improve the ability to adapt to environment change. Boundary errors and type errors can both affect the reconstruction of I-V curve with cubic spline function. The influence from boundary errors can be shown as:

$$\begin{cases} |\delta_a(x)| \le \frac{4}{27}\left(\frac{1}{2}\right)^{1+i} h_i h_0 |\eta_0| \\ |\delta_b(x)| \le \frac{4}{27}\left(\frac{1}{2}\right)^{n-i} h_i h_{n-1} |\eta_n| \end{cases} \tag{4}$$

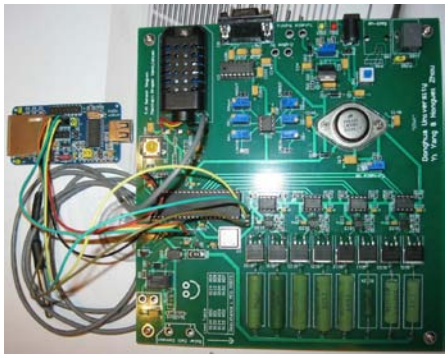
It can be shown in figure (4) that if the sampling errors (η_0, η_1) at the circuit voltage and the short circuit current can be restrained, the influence from boundary errors will be greatly reduced. Since the open circuit voltage is of stable transient response when sampling, we can get it with multiple methods. Since the voltage is zero at the short circuit, we can assign 0 to the spline function at that point and focus on the corresponding step size hn-1 to reduce errors, which is one of the benefits to use cubic spline interpolation function method. The type value error in each sampling data is:

$$|\delta_c| \le \begin{cases} |\varepsilon_k|\left(1-\frac{10}{9}\cdot\frac{h_k}{\overline{h_k}}\right) \\ |\varepsilon_k|\left(1+\frac{10}{9}\cdot\frac{h_k}{\overline{h_k}}\right) \\ |\varepsilon_k|\frac{20}{27}\Psi\frac{h_i}{\overline{h_k}} \end{cases} \tag{5}$$

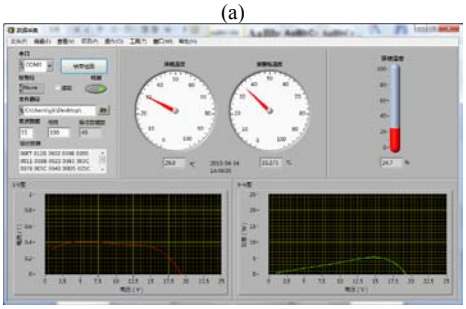
It can be shown from the formula above that we can improve the accuracy of curve resumption and reduce errors effectively. However, during the measurements, smaller step needs longer scanning time. If environment parameters change quickly, extra errors will be introduced into measurements even make the results useless. The shape of the I-V curve is usually organic and most parts of the curve are smooth. MPP appears at its inflection point. If boundary errors are limited well, the extreme value of errors of cubic spline interpolation appears beside the inflection point which changes greatly. As a result, it is more beneficial that the determining value process of MPP can also apply to optimize the I-V curve when designing algorithm. The process that MPP converges to its exact value is equivalent to cubic spline interpolation of the variable step size and the step size of spline interpolation lies on the topological structure of resistance matrix and its maximum individual load. According to the requirement from distributing measurement, measuring data collecting and transmission of control signals are based on ZigBee protocol. The nodes achieve control signals from the upper computer and return monitoring data. The timing interval of return monitoring data is controlled by the nodes. After they receive start control command from the upper computer, the nodes are synchronize with the command and produce a step value via pseudorandom sequence to determine the specific time to return data in periodic slot time.

PIC16F877 is used as MCU for each sensor node. ZigBee wireless communication module use CC2530 to connect PIC16F877 via a serial port. The current output of solar cells is

sampled by a 0.01Ω precision resistance then it will be magnified 50 times by MAX4376FAUK precision current applier for PIC16F877 to sample. The system uses an eight-size resistance matrix to get 256 scanning I-V curve values with 256 different load values. We use HCPL2630 optocoupler to isolate PIC16F877A and MOSFET drivers. The nodes use DS18B20 sensors and AM2301 sensors respectively to monitor the temperature characteristic of solar cells and ambient temperature and relative humidity to assess their working conditions. The data returned each time are sent back to the upper computer via ZigBee channels. The upper computer use LabVIEW to build the handler and the GUI interface to process and show the data. This program provides us with an interface to change the serial port and the sample interval. The data received are shown in four forms including instrumental panels, curves, a thermometer column and hexadecimal numbers. What's more, we can save the data to a certain file.



(a)



(b)

Fig. 1. Solar cell measurement node and GUI

IV. RESULTS AND DISCUSSIONS

In order to verify the feasibility and effectiveness of our I-V scanning algorithm, according to formula (1) to (3), we have simulated the process of monitoring solar cells in different temperatures with the 8-bit resistance matrix. The error of I-V curve reconstruction is shown in Figure.2 and Figure.3. It is shown that the refactoring I-V curves have high consistency with the original curves. The errors are nearly equal to AD quantization errors when voltage is small. The errors of the reconstructed I-V curves appear mainly beside the falling edge of boundary value MPP point, and especially beside MPP point, where the curves are with oscillation shape. The maximum of errors appears before MPP point as a result of the rapid change of the curve from smooth to fast falling, so that type error of cubic spline function increase with oscillation. At the falling edge with the change of the curves, errors reduce and gradually close to boundaries. Because of the accumulation of type error and the increasing influence of boundary errors, error curve grows a little. The reconstruction errors above can be reduced by multiple means and decrease of scanning step size. The errors beside MPP can be decreased especially when increase the resistance which is of the same order of magnitude with the series resistance of solar cells. It can be shown in the measuring result with the 8-bit resistance matrix that the errors of the refactoring I-V curves are smaller and the accuracy of the refactoring I-V curves are high. In addition, the method is of high operability so it can be taken as an effective way to measure parameters of solar cells.

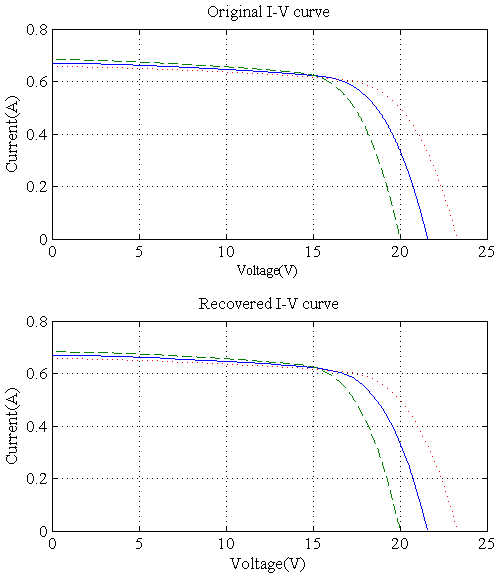


Fig. 2. Simulated results of dynamic resistance matrix scanning and interplolation reconstructions

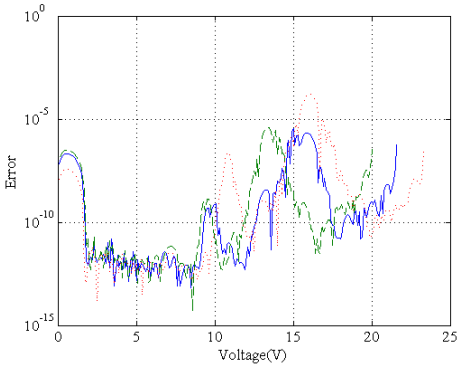


Fig. 3. Error of recovered I-V curve

Because of random access mechanism of polling, channel conflicts can't be avoided. Transmit buffers should be employed for conflict suppression. Figure.4 shows the channel conflicts of 1024 nodes in a transmit cycle with different buffer size (0, 4 and 16 cycle respectively). As can be seen from Figure.4, the channel conflict can be almost avoided even with 4-cycle buffer.

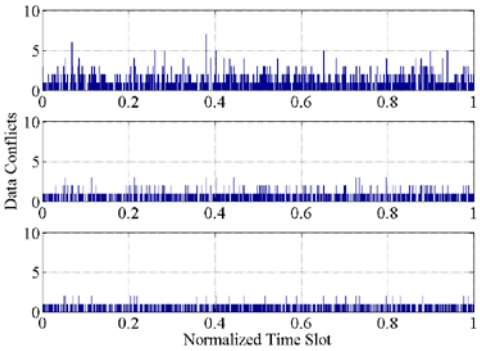


Fig. 4. Conflicts in ZigBee channel with difference cycle buffer

To test the performance of our system discussed above, I-V curve scanning about 16 solar cells has been carried out. The polling interval is set to 1 hour and the measuring period is 4 days (10:00 to 14:00 each day). The system sampling time set by real time clock and timestamps when data returns. The invalid data affected by the random change of cloud are deleted automatically. The measuring data are saved in database so that users can analyze them. The measuring I-V results are shown in Figure.5. In Figure.5, the shapes of I-V curve are achieved by dynamic resistance matrix scanning with spline interpolation algorithm shown above. The curves are more continuous around MPP points, which make it easier to get the accurate MPP value and the continuous I-V curve to analyze the solar cells performance. The MPP of solar cell is very important for solar cell performance test. The MPP measurement error is shown in Figure.6. Errors between the values calculated from our algorithm and the real values are less than 1% in most cases. The errors are in direct proportion to the scanning step size beside the MPP points. The errors can be further reduced if we increase the number of resistances in resistances matrix to shorten the scanning intervals.

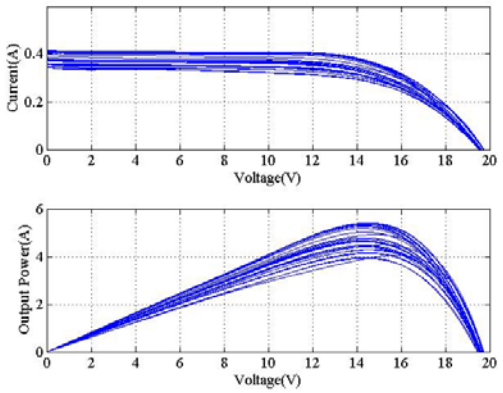


Fig. 5. Result of I-V curve scanning experiment

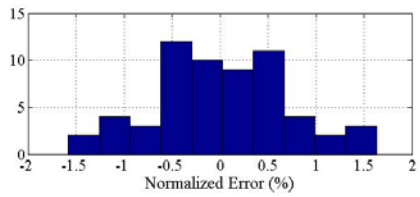


Fig. 6. Normalized MPP error between reconstruction MPPs and direct measurements

V. CONCLUSION

Outdoor solar cell measuring is important for PV system design. I-V curve and MPP are key parameters to determine the solar cell performance in outdoor condition. We report a low-cost ZigBee based measuring system for solar cell outdoor testing. A novel I-V curve dynamic scanning method is implied through the low-cost resistance matrix and revised discrete spline interpolation. The I-V curves reconstruction, ZigBee conflicts suppressing and MPP calibration are discussed through theoretical analysis and experiments. The results show that this method can provide high-precision solar cell characters measuring with low complexity.

ACKNOWLEDGMENT

The authors would like to thank Jiang Xue-qin and Zhu Ming-da from Donghua University for their useful discussion and programming supporting.

Supported by 2014 education department of Hebei Province (QN2014212), and 2013 Young Teachers Independent Research Program of Yanshan University (13SKA001) . Corresponding author: ztao@ysu.edu.cn

REFERENCES

[1] V.V. Tyagi, Nurul A.A. Rahim, N.A. Rahim, Jeyraj A., L. Selvaraj, “Progress in solar PV technology: Research and achievement,” Renewable and Sustainable Energy Reviews, vol. 20, pp. 443-461, April 2013.

[2] S. Zhang, P. Andrews-Speed, M. Ji, “The erratic path of the low-carbon transition in China: Evolution of solar PV policy,” Energy Policy, vol. 67, pp. 903-912, April 2014.

[3] T. Nielsen, C. Cruickshank, S. Foged, J. Thorsen, F. Krebs, “Business, market and intellectual property analysis of polymer solar cells,” Solar Energy Materials and Solar Cells, vol.94, pp. 1553-1571, 2010.

[4] M. Berginc, U. O. Krašovec, M. Topič, “Outdoor ageing of the dye-sensitized solar cell under different operation regimes,” Solar Energy Materials and Solar Cells, vol. 120, Part B, pp. 491-499, January 2014.

[5] J. K. Kaldellis, M. Kapsali, K. A. Kavadias, “Temperature and wind speed impact on the efficiency of PV installations. Experience obtained from outdoor measurements in Greece,” Renewable Energy, vol. 66, pp. 612-624, June 2014.

[6] B. Marko1, K. Opara1, T. Marko, “Outdoor ageing of the dye-sensitized solar cell under different operation regimes,” Solar Energy Materials and Solar Cells, vol. 120, pp. 491-499, January 2014.

[7] K. Naohiko, H. Kazuo, T. Hiromitsu, N. Junji, S. Toshiyuki, T. Tatsuo, “Improvement in long-term stability of dye-sensitized solar cell for outdoor use,” Solar Energy Materials and Solar Cells, vol. 95, pp. 301-305, January 2011.

[8] V. Vaidya, D. Wilson, “Maximum power tracking in solar cell arrays using time-based reconfiguration,” Renewable Energy, vol. 50, pp. 74-81, February 2013.

[9] M. Koehl, M. Heck, S. Wiesmeier, J. Wirth, “Modeling of the nominal operating cell temperature based on outdoor weathering,” Solar Energy Materials and Solar Cells, vol. 95, pp. 1638–1646, July 2011.

[10] M. Zagrouba, A. Sellami, M. Bouaïcha, M. Ksouri, “Identification of PV solar cells and modules parameters using the genetic algorithms: Application to maximum power extraction,” Solar Energy, vol. 84, pp. 860-866, May 2010.

Integrating Biometric Sensors into Automotive Internet of Things

Need and Proposed Implementation

Rupak Rathore, Carroll Gau
ATCS (Beijing) Technology Consulting Co., Ltd
Beijing, China
rupak.rathore@atcs.com; carroll.gau@atcs.com

***Abstract*—A rapidly ageing population, the prominence of obesity and associated medical conditions, widespread incidents of drunken driving, advent of telematics, advances in medical devices and emergence of 4G mobile networks necessitates and enables the convergence of biometrics into automobiles that can facilitate rapid and relevant responses to save precious lives. In this paper, the authors describe how such a convergence can be achieved by proposing an automotive healthcare and safety framework controlled by a dedicated healthcare systems control unit and its integration into an automotive Internet of Things containing telematics and other systems. The paper further explores the fulfillment of emergency response using cloud computing and vehicle to anything technologies as well as the potential impact such a framework can have on safety of our cities.**

Keywords—Biometrics; Sensors; Telematics; Internet of Things

I. INTRODUCTION

With continuous advances in technology, an automobile today is no longer just a mechanical means to take humans from one place to another but rather an "Internet of Things" in motion with numerous subsystems and hundreds of sensors that communicate with each other and, using 4G as well as vehicle to anything (V2X) technologies, with other automobiles.

Asimov's First Law of Robotics in [1] states that "A robot may not injure a human being or, through inaction, allow a human being to come to harm." In USA alone, there were over ten thousand fatalities in automobile crashes associated with drunk driving in 2012 [2] and close to fifty thousand accidents caused by medical emergency from 2005-2007 [3]. As the global vehicle population grows and disposable incomes rise in emerging economies, global fatalities due to impaired driving may go up even higher [4] [5]. It is an opportune time to apply Asimov's First Law of Robotics to the automobile. An ideal characteristic for automobiles would be the ability to prevent such accidents before they occur and to provide necessary inputs for a rapid and effective response after they do.

In this paper, authors conceptualize an "Automotive Healthcare and Safety" (AHS) framework that can fulfill this behavior.

II. AN AUTOMOTIVE HEALHCARE AND SAFETY FRAMEWORK

Advances in medical devices have made it possible to build sensors that are small enough to be embedded into a vehicle without causing an undue impact on aesthetics. In addition, an automobile is less restricted by size and weight in comparison to consumer electronic gadgets. A modern automobile can thus be unobtrusively embedded with biometric sensors into various parts of an automobile's interior. The proposed AHS framework is envisioned as a set of biometric sensors feeding real-time data into a dedicated control unit that can be configured to alter vehicle behavior based on configurable parameters.

As depicted in "Fig. 1", the AHS framework consists of a healthcare systems control unit (HSCU) that communicates and collaborates with other in-vehicle networks such as the telematics control unit (TCU), V2X, the engine control unit (ECU), the global positioning system (GPS), etc. The HSCU controls sensor arrays established within the vehicle and leverages in-vehicle displays such as the head unit (HU) along with dedicated displays for the AHS. The AHS framework is configurable in order to be able to tap into other devices including consumer electronics such as smart watches and smartphones or medical devices such as glucose monitors using near field communication (NFC), Bluetooth, Wi-Fi as well as wired channels such as universal serial bus (USB).

The HSCU consists of a real time process controller to continuously monitor the sensor arrays, an event management subsystem to address triggers received or generated using algorithms, a rules engine to host the configurable parameters controlling AHS operations, a security and encryption framework providing at-rest and in-transit data security, and an authentication and profile management subsystem providing a way to automatically segregate and protect sensitive information. The personal health profiles are stored in the AHS cloud that is accessible via the TCU and/or the user's smartphone. Once properly authenticated the AHS can use and load appropriate health profiles from the AHS cloud into local storage within the HSCU.

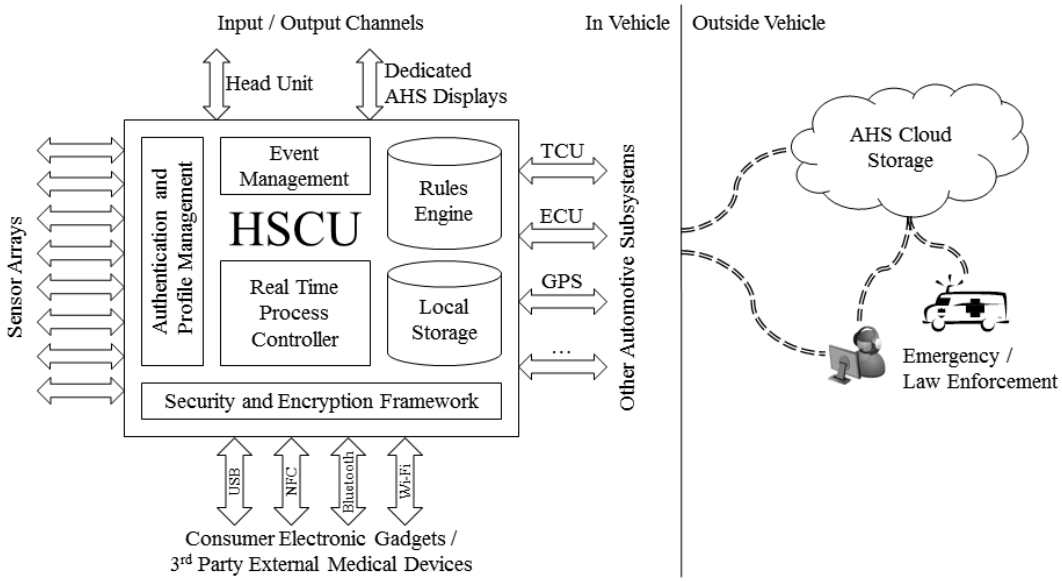


Fig. 1. The Automotive Healthcare and Safety Framework

The proposed framework is naturally extendable on the left as more and more biometric sensors are made available. These in-car biometric components could be powered by the vehicle's power systems and continuously record at preconfigured intervals into HSCU.

The AHS framework is envisaged as a multi-tenant framework and thus has a potential to cover all occupants of the vehicle, provided adequate sensors are available and configured for passengers. Privacy concerns may require a dedicated set of biometric sensors for each passenger.

The AHS framework allows for the use and loading of appropriate health profiles from AHS cloud. This enables interoperability across multiple vehicles and also provides the ability for appropriate third parties to use the information in case of an accident for a rapid and effective response.

III. Possible Implementation of AHS Framework

A possible single-tenant implementation of the AHS framework is depicted in "Fig. 2". Sensor arrays consists of heart rate, blood pressure and body temperature sensors on the grip of the steering wheel, a breath analyzer in the center of the steering wheel, a dashboard visual light and infrared cameras to detect pupil dilation and facial symptoms, air quality sensors on the center console, a wide angle camera on the rear view mirror, an electrocardiogram monitor in seat belts, weight and movement sensors in the seats, a fingerprint scanner on the gear selector knob, etc. Output from the HSCU displays real time data and analysis from the sensors on the windscreen. The HSCU is also connected to the HU to allow interaction with the driver.

IV. Possible Use Cases for AHS Framework

A. *Proactive Personal Health Monitoring*

As a population grows older, keeping track of an individual's health becomes more important. People spend a significant amount of time in their automobiles on a regular basis. This allows the AHS framework to collect a person's vitals on a regular basis using the vast array of sensors at its disposal.

When the automobile is powered up, the AHS uniquely identifies the driver using an authentication and profile management subsystem. The appropriate medical records are then loaded into memory and configured into to the profile, after fetching and synchronizing with the AHS cloud. If paired

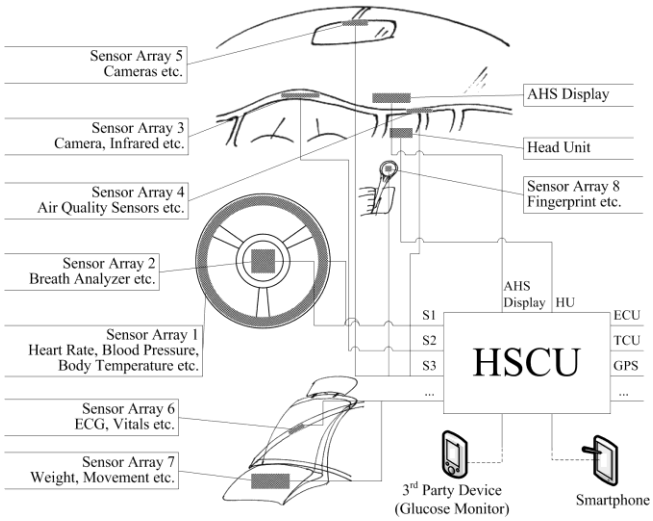


Fig. 2. A single-tenant implementation of AHS framework

with smart devices, a unidirectional or bidirectional communication channel is also established. The AHS framework then triggers collection of data from its sensors and keeps monitoring continuously at preconfigured periodic intervals. In collaboration with health applications that are installed on consumer electronic devices, a comprehensive personal medical history can be maintained. The care givers can be authorized to collect the medical history from the AHS cloud. This improves the efficiency and effectiveness of trips to the physician.

B. Medical Emergencies

During the course of driving the vehicle, the AHS framework is continuously collecting vitals from the driver at preconfigured periodic intervals. The data is being continuously analyzed in HSCU and if the HSCU detects an anomaly in medical data, it prompts the driver over the HU and/or audio systems in a non-distractive way and waits for a response. The framework can be configured to automatically trigger medical emergency procedures in the event no response is received from the driver within a certain time frame and/or erratic vehicle behavior such as sudden acceleration or deceleration is detected by other systems in the automobile.

When a medical emergency is triggered, either by the driver or by the AHS framework, the HSCU in conjunction with the TCU establishes communication with emergency services. Once a voice and data connection is established, all relevant medical data is automatically passed on to emergency responders in an encrypted form and a live data feed is provided to allow emergency responders to continuously monitor the occupants' vitals until assistance arrives. The responders will be also able to advise the passengers of basic self-sustenance measures for initial care if deemed appropriate, saving precious time. Finally, using identity, vitals and medical history, emergency responders can better prepare themselves for the patients' status when they arrive.

The AHS will also establish communication with the ECU and other automobile systems such as assisted driving or semi-autonomous driving to switch on hazard lights, alert law enforcement, safely slow, and maneuver the vehicle into a safe position. The AHS framework is also capable of using vehicle to vehicle (V2V) and V2X communication channels, if the automobile is appropriately equipped, to search for nearby medical facilities as well as emergency responders by using simulcast over multiple channels to shorten the emergency response time.

C. Proactive Preventive Scenarios

The AHS framework is capable of triggering proactive preventive responses based on configurable parameters. As the vehicle is powered up, the AHS framework kicks in and initiates a data collection from a preconfigured set of sensors. The AHS can prevent vehicle movement in case the driver's vital parameters are beyond the acceptable range. These thresholds either can be personally configured or mandated by regulatory requirements of a given location as obtained by GPS sensors. Safely and proactively disabling vehicle movement coupled with driver assistance can help prevent thousands of accidents per year.

For example, if the breath analyzer detects that the driver has a blood alcohol content (BAC) exceeding a preconfigured level (such as 0.08), a connection with the ECU can be established, and instructions for disabling the engine are sent along with a warning message to the vehicle's HU. As a secondary action, the AHS can communicate with the TCU to initiate an assistance call for services such as "designated driver" or "taxicab services."

D. Reactive Preventive Scenarios

Since the AHS framework is continuously monitoring a driver's vital health parameters, when coupled with other automobile systems, the AHS forms a strong sensory network that can prevent accidents caused by distracted driving and bad driving practices by warning and later enforcing, if so configured.

For example, the AHS framework can be configured to detect drowsiness using blood pressure and pulse monitors coupled with detection of abnormal steering wheel behavior. It can then not only provide feedback to the driver using cabin lights, horns, music volume, and other mechanisms to alert the driver but also compensate for impaired motor skills by increasing the dead zone on steering wheel or increasing the effectiveness of advance driver assistance systems.

In another example, the AHS can use eye movements, steering wheel responses, and noise levels to determine a lack of focus on the road and appropriately warn the driver and, in extreme scenarios, request other automobile systems to prepare for emergency reactions such as hard breaking commensurate with environmental factors. In such a case, V2X communication can also be utilized to warn other vehicles in the vicinity of the potential danger posed by an impaired driver. Other vehicles can then indicate to their drivers of the presence of a dangerous vehicle.

E. Reactive Permissive Scenarios

In the unfortunate event of an accident, the AHS framework will work collaboratively with the TCU and other systems to supplement automatic collision notification (ACN) messages and subsequent emergency protocol behavior.

As demonstrated earlier in the "Medical Emergencies" section above, the occupants' vitals will be sent over to emergency responders and a live feed will be established to allow responders to continuously monitor the situation. For multi-tenant implementations, all occupants' data can be shared to emergency responders. In case of single tenant implementation, the AHS can still use the infrared sensors and cameras coupled with other vehicle sensors to provide inputs regarding other occupants to emergency responders.

F. Other Use Cases

The AHS framework can be configured to automatically record video feeds from vehicle cameras, both inward and outward, in conjunction with other systems, and automatically upload video to the Cloud in the event of a traffic accident. That data can then be utilized by law enforcement agencies to serve as evidence if required.

The AHS framework can even be utilized while vehicle is not in use by using "standby" mode. A standby mode can be configured for limited usage of censors and reduced power consumption. This standby mode can be configured to detect when and if there is a human or animal present in the vehicle when the automobile is locked and all windows are shut. This can help reduce the number of incidents where infants or pets are accidently left in parked vehicles and die due to exposure. The AHS framework can be configured to page the owner of the vehicle to notify them. In the extreme event that the vehicle detects abnormally high temperatures or loud noises in the vehicle, emergency services can also be automatically notified.

V. CONCLUSION

As we enter an era of ubiquitous sensing and connected systems, it is important that automobiles evolve alongside other technologies. The AHS framework provides a possible means for developing an extensible healthcare subsystem for deployment in automobiles.

In 2012, there were over ten thousand fatalities in crashes involving a driver with a BAC of 0.08 or higher in the USA alone [2]. These scenarios can be prevented proactively by disabling the vehicle temporarily. The AHS framework allows for safe limits of BAC to be configured by well-wishers and/or mandated by regulatory policies, potentially saving thousands of lives across the globe. Similarly, a study conducted in 2007 discovered that there were close to fifty thousand incidents of traffic accidents caused by medical emergencies from 2005-2007 in USA [3] and in 2012, the National Highway Traffic Safety Administration estimated 3,328 people were killed and 421,000 injured in traffic accidents involving distracted driving. Finally, 10% of 15 to 19 year olds involved in traffic accidents reported being distracted at the time of the incident [6]. Coupled with assisted driving or semi-autonomous driving, there is a significant potential to reduce the number of accidents caused by preceding medical emergencies and distracted driving and help drivers in distress.

Research suggests that a patient will lose consciousness 8 seconds after a cardiac arrest event occurs and the brain damage occurs after the first 4-6 minutes following the event [7]. The AHS framework can detect symptoms of cardiac arrest and automatically notify emergency responders of the situation while the framework continues to actively monitor the driver's vital health parameters and relaying this data to emergency responders enabling them to better prepare themselves for arrival at the scene. While there is no correlation between EMS response times with mortality rates, but there is a need for identification of patients who could benefit from rapid EMS response [8]. Additionally, EMTs arriving on scene have to spend precious time to determine the mechanism or nature of the injury, the total number of patients, the necessity to request for additional units if required, and determine the priority for required emergency care [9]. The AHS framework can be configured to inform emergency responders of the time-sensitive nature of the medical emergency and provide initial information to the responders, reducing delay for the onset of the medical emergency to treatment being administered. More than six hundred children have died in vehicles due to heat stroke in USA [10]. The standby mode of AHS can help reduce the number of incidents where infants or pets are accidently left in parked vehicles and die due to exposure.

As demonstrated, solutions developed using the AHS framework will allow for more rapid and appropriate reactions when faced with critical decisions involving human life by preparing emergency responders to the circumstances while en route. The AHS will also have compatibility across multiple automobile manufacturers as well as emergency service providers, potentially saving thousands of precious lives.

REFERENCES

[1] I. Asimov, "I, Robot," Gnome Press, 1951.

[2] National Highway Traffic Safety Administration, "Traffic safety facts 2012," DOT HS 811 870, U.S. Department of Transportation, December 2013, pp. 42.

[3] National Highway Traffic Safety Administration, "The contribution of medical conditions to passenger vehicle crashes," DOT HS 811 219, U.S. Department of Transportation, November 2009, pp. 4.

[4] J. Sousanis, "World vehicle population tops 1 billion units," Wards Auto, August 2011.

[5] M. Jha, S. Amerasinghe, J. Calverly, "Taming the Gini: Inequality in perspective," Standard Chartered, July 2014, pp. 10.

[6] National Highway Traffic Safety Administration, "Traffic safety facts 2012," DOT HS 812 012, U.S. Department of Transportation, April 2014, pp. 42.

[7] J. Mayer, ""Response time and its significance in medical emergencies", Geographical Review, Vol. 70, No. 1. (Jan., 1980), pp. 79-87

[8] I. Blanchard, C. Doig, "Emergency medical services response time and morality in an urban setting", Prehosp Emerg Care. 2012 Jan-Mar;16(1):142-51. doi: 10.3109/10903127.2011.614046. Epub 2011 Oct 25.

[9] New York State Department of Health, "Job description- emergency medical technician - basic", New York State, retrieved on November 11, 2014 from https://www.health.ny.gov/professionals/ems/pdf/srgemt.pdf

[10] J. Wiesenfelder, "Child deaths in hot cars: 10 key facts" US News, August 2014

[11] C. Mclaren, J. Null, "Heat stress from enclosed vehicles: moderate ambient temperatures cause significant temperature rise in enclosed vehicles," PEDIATRICS Vol. 116 No. 1 July 1, 2005 pp. e109 -e112

[12] Di Liu, Zhi-Jiang Zhang, Ni Zhang, "A biometrics-based SSO authentication scheme in telematics," 2013 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery, pp. 191-194, 2012 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery, 2012

Study and Implementation of MVC Pattern upon Extended Function in the Management System of University Laboratory Project Declaration

Weihong Wang
College of Computer Science and Technology
Zhejiang University of Technology
Hangzhou, China
wwh@zjut.edu.cn

Wentao Xu
College of Computer Science and Technology
Zhejiang University of Technology
Hangzhou, China
elfxwt@163.com

***Abstract*—Based on the management system of university laboratory project application in ZJUT, this article studies the technical principle and core things deeply so as to know the actual effect of web application developed by WebWork framework and extended functions by jFinal framework. At the same time, a MVFO model which is suitable for this project is provided according to the insight of MVC and specifically analyzed by the function of exporting form view in Excel spreadsheet, which strengthen the comprehension and study of software architecture theory.**

Keywords—MVC architecture; Work flow; Jfinal Framwork

I. MVC PATTERN

With the development of software pattern and framework, a lot of outstanding design patterns have been provided to enhance the scalability and maintainability of project and optimize the logic of codes, which have been already employed to many fields widely. Among those, the MVC pattern, first made available by Gamma, has played a great role in web application development gradually. It makes use of the object-oriented thinking to divide the whole software responsibilities into three parts in order to reduce the complexity of development and supply a clear way to design web application [1]. At the same time, it even becomes the standard rule of software engineering to some extent.

A. The Composes of MVC

There are three parts in MVC patter: Model, data model, which transports the input and output data to uniformly process and saves them as the abstract object, encapsulating the data structure and operation [2]; View, the outside presentation of data, which provides users the visualization by many relevant views and make active interaction with outside world to take on the responsibility to trigger the implementation of inner logic; Controller, the bridge of View and Model, which maintains the views by processing the input data according to corresponding rules and requirements so as to achieve business logic [3]. The specifics of the three parts' relationship are as followed:

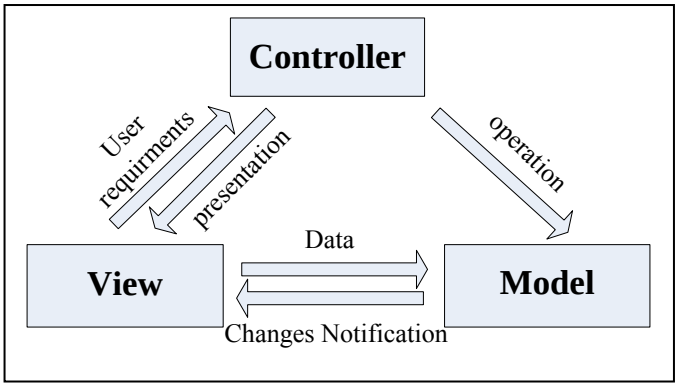


Fig. 1. MVC pattern

B. Pros and Cons of MVC

As a structural and creational software design model, MVC allows for reusable components to be applied in an expandable system and reduces the development complexity of Web-based application by separating data logic and business logic of projects, leading the general and normative way for large web service applications [4]; And the three parts not only have a close connection but also have a pronounced division of labor, ensuring the realization of business logic and the requirements of project by the way of controller connecting the view and model, which make full use of the object-oriented idea; Furthermore, Model, View and Controller can be concurrent and has no relationship with platform, improving the efficiency of development partly [5].

Nevertheless, a clear division of labor and objectified design will inevitably lead to the tedious code and extra overhead, which produce the increasing of workload for small web-based applications that not need the strict divisive structure; at the same time, there are also some issues exist in large scale projects, for example, the unclear part belong to some functions and classes when there have the increasing number and complexity of middle operations and the bulky Controller which is difficulty to control and organize [6].

Therefore, the individual architecture for different project should be designed according to specific requirements on the base of MVC model. In this article, a MVFO model has been proposed by the extend function of management system in university laboratory project application, which will be analyzed with details in chapter three.

II. MVC MODEL OF PROJECT

A. Project Introduction

University Laboratory Project Declaration System is an innovation responding to "National Five-year plan", which combines Education and Information industry. This project, based on the department of laboratory and property management in Zhejiang University of Technology, realized the whole application process of laboratory constructed project declaration in Web System, involving many roles of the departments and four main workflows: the academy long plan of laboratory construction project application, the laboratory construction project application, the equipment purchase application, large equipment and self-made equipment purchase application. As a result, the capability of university to undertake large projects and manage the purchase of equipment has been improved.

MyApps software developed quickly platform has been employed in this project and its in-built components, including workflow engineer, Form constructor, Report Constructor and other visual tools deal with the business functions, such as the achievement of data, workflow processing and reports presenting. In addition, the system is divided into two parts, foreground processing and background management. The logging user in the foreground could set his department and role in the background so as to present different page view according to different users. At the same time, the powerful workflow engine of MyApps makes the easy development of management system based on workflow. Take the purchase management module as example, whose design Figure is as follows:

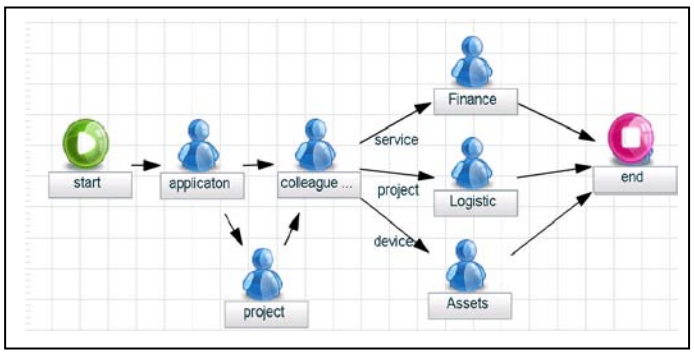


Fig. 2. The workflow of purchase management

B. Analysis of the Whole Functions

The bottom framework of myApps has employed the WebWork framework and made best use of MVC model thinking [8]. This project, which makes the secondary development based on myApps, finish the basic functions of the University Laboratory Project Application management system, such as fill out form, data storage, conditions search, view presentation, review based on process and so on. Next we will analysis a whole function in the project according to MVC model thinking.

Take the purchase application of large device as the example, firstly input data in the view and send out data request, then show up the view page, which is realized by JSP technology. The data request is encapsulated into Action object by Servlet controller, which will be the instance of many data models by the time of the building of many view objects and transferred to background to prepare for corresponding business logic operations.

Among the purchase application of large device, different login roles correspond to different menu functions, for example, the leader and relevant department need to check the revised form, which requires the form that satisfied the conditions searched out and showed up on the front page. On the myApps developed platform, some complex functions can be realized by easy script language, just as the SQL script to query the results from database.

```
var doc = getRelateDocument();
var id = doc.getId();
var sql = "select * from tlk_m05_spending_plan where
ITEM_COMMON_PROJECT_ID = '" + id +  "'"
sql
```

Fig. 3. SQL script

MyApps platform employs the Hibernate Framework, which maps the Java object to database by corresponding configuration files and mapping files so that it can make the operation of modification and update and so on for data [9]. Here, the SQL statements query the qualified data objects by Servlet controller of Hibernate and commit the corresponding operation to database, then return the query results and show to the explore by View level. The View level makes the encapsulation of data request to data model by controller, and then executes corresponding operations in databases by data model and some business logic of request in controller, finally shows the return results to explore in view level by controller.

III. MVC OF PROJECT EXPANDED FUNTION

The basic requirements of University Laboratory Declaration Management have been finished by employing the myApps platform effectively at the beginning. However, with the increasing of requirements, the platform cannot satisfy the real requirements, which is embodied by the imperfection of platform statistic function and the lack of document import and export functions. Because of the complexity of original framework, we explore an extended package by another simple MVC framework, Jfinal. Through the interaction of data and operation with original system, we has already explored many extended function modules, such as Form-View export Excel

Module, Form content export Excel Module and Form status statistic Module.

A. *Jfinal Framework*

JFinal is a simple, light, rapid, independent, extensible Java WEB+ORM framework, which is also a MVC structure like Struts2 Framework [10]. During the process of development, there are no complex operations of xml configuration and database. So we choose JFinal framework to develop the extended functions of this project.

There are five parts of this JFinal framework, including Handler, Interceptor, Controller, Render, Plugin [11]. Among that, Controller is the controller of MVC model as one of the kernel classes, which handles the definition and organization of Action and plays an important role. At the same time, JFinal employs ActiveRecord to operate data and Model component to be the model of MVC, which can have the database operations by inheriting them and have simpler operations and higher efficient than Hibernate Framework [12]. This project takes best use of the convenience supplied by JFinal Framework and realized the corresponding business logic directed by MVC model.

B. *Analysis of Extended Funtions*

Take the Form-View export Excel Module as an example. We need export the represented or queried data of form-view to Excel. The details are as followed:

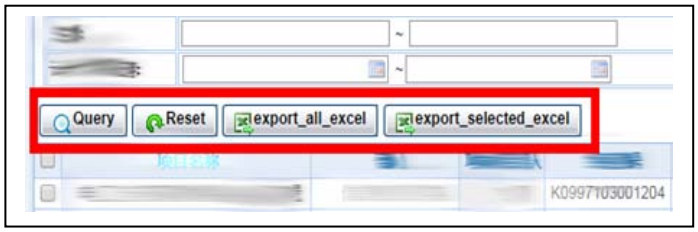


Fig. 4. The view of Export Excel Button

The above picture is a part of View in the MVC model of the Export Excel functions and it employs the original view of myApps resulting from the base on myApps software platform. In this view, we add two extra buttons, export all Excel and export selected Excel, which can realize the function of exporting Excel, to post the wanted data by Controller component. With the demand of actual requirements, two data models were designed because the represented data and searched data by conditions should be exported. At the same time, other models, such as Columns, were also designed to map different views, including project implementation schedule Form, project device items Form, fund plan Form. The Models extend class Record of ActiveRecord in JFinal framework and provide many other functions adapt to MVC operations. Despite that, the function of exporting Excel needs relevant database operations to achieve the right data other than the direct data passed by View, so we build up xml document to assistant the database query in order to expand data Model. The function of View and data Model are shown as following.

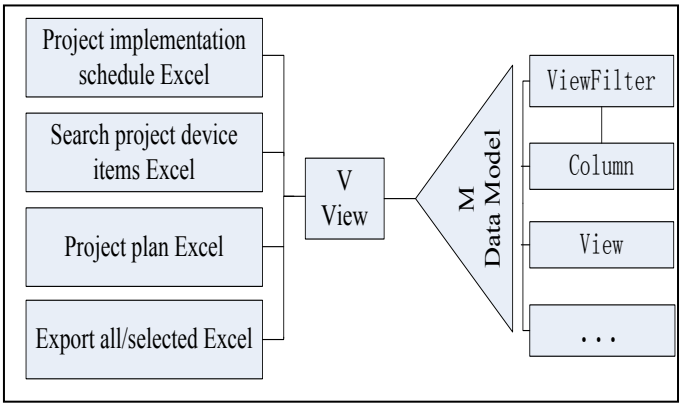


Fig. 5. The function of View and model

We design ExportView Class to inherit kernel Controller class, whose design principle is to accept the data transported by View, and have a lot basic interacted functions of View and Model. This Controller of exporting Excel is mainly to accept data posted by View which will be build up Data Mode object of View and ViewFilter. And at the same time, the data achieved by relevant database operations and document format to shape Excel form view which can get from web. There have the renderFile function owned by Controller in JFinal framework to realize the specific implementation. The project documents of Controller and Data Model are shown as followed.

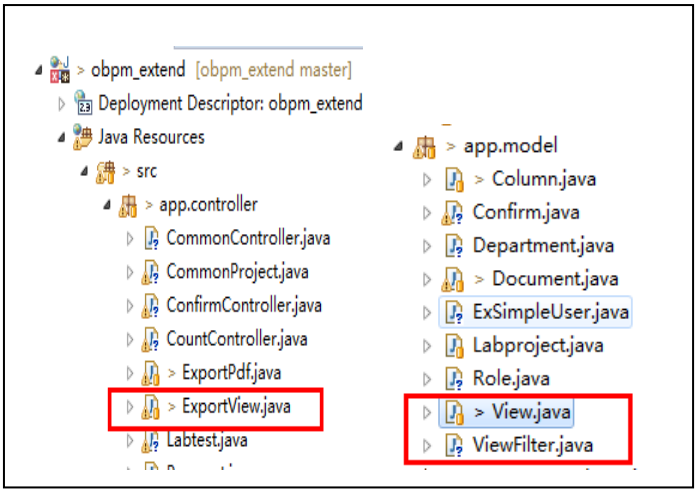


Fig. 6. The project organisation

Next illustration is the whole process of specific design.

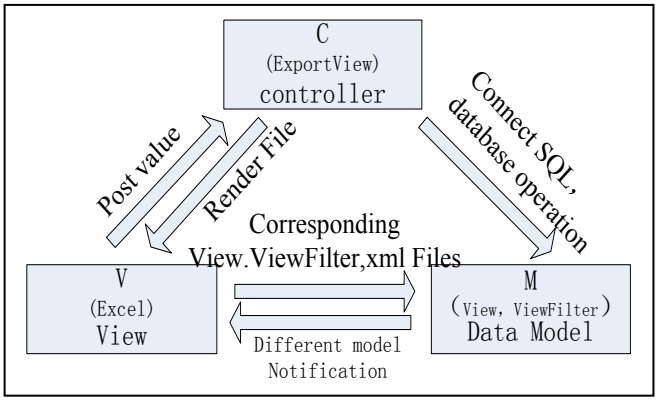


Fig. 7. The whole process of specific design

C. MVFO Model

Although we can manage different and variable requirements by separating Model Data and View part through Controller in the time of designing system extended function by MVC framework, it is hard to label some functions in the specific background operations and used to be classified to Controller leading to complexity and redundancy. And at the same time, this also makes the whole project hard to maintain and expand. Therefore, this paper presents a new framework MVFO to deal with these problems according to the extended functions of this project. In this framework, M and V are still original meaning, Model and View, while F is Function, including some functions that get data from view or generate object data, and O is Operation, mainly focusing on the middle and necessary operation to achieve the object data, such as some operations of database and so on. That is to specific the details of Controller to two classes. It is no doubt that there have different missions in F and O by different projects. In this extended function project, F includes the functions of accepting view data and generating Excel, while O is responsible for some related database operation such as the splice of SQL and so on. The following is the design figure:

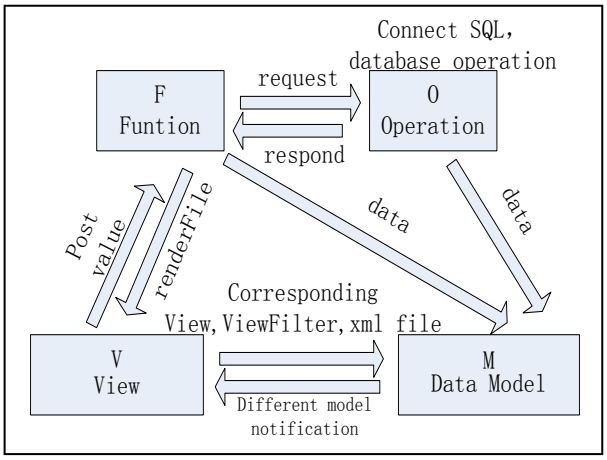


Fig. 8. the design figure

This Framework is just adapt to some large and complexity project but not all the projects, and the above specific operations is on the Laboratory project declaration and make relevant operations more clear, providing convenience to the maintainability and expansibility of follow-up work.

IV. CONCLUSION

This paper has studied the principle and thinking of MVC framework deeply based on the implementation of Laboratory Project Declaration management system and the details of Java Web upon the MVC framework, including WebWork framework and JFinal framework, by analyzing the design of MVC in this project. What's more, we developed the extended functions of this project employing the JFinal framework which integrated well to myApps platform, and analyze the MVC framework specifically by taking export reports functions as the example to propose the MVFO framework adapted to this project further, separating the functions and operations from controller to get better improvement of expandability and maintainability.

Because the emphasis of this paper is the design and study of MVC framework, there have no specific introductions and references of design patter.

ACKNOWLEDGMENT

This paper is supported by the National Natural Science Foundation of China (60873033), the Natural Science Foundation of Zhejiang Province (R1090569 ,LY12F02039), and the State Key Laboratory of Software Development Environment Open Fund (SKLSDE-2012KF-05).

REFERENCES

[1] Chengwang He,Qiuhui Yu, "Study of MVC2 Archtecture and Softeware Framework," Computer Engineering. 2002, 28(6), pp. 274-275.

[2] Zhongfang Ren, Hua Zhang, "Summary of MVC Architecture", Study of Computer Application, 2004, 10(528), pp. 12.

[3] Wei Zeng, Baoping Yan, "Summary of Work Flow Model Study", Study of Computer Application 2005, 22(5), pp. 11-13.

[4] Keke Yi, Zhigang Chen, "Design and Study of Web OA System based on MVC Archtecture" ,Computer Engineering and Application, 2005, 4, pp. 112-115.

[5] Fujuan Wang, "Design Model of MVC", Computer Engineering, 2005, 31(9),pp. 96-97.

[6] Yijian He, "MVC Application Study based on Struts Framework",Computer Knowledgy and technology:comunication of study. 2010, 6(005), pp. 3534-3536

[7] Yongliang Li, Shewu Cui, "The Improvement and Application of MVC Design Pattern",Computer Engineering, 2005, 31(9), pp. 96-97

[8] Qin Lin, Junshan Tan, "Design and Implementation of Web Report Show Based on Struts Framework",Computer System Application, 2006, 11, pp. 25-28..

[9] Xiangzhong Fen, "The Study of Struts Framework and Application Based on MVC Design Pattern",Computer System Application. 2006, 11: 25-28.

[10] Chengwang He,Qiuhui Yu, "Study of MVC2 Archtecture and Softeware Framework," Computer Engineering. 2002, 28(6), pp. 274-275.

[11] http://www.yiiframework.com/doc/guide/1.1/zh_cn/basics.entry

[12] Huiguang Yao,Yuesong Zhao, “Application of MVC Model in Web Programming”, Development of Micro Computer 002, 11(3) pp. 9-10.

[13] Raible M. Comparing Web Frameworks Struts, Spring MVC, WebWork, Tapestry & JSF[J]. Virtuas Presentation. Available at: http://www. chariotsolutions. com/slides/spring-forward-2006-web-frameworks. pdf, 2006.

[14] Shancheng Tang, “Preliminary Study of Webwork Principle”, Computer Knowledgy and Technology , 2005 (2),pp. 82-85.

[15] Yang Yu, Xingdong Lu, Ma Fang, “The MVC Design Model of Applying Struts”,Computer Application . 2003, 12, pp.346-347

[16] Wanlong Li, Xueli Wu, Yanxia Wang. “The Implementation of Web Application based on Struts Framework”, Computer Technology and Development, 2006, 16(4),pp. 102-104.

[17] Yi Kou, Liwen Wu, “The Application Function of Struts Framework Based on MVC Design Model”,Computer Application, 2003, 23(11), pp. 91-93.

[18] Fujuan Wang. “MVC Design Pattern”,Silicon Valley.2009(007).

[19] Baohua Qin, Yongjin Zhang,Yi Sun, “The Study of Web Information System Framework Based on MVC Framework and J2EE Architecture”. Modern Electronic Technology, 2005, 3, pp. 12-l4

[20] Xiaopeng Ren, Wenbin Zhao, Chunping Zhang. “The Study of Web System Develop Based on MVC Framework”. Computer Engineering and Design. 2010, 31(4),pp. 772-775

Using V-Model Methodology, UML Process-Based Risk Assessment of Software and Visualization

Muhammad Rashid Naeem
Weihua Zhu
School of Software Engineering
Chongqing University
Chongqing, China
rashidnaeem717@yahoo.com

Adeel Akbar Memon
Adeel Khalid
School of Software Engineering
Chongqing University
Chongqing, China

***Abstract*—Risk Assessment is one of the most critical parts of software engineering process and Risks are the factors that could be results in software failure if they are not correctly handled. In this Paper we propose a solution to reduce risks using UML visualization of software processes in detail and V-Model (Software Development Lifecycle Model) techniques are performed to produce verified results in each phase of development lifecycle. Using all above aspects and research findings we propose "Risk Assessment V-Model". Furthermore, we use "Do Sale Process" to discuss different properties of our model. This model is quite in general and can be applicable and expendable for many kinds of software systems.**

Keywords—Risk Assessment, UML, Risk Identification, Software Project Risks, V-Model

I. INTRODUCTION

One of the critical concerns in software industry today is risk management. Risks could results in software failure if they are not managed accordingly. Risks are basically combination of two factors where first is software malfunctioning. It focuses on failure to deliver at certain part of software development cycle. Second factor is severity. It shows effects of malfunctioning in software processes. Sometimes malfunctioning cannot be manageable. For example if all software processes are dependent on malfunctioned process and it need to be change. It may affect whole software and may results in software failure.

Software risk management is the practice of assessing and controlling risks that affects the software projects, process or products. In [1], S. Nurazlina et al describes that risk management as practice to avoid software risk and using software visualization it may helpful for software risk assessment in software development life cycle.

In [2], M. Bosan et al defined possibility of loss by software risks in form of poor quality and increased cost and also as software failure or delay in completion. These days as Information Technology is expanding, there are more chances of risks in software development then before. In [3], Z. Kremljak and C. Kafol describe different types of risks to which companies are exposing day by day and their effect on software projects

Risks are affecting largely to software industry. W. Han and S. Huang also describe its effect on software projects and show report [4] by Standish Group in 2004 which indicated that 53% of projects were unable to deliver on time, within budget or within required functionality whereas 18% projects were cancelled due to software risks. According to this report risk assessment is one of the most critical parts of software engineering.

For risk assessment there are various measures have been taken and some of them are implemented and currently in use by many organizations. Most of these measures focus on specific area of risks. In next section we will discuss some of related approaches proposed by authors to handle risks.

II. RELATED WORK

Risk assessment is done at different levels and different stages. There are several methodologies are introduced to handle different kinds of software risks associated with software and software development lifecycle.

Ranking risks is the most popular technique to measure severity of software risks. In [5], Y. Wang et al proposed a technique to rank risks on historical data. Assessment is divided into two parts, First part is gathering assessment set from Historical data. Secondly achieve matrices values for risks in the set.

Trustworthiness is also a major factor in software development model. In [6], J. Li et al propose trustworthiness metrics and risk management effectiveness calculation methods with risk transfer assumptions and cost constraints to improve software process risk assessment and enhance trustworthiness for software process management.

Software Visualization technique is also a good way to reduce software risks as we mentioned previously in [1]. Using software visualization there is more possibility of assessment accuracy.

Besides, these are many assessment models for risks some of them based on probability matrices, mathematical matrices or on questionnaires based risk assessments and most of them are for specific field of use.

A survey conducted by K. Georgieva et al in [7] gives analysis of different assessment models and a timeline from 1995 to 2008 about risk assessment. Some of them ensures qualitative and some quantitative assessments.

From all above observations we conclude that software risk assessment is critical part and as risks involve, there is possibility of failure or delay in software completion. There are many models to overcome but we use assessment based on UML Specifications. As UML can better visualize software processes, therefore it's easy to determine risks at every phase.

Our model is methodological base which main goals are to identify and monitor risks and also ensures traceability at each phase.

III. RISK ASSESSMENT MRTHODOLOGY

In this model we use methodology based on UML and V-Model development lifecycle. As we know UML is the most widely used for software development and it is understandable by developers and designers. UML diagram also helps to manage complexity of system and architectural problems and enhances comprehensibility [8] and [9].

On other hand V-Model is widely accepted development lifecycle for software development [10]. It is best model for test driven software development. Each phase of V-Model have two objectives. One is Validation and second is Verification as defined by IEEE in [11]. Validation refers to requirements analysis and verification refers to evaluation of requirements at each phase with respect to validation.

In our methodology we assume validation as identification of risks and validation as monitoring or evaluation of risks. We have seven phases of risk assessment in our model as shown in Fig. 1.

Each phase of risk assessment V-Model shows identification and evaluation of risks at start and end respectively. First phase is about requirements mapping as we know risks are common in requirements phase. In [12], S. Li and S. Duo describe hazards caused by risks in software requirements in models and process.

Second phase is about detail process analysis focuses to identify risks within processes that could be severe in future. Third phase is to identify the risk elements in list form based on possibility in phase two.

Risk Prioritization [13] is one of modern ways of risk assessment in software development. X. Lin, D. Mingrong in [14] focuses to indexing of risks to gain more accuracy in risk assessment. In third phase we assume to prioritize risks. Solving risk items is also a critical part of software development. Phase five focuses on difficulty of risk solution by means of risk occurrence probability. We can identify solution difficulty by applying weights.

Phase six focuses towards use of matrices to be applied to measure risks. As described by X. Lin and D. Mingrong in [14], there are different methods to measure risks but accuracy is not possible from all of them. Therefore matrices must be selected basis on type of risk. The seven and last phase is to calculate and risks and takes measures to solve them.

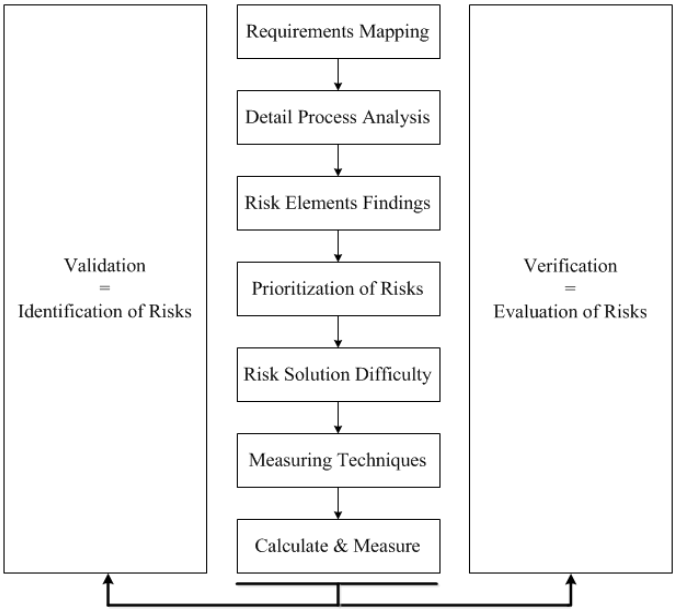


Fig. 2. Phases of Risk Assessment V-Model

In this section we discussed our risk assessment methodology and importance of phases to be used in our model with respect to V-Model methodology. In next section we will discuss our model and its phases in detail with application

IV. RISK ASSESSMENT V-MODEL & APPLICATION

As we discussed before our model is methodological based and its phases overview. In this section, we will discuss its aspects in detail as The Risk Assessment V-Model is shown in Fig. 2.

A. Requirement Mapping

Using requirement mapping is one of the good ways to eliminate risks at initial level. But the most challenging task is to map requirements in right way. As we know use case diagrams are considered to be the best way to map requirements in visual form. In [15], F. L. Siqueira and P. S. M. Silva propose to transform stakeholder requirements into software requirement to ensure requirement refinement. As we know use case diagram is best way to map user requirement so in first phase we propose to evaluate initial risks using use case diagram.

In our strategy, we assume that if risks are highlighted in use case diagram it will be more convenient way to evaluate risk elements at requirement level. Fig 3 shows the use case diagram of sale process. In this figure, when buyer calls "do sale" process and other sub system or provider calls for validation of sub processes that are connected to sale process. This will help software engineers to visualize possibility of initial risks at requirement phase.

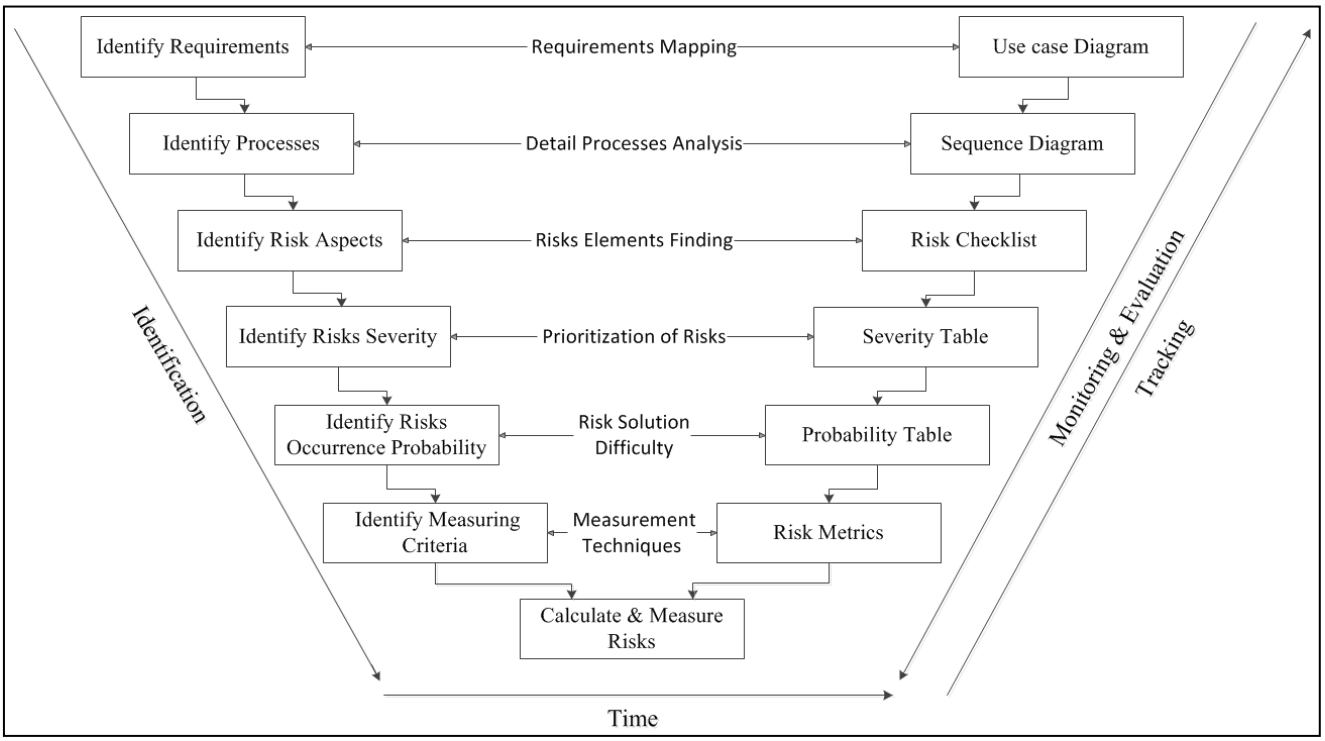


Fig. 2. Risk Assessment V-Model and its phases showing identification and evaluation of risks with risk tracking.

B. Detail Process Analysis

In this phase, we do detailed analysis of business processes. As we know, the error occurs between process can lead to be failure of software project therefore it is needed to identify risk elements within processes so they could be handled in future development. Sequence Diagram is considered to be best to visualize business processes.

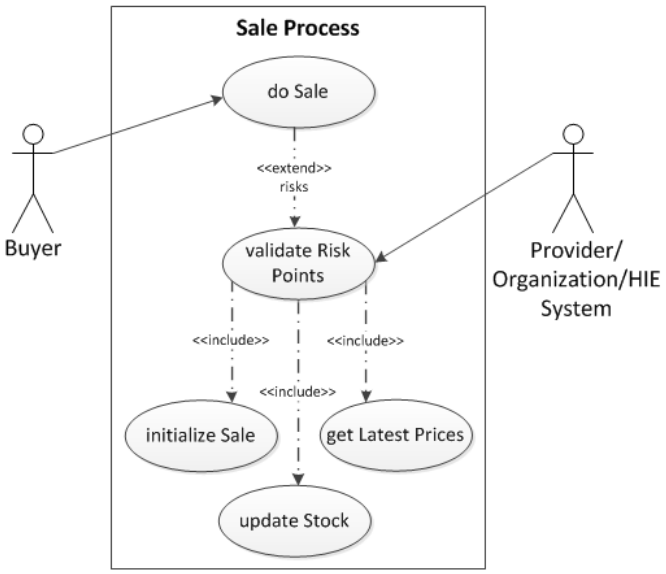


Fig. 3. Fig. 3: Use case Diagram of Sale Process

In [16], P. Luo and Yang Hu do analysis of system risk and identification using event based sequence diagram. A. Refsdal and K. Stolen also describes in [17] usefulness of sequence diagrams to make trust worthy risk analysis because they can be understood by users, decision makers, engineers and parties involved in risk analysis.

We use "do sale" process sequence diagram for detail process analysis. Fig. 4 shows the sequence diagram of customer buying sale item. In this figure we assume that those obstacles which can affect do sale process as risks. If we point out these threat points in sequence diagram it could be helpful for developers and software engineers to deal with upcoming threats. In this figure we use "TP" as threat point where there is a possibly of likelihood of occurrence. For example threat point 3 refers to find an item in the stock. There is possibility of non estimated products in stock, out of date prices in stock, connection failure between prices API and stock database. These are some technical risks in sale process. Besides this, there are also performance risks like time taken in processing. As we know there is internet connection involve for updating price between stocks and prices API.

C. Risk Element Findings

In this phase, we identify risk aspects from threat points as we discussed in detail process analysis. For this purpose we make checklist of the risks we find in sale process and other findings affecting process.

Making risk checklist is one of tmodern ways to make assessments. C. Lopez and J. L. Salmeron [18] empathize on the risk strategies to reduce software risks. They gather data about risks in software projects based on average and make a checklist. They also focused on suitability of this strategy for managing each risk in effective way. M. Keil et al also in [19] make investigation on different software projects and they propose that risk checklist help to identify more risks then they would find without checklist strategy.

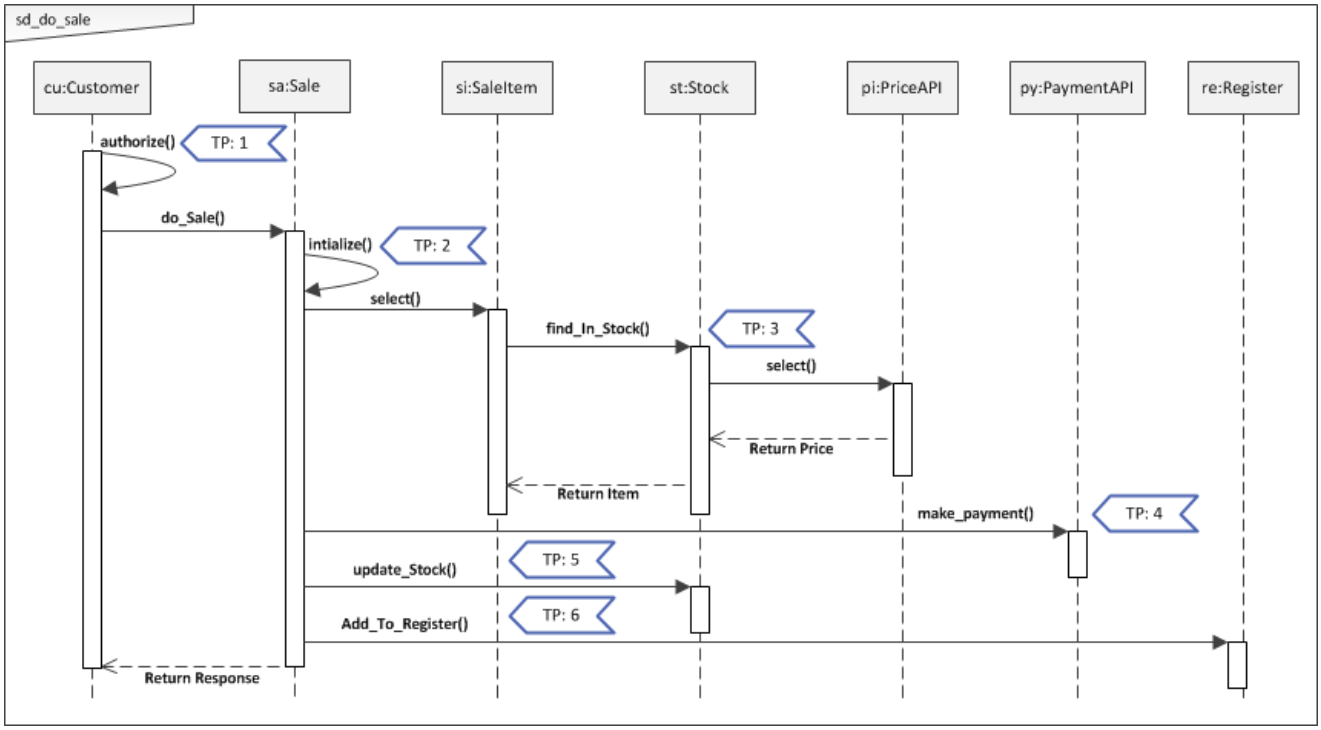


Fig. 4. Risk Assessment V-Model and its phases showing identification and evaluation of risks with risk tracking.

The checklist given in Table I is a sample checklist for sale process. In this checklist we also listed risk items related to performances, interfaces, compatibility measures as well as technical risks. We also mention some requirement mismatch risks compared to use case diagram in Fig 3. Because our model ensures traceability so we also try to cover use case diagram risk points.

TABLE I. SAMPLE RISK CHECKLIST FOR SALE PROCESS

	Risk Checklist
	Sale Process risks check list
1	Non estimated products
2	Out of date prices
3	Authentication security
4	Connection availability
5	API response time
6	Process works as requirments
7	Failure to update stocks
8	Process works stable as requirements
9	Command failure
10	Vulnerability in process execution

D. *Priortization of Risks*

Risks are prioritizing according to some defined criteria. In this section, we use severity scales for risk prioritization. Severity refers to the potential effect of failure over the process. Mostly rating scales are used for finding severity of risks. Risk severity is one of the modern ways to prioritize risk. The general severity criteria for prioritizing risks are given in Table II.

TABLE II. RATING CRIETERIA FOR SEVERTY

Rating	**Description**	**Severity of Effect**
5	Critical	give customer loss
4	High	Function loss
3	Moderate	Overall performance degradation
2	Low	Reduced performance
1	None	Failure wouldn't noticeable

E. *Risk Solution Difficulty*

This section refers to effort required to avoid risks. Some risk are so twisted they are difficult to find and solve because they occurred frequently. For example in sale process there are more chances of command failure due to high data transfer and internet communication and make system stress less.

TABLE III. RATING CRITERAIA FROM PROABABILIY

Rating	**Description**	**Probability Estimate**
5	Almost certain	26% to 99%
4	Likely	11% to 25%
3	Occasionaly	6% to 10%
2	Unlikely	1% to 5%
1	Seldom	<1%

Therefore, it is more difficult to solve such risk that comes frequently. In this section we identify complexity as probability of risk occurrence and prioritize them according to defined criteria.

In Table III the rating defined 1 to 5 as probability of occurrence for complexity of solution. This is generic criteria and can be changed according to different projects.

F. Measurement Techniques

After getting Risk matrix from severity and complexity table the most important thing is to use right technique for measurement. There are lots of measurements tools widely use for risk analysis. FMEA in [20] describes the RPN (Risk Priority Number) measure to find criticality of risk. They describes each RPN as

*Severity * Occurrence * Detection = RPN*

Besides this, there are many graphical user interface tools like excel to graphical measure the risks hazards in proper ways

Thirdly, decision trees are also common to measure criticality of risks because they can give more visualization of tracing then others.

G. Calulate and Measure

To measure and visualize risks we assume the following data based on Sales process. Table IV shows the risk assessment from Table I risk checklist.

TABLE IV. RISK ASSESSMENT TABLE BASED ON SALE PROCESS

	Risk Assessment Table		
	Sale Process risks check list	***Severity***	***Probability***
1	Non estimated products	Moderate	Likely
2	Out of date prices	Moderate	Occasionally
3	Authentication security	High	Unlikely
4	Connection availability	Critical	Seldom
5	API response time	Low	Seldom
6	Process works as requirments	Low	Unlikely
7	Failure to update stocks	High	Occasionally
8	Process works stable as requirements	None	Seldom
9	Command failure	None	Seldom
10	Vulnerability in process execution	Moderate	Unlikely

From Table IV, we visualize risk measures using Microsoft Excel Tool. The Figure 5 shows assessment Graph for Sale Process.

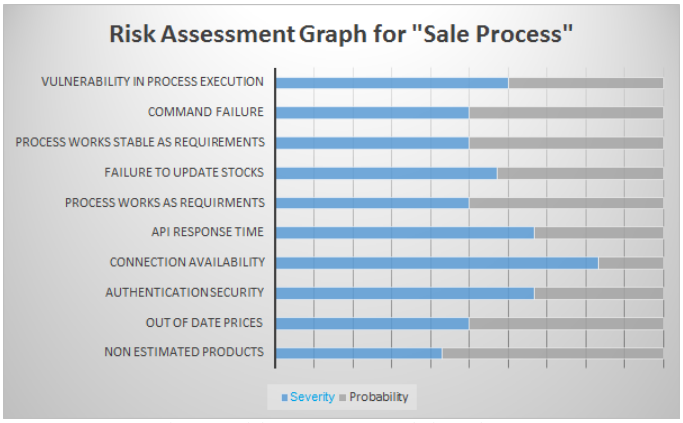


Fig. 5. Risk assessment graph for Sale Process

In Fig. 5, on left side it shows the risks in sale process whereas blue lines show severity of risks as rating whereas gray lines show the occurrence probability of risks as probability.

We can also visualize our calculations using design tree as shown in Figure 6. We use classification to make the following decision tree using Weka tool from data in Table IV.

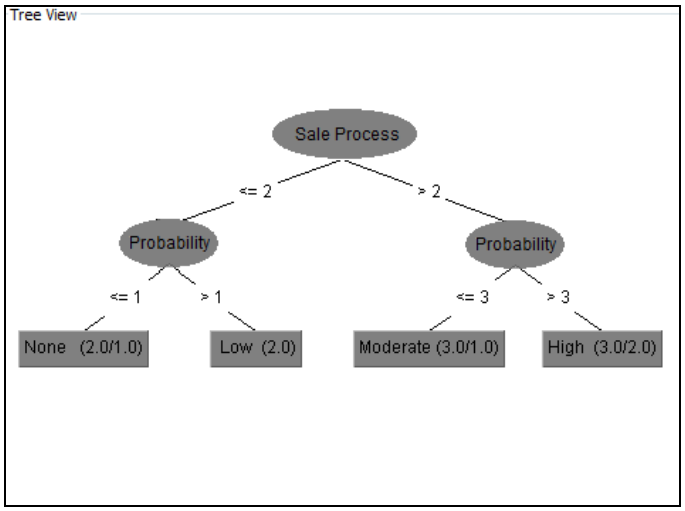


Fig. 6. Decision tree generated by Classification

V. CONCLUSION

In this paper, we propose an assessment model for software risks using V-Model Methodology. We hope this model not only helps software outsourcing companies to do assessment of risks in the software solutions with maximum accuracy but also facilitate its usability and extendibility according to their own software assessment criteria.

Furthermore, based on research data our assessment model identifies risks coming at each phase of development and focuses to evaluate them accordingly. The big goal of our assessment model it ensures the traceability of risks founds in software solution as all assessments found in each phase are linked to previous phase.

REFERENCES

[1] S. Nurazlina, M. Nuridawati1, A Visualization Tool for Risk Assessment in Software Development, International Symposium on Information Technology 2008, Vol 3, Conference code:74115

[2] M. Boban, Z. Poagaj, H. Sertic, Strategies for Successful Software Development Risk Management, Management 2003, Vol 8, pp. 77–91

[3] Z. Kremljak, C. Kafol, Types of Risks in a System Engineering Environment and Software Tools for Risk Analysis, Procedia Engineering 2014, Vol 69, pp. 177-183

[4] W. Han, S. Huang, An Empirial Analysis of Risk Components and Performance on Software Projects, Journal of Systems and Software, Vol 80, pp. 42-50

[5] Y. Wang, S. Fu, T. Zhong, Ranking Software Risks Based on Historical Data, Advances in Intellignet and Soft Computing 2012, vol 169, pp. 393-398

[6] J. Li, M. Li, D. Wu, J. Song, An Integrated Risk Mesurement and optimization Model for Trustworthy Software Process Management, Information Sciences 2012, vol 191, pp. 47-60

[7] K. Georgieva, Ayaz Farooq, R.R. Dumke, Analysis of the Risk Assessment Methods, Lecture NOtes in Computer Science 2009, Vol 5891, pp. 76-86

[8] M. R. Naeem, W. Zhu, A. A. Memom, New Approach for UML Based Modeling of Relational Databases, International Journal of Computer Science and Telecommunications 2014, vol 5, Issue 5, pp. 18-23

[9] J.A. Cruz-Lemus, M. Genero, M.E. Manso, S. Morasca, M. Piattini, Assessing the Understandability of UML Statechart Diagrams with composite states – A family of Empirical Studies Empirical Software Engineering 2009, pp. 685–719

[10] J.V. Moll, J. Jacobs, R. Kusters, J. Trienekens, Defect Detection Oriented Lifecycle Modeling in Complex Product Development, Information and Software Technology 2004, vol 46, Issue 10, pp. 665-675

[11] IEEE Standard for Software Verification and Validation[S] IEEE, Std, 1012 2004

[12] S. Li, S. Duo, Safety Analysis of Software Requirements: Model and Process, Procedia Engineering 2014, vol 80, pp. 153-164

[13] Z. Salarian, H. Rashidi, Improving Offshore-Outsourced Software Development requirement Risk Prioritization, Scheduling and Prediction with Using a Fuzzy Analytic Hierarchy Process, Int. J Latest Trends Computing 2011, Vol 2, no 4, pp. 478-495

[14] X. Lin, D. Mingrong, A New Risk Evaluation Method for Software Development Project, The 5th International Symposium on Managment of Technology 2007, pp. 874-878

[15] F. L. Siqueira, P. S. M. Silva, Transforming an Enterprise Model into a Use case Model in Business Process Systems, Journal of Systems and Software 2014, vol 96, pp 152-171

[16] P. Luo, Y. Hu, System Risk Evaolution Anayslis and Risk Critical Event Identification based on Event Sequence Diagram, Reliability Engineering & System Safety, vol 114, pp 36-44

[17] A. Refsdal, K. Stolen, Extending UML Sequence Diagrams to Model Trust-Dependent behavior with the Aim to Support Risk Analysis, Science of Computer Programming 2008, vol 74, issue 1-2, pp 34-42

[18] C. Lopez, J. L. Salmeron, Risk Response Strategies for Supporting Practitioners Decision-Making in Software Projects, Procedia Technology 2012, vol 5, pp 437-444

[19] M. Keil, L. Li, L. Mathiassen, G. Zheng, The influence of Checklists and Roles on Software Practitioner Risk Preception and Decision-Making, Journal of System and Software 2008, vol 81, issue 6, pp 908-919

[20] FMEA-FMECA, FMEA Risk Priority Number (RPN), http://www.fmea-fmeca.com/fmea-rpn.html

Optimization of Neural Network Based on Genetic Algorithm and BP

[1]Shiwei Zhang, [1]Hanshi Wang, [1]Lizhen Liu, [1]Chao Du, [2]Jingli Lu
[1]Information and Engineering College, Capital Normal University, Beijing 100048, China
2Agresearch Ltd, New Zealand

***Abstract*—In order to improve the intelligence, high efficiency, humanization of the type of the search and eliminate games, and also to improve the search performance and rule out the accuracy of the target during intelligent games running. This paper puts forward a comprehensive method that combines Genetic Algorithm, Neural Network and Back Propagation (BP) to solve the insufficient of computing power and low efficiency by using a single algorithm in Intelligence games. In this method, Genetic Algorithm will be used in weight training of Neural Network first of all. It will not stop iterating until Genetic Algorithm evolves into a certain degree or network errors satisfies the requirements, and delivers the best chromosome we get to Neural Network. Then BP trains the data that runs through the Neural Network, which is Neural Network's second training. Finally, the paper applies the new way in the Mine Clearance experiment. By comparing this experiment with only using Genetic Algorithm or Neural Network, finds out the proposed method significantly improves the minesweepers accuracy.**

Keywords—Genetic Algorithm; Neural Network; Back Propagation

I. INTRODUCE

The processing efficiency of games has always been a topic of discussion among users. One of the factors that influences efficiency of games is the use of the algorithm. Genetic algorithm is an efficient parallel global search algorithm, which is simple, universal, robustness and suitable for parallel distributed processing and has potential to be widely utilized[1]. Neural Network is widely connected by a large number of neurons, forming a complex network system which can massively process concurrently and store distributed information[2]. Each algorithm has its own advantages, however, when the algorithm is separately used in games there will be some problems we cannot avoid, such as the inferior computation, low efficiency and so on.

In order to solve the problem of low efficiency caused by single algorithm, an improved algorithm was put forward in this paper. The algorithm optimizes neural network through the genetic algorithm, then trains the training set by BP, which makes the Neural Network get secondary training. Then apply the data again after the second training to minesweepers. The new algorithm takes advantage of the characteristics of each algorithm effectively, which improves the running efficiency of the game.

The remainder of the paper is organized as follows. In Section 2, we introduce some related works. The introduction of the basic algorithm is proposed in Section 3. Experiment and analysis reported in Section 4. The last part is the summary of the experiment.

II. RELATED WORK

Here are two classical algorithms of Artificial Intelligence, Neural Network and Genetic Algorithm. Application of these two algorithm is quite extensive. But Genetic Algorithm cannot avoid such problems as the rough datasets. At the same time, Neural Network also exists two inherent problems: one is that it is easy to fall into local minimum, and the other is slow convergence speed [3]. BP is also with the same problem. A new algorithm was put forward in this paper synthesizing these three algorithms with the characteristics of their respective.

For the generation of new algorithm, the paper mainly has two aspects: Firstly, the paper combines the Genetic Algorithm with Neural Network. The combination mixes the Neural Network's rapid parallelism and global searching capability of Genetic Algorithm together[4], which solves the randomness of the Neural Network structure and parameter design and the shortcoming of depending on person's experience, and optimizes the Neural Network .in other words, trains the Neural Network. Secondly, use BP to train the first training data of the Neural Network which makes the Neural Network get secondary training.

III. ALGORITHM

A. Genetic Algorithm

First of all, Genetic Algorithm generates initial population is the basic idea of, then according to the objective functions of problems structures the fitness function. With stand or fall of fitness, Genetic Algorithm continuously selects and breeds. At last we will get the best individual fitness which is the optimal solution after several generations. That is the basic idea of Genetic Algorithm. That Genetic Algorithm forms the initial population based on the chromosome coding, then exerts a certain operation to it according to fitness on the environment of the individual groups, enabling to realize the evolution process of survival of the fittest.

In the process of achieving the survival of the fittest, Genetic Algorithm establishes an iterative process with coding

space instead of the problem space, putting the fitness function as an evaluation criterion, taking the evolution of code groups as the basis, realizing select and genetic mechanism by genetic operation of the individual bit string. Finally, pattern gradually evolved to a better direction and found the optimal solution through these genetic operation.

B. Neural Network

Neural Network is abstraction and simulation to some basic features of human brain or Natural Neural Network[5]. Neural Network is a model to explore and simulate the model of brain nerve system function through modeling and coupling of the basic unit of the human brain–neurons, and to develop artificial systems with learning, thinking, memorize, pattern recognition and other information processing functions. Neurons possess three main functions that are signal input, integration and output.

An important characteristic of Neural Network is that it can learn from the environment, and storage the results of learning to the synaptic connections of the network. Learning of Neural Network is a process. Under the stimulus of its environment, we input some sample models one after another to NN, and adjust the network weights matrix of each layer according to certain rules. The learning process ended when the network weights of each layer all converge to a certain value. Then the generated Neural Network can be used to do the real data classification.

C. BP neural network

BP algorithm is a supervised learning algorithm[6]. It is a multilayer feedforward networks which is composed of back propagation and error correction. BP is composed of two process that includes forward propagation of information and error back propagation. It has characteristics of good self-learning, self-adaption, robustness and generalization [7]. BP network can learn and store a lot of input-output model mapping relation without prior revealing description the mathematical equations of the mapping relationship. The essence of its learning is the minimum value of seeking mean square error function[8]. BP adjust the network weights and thresholds constantly by error back propagation to make the minimum error sum of squares of the network to make the Sum of square error minimum

IV. A NEW ALGORITHM -AGB

A. Optimization of Neural Network Based on Genetic Algorithm

Basing on the global searching capability of Genetic Algorithm and the fast parallel of Neural Network, we integrate Genetic Algorithm and Neural Network by the way of using Genetic Algorithm to optimize the initial weights of Neural Network, which achieves complementary advantages and solves the problem of low efficiency caused due to the disadvantages of Neural Networks and Genetic Algorithms. Learning steps of Genetic Algorithm combines with Neural Network:

1. Set parameters: Set the population size, crossover probability and mutation probability, the network layers, the parameters of each layer such as number of neurons.
2. Initialization: Initial population generated randomly $P=\{x_1, x_2, x_3,\ldots\ldots x_n\}$,n is the number of connection weight. P is comprised of n weight vector and a threshold vector. The weight vector is real numbers vector, the threshold vector is an n-dimensional real numbers vector.
3. Calculation: Update weights and adaptive scores of Neural Network according to the environmental information
4. Rank: Sort According to the fitness from small to large, that is, sort it according to the chromosome from good to bad.
5. Keep the best chromosomes.
6. The process of cross and evaluation: Do crossover operation and evaluate the new individual group.
7. Variation and evaluate operations: Do mutation operation and evaluate the new individual group.
8. If the time beyond the prescribed clock, then turn into the next generation and go to the step 3, until the users launch program.

In conclusion, that is, apply Genetic Algorithm to training the weight of Neural Network, and stop iteration when Genetic Algorithm evolved to a certain extent or the network error meet the requirement, then deliver the optimal chromosome to Neural Network[9]. Make Neural Network getting trained until the better result is got.

B. The Formation of AGB

Get data from the Neural Network and randomly assign weight for these nerve cells, then send an insert mode to input terminal of network, and observe the error value between actual network output value and the expected output value. We Use the error value to adjust the weight which comes from the output of the hidden layer so that the output value can be more close to the right answer when we use the same input terminal into the network again. It can do the same thing for the front of the layer after adjusting the weight of the current layer. So to ensure the weight of all layers have been adjusted, we push in the direction of the input layer starting from the output layer one by one, until you reach the first hidden layer location. If the work was finished, the actual network output value will be more close to the expected output values when the insert mode is send to network input terminal. All different input patterns need to be repeated many times according to such a process, until the error reduced to the acceptable within a limit value what we deal with. By this time, it can be said that the network has been trained well, this is the second optimization of the Neural Network. The following step is training the data of Neural Network by BP:

1. *Create network.*
2. *Initialize the weight to the small random values that average is almost to zero.*
3. *repeat steps (i) to (v) for each pattern:*

i) Put the training set into the network, and calculate the actual output ;

ii) To calculate the output error between actual output and desired output ;

iii) Adjust the output layer weights. Repeat steps (iv) and (v) for each hidden layer ;

iv) To calculation the error of the hidden layer neurons;

v) Adjust the weights of hidden layers

4. Repeat step 3 until the error in the step (ii) is in the acceptable range.

The data that trained by BP can adapt to the surrounding environment more preferably. Putting the data trained by BP into Neural Network of minesweepers makes the minesweeper can find and remove the target more quickly.

Summing up the above, the formation of the kind of new algorithm combines with the characteristics and advantages of Genetic Algorithm, Neural Network and BP.

V. The Experiment and Analyze

A. *Initialization Parameter Settings*

Before the experiment, the parameters must be initialized. The following table shows the initial parameter values of this experiment:

TABLE I. The Initial Parameter Table

ParamName	Set Value
iFramesPerSecond	60
iNumInputs	4
iNumHidden	1
iNeuronsPerHiddenLayer	10
iNumOutputs	2
dActivationResponse	1
dBias	-1
dLearningRate	0.5
dErrorThreshold	0.003
dMaxTurnRate	0.3
dMaxSpeed	2
iNumMines	40
iNumSweepers	30
dCrossoverRate	0.7
dMutationRate	0.1
dMaxPerturbation	0.3
iNumElite	4

The nature of the Genetic Algorithm is a series of operations on chromosome patterns, that is, pass from the excellent pattern to the next group, then use crossover operator to pattern mutations. Therefore, selection, crossover and mutation are three genetic operators of Genetic Algorithm. The genetic algorithm can be set through the three parameters. Input and hidden layer, hidden layer neurons and output response value, motivation can be set in neural network. The rest of the parameters are conducted on experimental ontology and experimental target.

B. *The results of experimental and analysis*

1) *The results of experimental*

Run the experiment and choose one of the three algorithms on the interface, then show the highest fitness scores of genetic algorithm and the average fitness score line chart through the keyboard control "F". With time, the algebra of machine get changed, also the highest fitness and the average fitness changes. Here is the line chart of GA-AN, AGB in the case of initial parameter value.

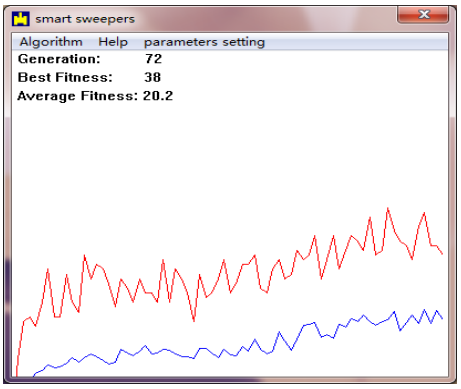


Fig. 1. The line chart for 72 GA-AN algebra

As can be seen from fig1, there is a growing trend in highest fitness score and the average fitness score with the increase of learning algebra generations.

The fitness score and the average fitness score showed a rising trend, but there will be fluctuations up and down, the reason for this is that the hybrid between chromosomes and its variations exist great randomness which may lead to a relatively good chromosomes instead become a general or poor chromosomes.

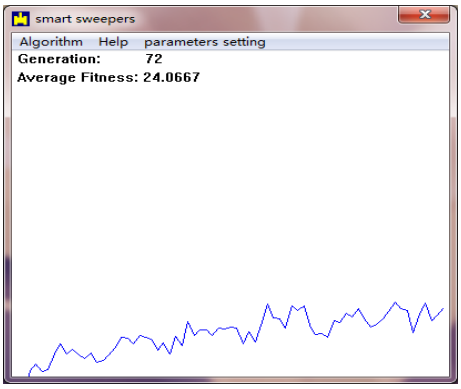


Fig. 2. The line chart for 72 AGB algebra

In the case of parameter value unchanged, considering the data is trained well by BP, then applied to the machine again, the adaptability of machine should be more average. There is no relative the highest and lowest fitness. Therefore, only the average fitness is set for BP in this experiment.

AS can be seen that there is no obvious peaks and valleys in fig2. Although the average fitness scores appear the phenomenon of fluctuated, but it is relatively stable on the whole, the change is not obvious. The reason is that after training Neural Network by BP, each connection weight has been adjusted for the best value in the range of allowable error. At this time the learning of BP Neural Network has been established. After applying the training set to minesweeper games again, the average score of the minesweeper would not change too obviously.

2) Comprehensive comparison and analysis of the experimental results

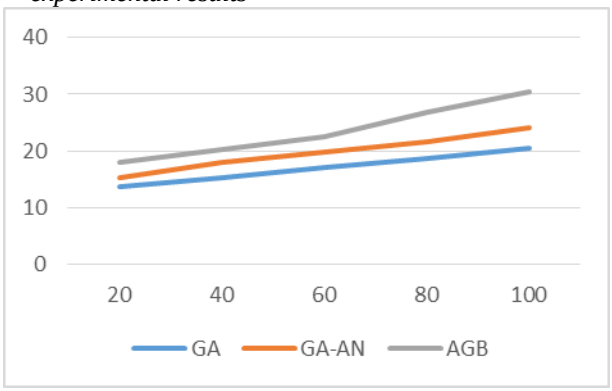


Fig. 3. Comparison of three methods of calculation results

In the figure, the abscissa is the algebra of minesweepers, and the ordinate is the average fitness scores of three algorithms. Genetic Algorithm, GA is used alone in the experiment of minesweepers, the disadvantage is that the data set is rough. Relative to the algorithm which is combined with Neural Network, with the increase of algebra, although average fitness scores increases, it always is not higher than GA-AN. Likewise, on the basis of the two algorithms, the average fitness scores of Neural Network trained by BP will be higher. Therefore, the highest is the average score of AGB algorithm.

a) In the case of continuous changing the same parameter value, comparing the three algorithms:

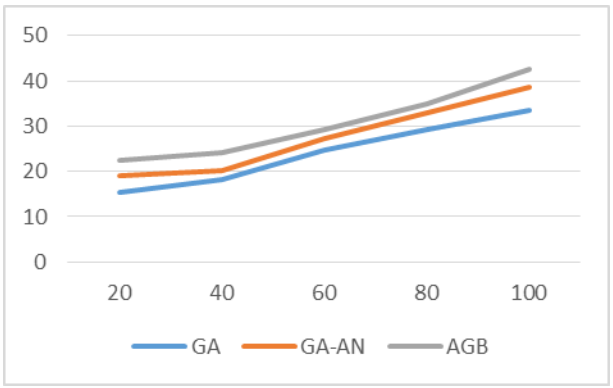


Fig. 4. Gradually increasing the number of mines

From the fig4 we can see that the average score of three adaptive algorithms increases with the increase of the number of mines. This is because minesweepers need to judge more and more information of the surrounding environment with the increase of the mines. In the iterative process of Genetic Algorithm, after a good learning of the minesweepers, put the information into the Neural Network and change weights of the Neural Network to increase the adaptability scores. So the point of training set increased. Apply data that was trained well to the minesweepers, the average fitness scores increased compared to the previous.

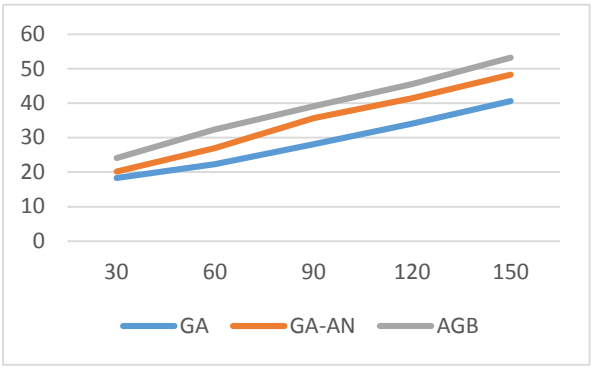


Fig. 5. Gradually increasing the number of minesweeper

In the figure5, the abscissa is the algebra of minesweepers, and the ordinate is the average fitness scores of three algorithms. What can be seen from the figure is that the average fitness score of algorithm increased with the increasing number of minesweepers. This is because the increasing number of minesweepers means that the number of population increased relative to the Genetic Algorithms, which lead to the increasing number of eugenics individual that chosen from population by Genetic Algorithm, so that the data which is applied to the Neural Network weights training becomes more variable and the input information gets more and more, then more information is obtained by minesweepers. So the average fitness scores increased with determining the location of the target more accurate. Therefore, BP data is extracted more optimized. Data is applied to minesweepers after training, then the final result is that average fitness score increase.

b) Comparing average fitness scores of three algorithm through changing four parameter values, the comparison result is in the following table.

TABLE II. THE AVERAGE FITNESS SCORE UNDER DIFFERENT PARAMETERS IN THE SAME ALGEBRAIC

Parameters / algorithm	Mines number (30)	Minesweeper number (50)	CrossoverRate (0.4)	Mutation rate (0.4)
GA	17.20	19.86	17.83	29.07
GA-AN	20.33	24.28	19.03	32.47
AGB	22.17	25.66	24.33	36.39

Results analysis:

The results in the table is the average fitness score of three algorithms in the 72-generation with changing only one parameter.

As can be seen from the table, the number of minesweepers reduced relative to initial parameter, and the highest fitness score decreased either. The fewer the number of minesweepers

is, the lower the highest suitability score is. This is due to the number of minesweepers reduced, the amount of information network input greatly reduced, thus affecting the output of the Neural Network.

The more the number of mines, the highest fitness score and the average fitness score are higher. This is due to minesweepers cannot judge the location of mines accurately, so that the minesweeper cannot remove mines preferably when the number of mines are less. On the contrary, when the number of mine are more, minesweepers receive much more information, and the input of Neural Network increase, so the minesweeper can learn timely, and it can judge the location of mines more accurately after learning. At last, the minesweeper have a higher chance to remove mines, thus increasing the adaptation degree scores.

Hybridization, i.e. the cross hybridization, refers to two overlapping chromosomes to cross some of its genes in some manner, to form a new individual[10]. The hybrid rate become low, and the number of iterations will reduce, so the chromosome that was delivered to Neural Network is not the optimal, which affects fitness scores.

Mutation , replace some genic value of locus in the individual chromosomes encoding string with the other on the loci alleles, to form a new individual[10]. Mutation rate relative to the initial parameter value becomes higher, which will lead to the number of iterations increasing. The iteration will stop when the evolutionary computation of Genetic Algorithm reach a certain degree, then pass the optimal chromosome to neural networks, until a better result. Therefore, the higher mutation rate is, the higher the fitness score will be.

On the whole, comparing three algorithm with each other, no matter what changes, the average fitness score of AGB is the highest, and the next is GA-AN, the last is GA. The reason is that Neural Network is based on Genetic Algorithm, which makes the Neural Network is superior to Genetic Algorithm. AGB is to optimize Neural Network that was trained by Genetic Algorithm, so that AGB is superior to the first two, therefore, there is no doubt that its average fitness score is higher than the first two.

VI. CONCLUSION

The experimental results can be seen from the figure contrast. Comparing three algorithms with each other, GA average fitness score is lower than the GA-AN, the highest fitness score is AGB trained by BP algorithm.

In genetic algorithms, the chromosomes of minesweepers can crossover and mutation to generate new individuals to produce a generation of evolution, so as to make it more "intelligent." Neural network is based on the Genetic Algorithm, it makes use of the characteristic of global search to get an initial weight matrix, an initial threshold vector and the optimal chromosome, then apply it to the minesweeper to update the “brain” of the minesweeper with purpose by collecting the surrounding information. So that the minesweepers can be more intelligent and remove mine quickly and accurately. BP update the minesweepers’ weight of neural network by training the data of Neural Network, which increase the efficiency and accuracy of minesweepers. Therefore, minesweepers can find and remove the target more quickly and efficiently in the process of superposition of the algorithm step by step.

ACKNOWLEDGMENT

This work was supported in part by National Science Foundation of China under Grants No. 61303105 and 61402304; the Humanity & Social Science general project of Ministry of Education under Grants No.14YJAZH046; the Beijing Educational Committee Science and Technology Development Planned under Grants No.KM201410028017; Academic Degree Graduate Courses group projects and the Beijing Key Disciplines of Computer Application Technology.

REFERENCES

[1] X. Wang, "Genetic algorithm and its application," Mini-Micro computer system, 1995.

[2] X.-l. he, "The similarity comparsion of Nerral Network and Genetic Algorithm," Network and Information, 2010.

[3] X.-j. GAO, J. ZHANG, Y. HONG, and G.-x. CHENG, "The Research on Technology Based on Genetic Algorithm and Neural Network," Equipment Manufacturing Technology, vol. 2, p. 007, 2010.

[4] Y. Tai-shan, "The Research on Technology Based on Genetic Algorithm and Neural Network," Network and information, 2007.

[5] P. Chunhua, "Shallowly Discuss the Artificial Networks," Computer Development, vol. 5, 2009.

[6] F. Xing-juan, "Developing the Superhighway ITS System on the Client End by Using Delphi [J]," Sci-Tech Information Development & Economy, vol. 3, p. 137, 2006.

[7] J.-J. Lu and H. Chen, "Researching development on BP neural networks," Control Engineering of China, vol. 13, pp. 449-451, 2006.

[8] L.-h. Jia and X.-r. Zhang, "Analysis and Improvements of BP Algorithm," Computer Technology and Development, vol. 10, pp. 107-113, 2006.

[9] H. Q. DENG Zheng-hong, Zheng Yu-shan, "Two-Times Training Algorithm of Neural Network Based on GA," Institute of Computer, pp. 252-254, 2005.

[10] J. Ge, Y. Qiu, C. Wu, and G. Pu, "Summary of genetic algorithms research," Application Research of Computers, vol. 25, pp. 2911-2916, 2008.

Performance Evaluation of Network Coding-Based Convergecast in Realistic Wireless Sensor Networks

Chun'e Ku[1], Hengyi Zhang[1], Xiaoqiu Shi[1*], Kezhong Jin[1], Zhenzhou Tang[1,2]
[1]College of Physics and Electronic Information Engineering, Wenzhou University
Wenzhou, Zhejiang Province, P. R. China
[2]School of Information and Communication Engineering, Dalian University of Technology
Dalian, Liaoning Province, P. R. China
*Corresponding Author

Abstract—Network Coding has attracted significant attention in recent years, and it has been widely applied in broadcast, multicast and unicast. Nevertheless, the study on network coding in convergecast is still in the theoretical analysis and the simulation stage. In order to investigate the network coding gain of convergecast in realistic wireless sensor networks, a network coding-based convergecast scheme is realized in this paper. The experiment is conducted in the basic size of convergecast whose size is limited below four, which is for the purpose of eliminating as much interference from the large size of convergecast network as possible. Performed in traditional convergecast scheme and network coding convergecast scheme, the experiment on collection rate and energy consumption are evaluated under different link delivery ratios and different scales of convergecast network. The experiment results show that network coding is certain to introduce benefit in the case of poor link equality. However, it is inferior to the traditional scheme when the wireless link delivery ratio is high enough due to the conscious overhearing and the prerequisite of successful decoding.

Keywords—wireless sensor networks; network coding; convergecast; collection rate; energy consumption

I. INTRODUCTION

Network coding (NC) is an information exchange technology which combines routing and coding. NC is put forward based on network information theory in [1], in which, butterfly network is researched, in addition, network throughput is improved as multicast information delivery ratio can be promoted to the top level with node's coding. The theory overthrows the view that processing transmission data in interior nodes can't improve the information delivery ratio. So improving the traditional information processing ways in communication network becomes a hot study. NC theory is replenished and completed in [2-3], the theory gives NC with high performance and low complexity. NC's application area is extended from wired multicast to wireless network in [4-11]. Contrast to wired network, wireless network has special properties, such as broadcast characteristic of wireless channel, non reliability of link and so on, and those properties can make NC play a more important role in wireless network area.

Currently, researches of NC in wireless network are concentrated on unicast [4] and multicast/broadcast [5-6], in [4] and [5-6], data transmission ratio can be improved. The data transmission of wireless network can be split to three ways: unicast of one to one, multicast/broadcast of one to many and convergecast of many to one. In wireless sensor networks (WSNs), convergecast network play the role of collecting information, specifically, the sensor nodes collect the environment information around, then transmit the information to the sink node through several hops as shown in Fig.1.

Several papers have been found to apply NC to convergecast in WSNs so far. The reliability problem of NC in convergecast is revealed from the view of theory and simulations [7-8]. A resolve scheme of convergecast in WSNs is come up with in [9], and the scheme can improve the network reliability and decrease energy consumption, but a re-transmission slot is added. NC can be used to improve the network reliability in [10], while there are two limitations, one is that network delay is increased, and the other is high energy costs and poor adaptability. Sensecode protocol is proposed in [11], and it can improve the reliability from 15% to 20%, but more energy is consumed. From those papers, it is obvious to see that NC can make contribution to convergecast in WSNs, but, at the same time, NC also does harm to the convergecast network in some impacts, such as, in NC-based convergecast, the encoding nodes need to continue to overhear the wireless links when other nodes overhear more original packets for encoding operation, the way of several sides overhear in the same time will consume more energy than traditional convergecast. In addition, NC has cliff effect, that is, once decoder can't receive enough combined packets, all original packets which participated in encoding wouldn't be recovered out.

In order to research the influence of NC to convergecast, a network coding-based convergecast scheme is realized in a basic convergecast network in this paper. The performance of NC will be evaluated in realistic WSNs with convergecast Network applied in the case of the practical situations are distinguished. With the experiments, the performance of NC

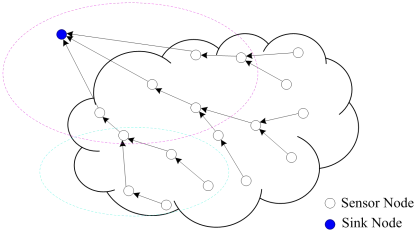


Fig.1 Convergecast scheme in wireless sensor networks

under different link qualities and different sizes of convergecast network is tested and contrasted with the traditional convergecast scheme. The results will bring a new idea about how to apply NC effectively in convergecast.

The rest of this paper is organized as follows: Section 2 presents the system mode and the network coding processing scheme. In Section 3, we will evaluate the performance of NC-based convergecast compared to traditional convergecast on collection rate and average energy consumption. Conclusions are given in section 4.

II. SYSTEM MODELS AND NETWORK CODING WORKING SCHEME

A. System models

The combination of intra-flow NC and converge processing is feasible, which has been proofed in [7]. There are many common characteristics between intra-flow NC and convergecast processing. In Fig.1, packets converge together at the interior nodes and sink node, when the packets collected is enough, the sink node can recover the original packets. So intra-flow NC can improve the security of information benefit convergecast.

In the processing of tree-based convergecast in WSNs, many sub trees can be found, as shown in Fig.2(a). The sub-tree contains n leaf nodes (L_i, i = 1, 2, …, n), n interior nodes (M_i, i = 1, 2, …, n) and one root node (*Root*). L_i and M_i is one-to-one relationship, that is, every interior node M_i only has one child node L_i, and every child node L_i only has one father node M_i. In fact, this sub-tree model is the basic component of traditional WSNs convergecast. The network coding-based convergecast model is also based on this sub-tree model, we called Converge-Tree (CT_n), in which, n represents the number of leaf nodes and interior nodes, as shown in Fig.2(b). In the CT_n, leaf nodes are treated as original nodes (S_i, i = 1, 2, …, n), interior nodes are treated as encoding nodes (E_i, i = 1, 2, …, n), and root node is treated as decoding node (D). The only difference between CT_n and the traditional convergecast sub-tree model is that encoding node will overhear the packets from other neighbors' child nodes and encode them.

In CT_n, original packets are generated periodically from node S_i (i = 1, 2, …, n) and sent to node E_i (i = 1, 2, …, n), asshown in Fig.2(b). Node E_1 can receive the packets from node S_1, and overheard the packets from node S_2. E_n can receive the packets from S_n, and overhear the packets from

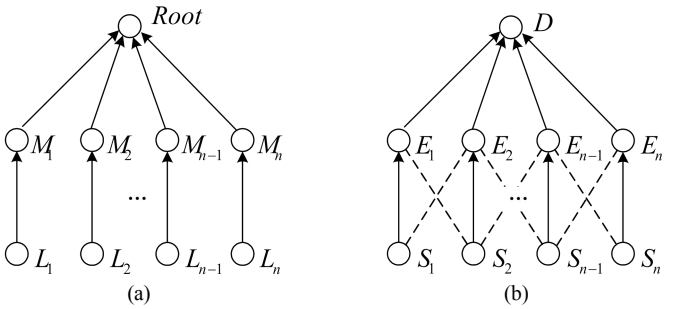


Fig.2 Convergecast structure. (a) Sub-tree model of traditional convergecast; (b) network coding-based convergecast model

S_{n-1}. E_i (i = 2, 3, …, n-1) can receive the packets from S_i (i = 2, 3, …, n-1), and overhear the packets from S_{i-1} (i = 1, 2, …, n-2) and S_{i+1} (i = 3, 4, …, n). E_i can combine the received and overheard original packets into one packet through linear network coding, and then send the combined packet to the decoding node. When receiving all the combined packets, the decoding node will process them and recover the original packets.

In view of the purpose of pursuing the influence of network coding to WSNs convergecast, in the realization of CT_n, fixed route is adapted to eliminate the influence of routing protocol. It's clear that the CT_n needn't to change the topology of traditional convergecast sub-tree model, but a network coding layer is added only. So, the combined technology of the CT_n and network coding can be used in practical environment.

B. Network coding working scheme

Linear network coding is adopted to perform encoding operation. In CT_n, node E_i (i = 2, 3, …, n-1) will receive or overhear at most three original packets from S_{i-1}, S_i and S_{i+1}. The way of the linear network coding as follows:

$$X_j = \sum_{i=1}^{3} g_j^i P_i \qquad (1)$$

where P_i(i = 1, 2, 3) is the original packet, g_j^i is the coefficient selected from Galois field, and X_j is the combined packet. The addition and multiplication in the formula should obey the finite field addition and multiplication. n encoders will generate n combined packets, and then the coded packets will be forwarded to the decoder D.

Gaussian Elimination will be used to decode the combined packets at the decoder [3]. When abundant combined packets are received, the original packets can be recovered. Obviously, in consideration of the processing of n dimensions network coding, n coded packets, at least, are needed for the decoder to recover the original packets, but due to the the packets loss, the total of the coded packets the decoder received will below n, then the decoding processing will fail.

III. EXPERIMENT RESULT

For eliminating as much interference from large scale of convergecast network as possible, the experiment is conducted in three more basic scale of convergecast networks, which are CT_2, CT_3 and CT_4. It is easy to get them from Fig.2(b). CT_2, CT_3 and CT_4 are elements of S_n when n equals to 2, 3 and 4 respectively. More than 12000 packets are sent out by original node under different link delivery ratios in each scale of convergecast structure, which means more than 12000 experiments are going to perform repeatedly in each convergecast network. TelosB nodes are used as the hardware platform. In the node, TinyOS is used as the operation system, which is developed by UC Berkeley University. CC2420 chip is used as the wireless module, and the power of the wireless radio is turned up to the maximal so that the nodes in the network can sense each other. In addition, because of the limited memory, the node will abandon the received packets at random. In the experiment, the link delivery ratio is under human control, simple time division multiple access medium

access control protocol is employed to synchronize all the nodes in the network and coordinate the access of wireless media. The duration of each slot is 100 ms. Fixed coding coefficients, as shown in (2), are used to network coding operation.

$$M = \begin{pmatrix} 2 & 1 & 0 & 0 \\ 3 & 2 & 1 & 0 \\ 0 & 3 & 2 & 1 \\ 0 & 0 & 3 & 2 \end{pmatrix} \quad (2)$$

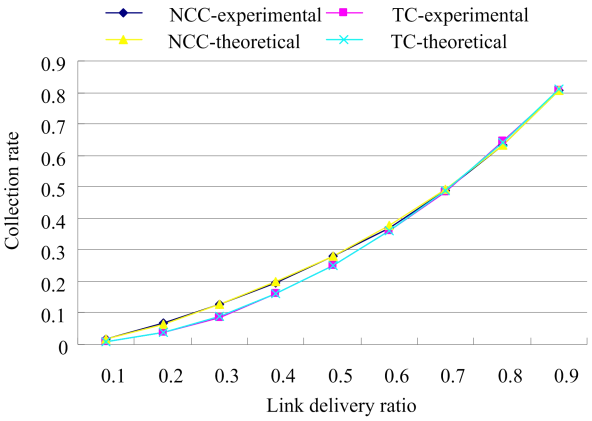
In the experiment, collection rate and energy consumption of Network Coding-based Convergecast (NCC) and Traditional Convergecast (TC) are discussed respectively.

A. *Collection rate*

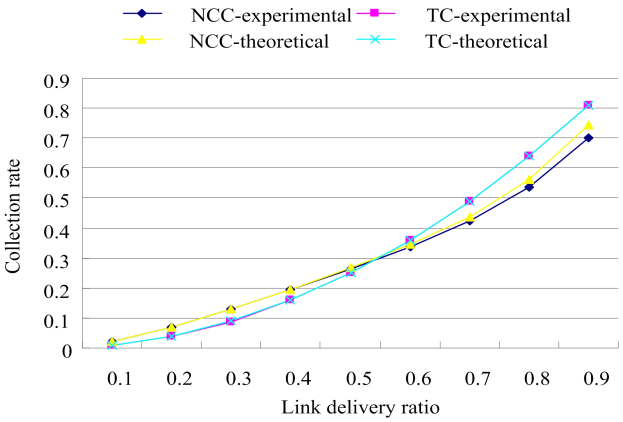
Collection rate R_c is the ratio of actually received packets N_a and the total received packets N_t in the root node in CT_n:

$$R_c = \frac{N_a}{N_t} \quad (3)$$

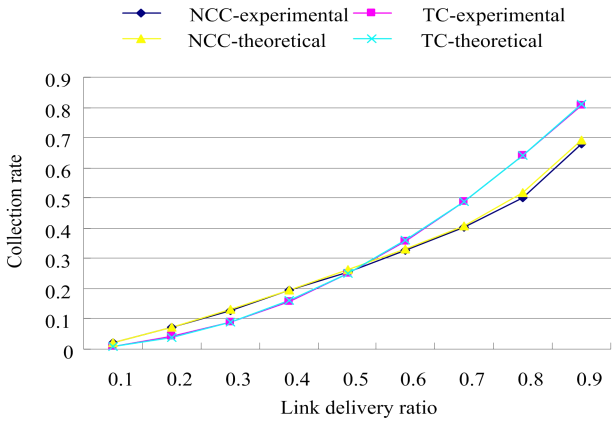
The convergecast network structures tested in the experiments are CT_2, CT_3 and CT_4. More than 12000 packets are sent out by original node under different link delivery ratios in different scales. The experimental and theoretical results are shown in Fig.3. From the figure, it's obviously to see that, the experimental results are corresponding with the theoretical results from [8].



(a) Collection rate of CT_2 structure



(b) Collection rate of CT_3 structure

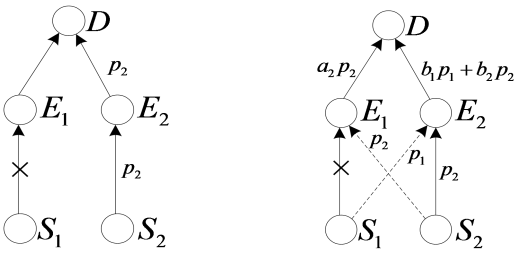


(c) Collection rate of CT_4 structure

Fig.3 Collection rate of NCC and TC

From Fig.3, it is easily to find that, when the link delivery ratio is low, the collection rate of NCC is higher than TC. That's because the overhearing wireless links are exploited by encoders, which bring the benefit of link reliability. For example, in CT_2, node S_1 and S_2 will send their packets p_1 and p_2 to node D, and they need interior node E_1 and E_2 to forward, E_1 and E_2 are in the communicating range of S_1 and S_2, as shown in Fig.4(a). p_1 is assumed to lose for transmission on the link from S_1 to E_1, and other links are good. In the traditional convergecast processing, only original packet p_2 is received by D, if we bring NC into the traditional convergecast network, because the packets from S_1 and S_2 can be overheard by E_1 and E_2, E_1 and E_2 can generate combined packets ($a_2 * p_2$) and ($b_1 * p_1 + b_2 * p_2$), the vector [0, a_2] and [b_1, b_2] are linear independent, then D can recover p_1 and p_2 according to linear network coding theory, as shown in Fig.4(b). Thus, NCC can bring reliability benefit.

The reliability is improved by NCC with the achieved network coding gain. But when the link equality is good enough, the collection rate of TC is higher than NCC, especially the large size of convergecast networks. That's because when the link quality is good, the reliability of the direct link is also high, then the link reliability benefit NC and redundant link bring is not very obvious. Besides, the cliff effect exists, that is, once decoder can't receive enough combined packets, then all the original packets which participating in encoding won't be recovered out. The factors lead to that the collection rate of NCC is not better than that of TC at the time when link quality is good enough.



(a) Traditional convergecast (b) Convergecast with network coding

Fig.4 Example of network coding used in convergecast

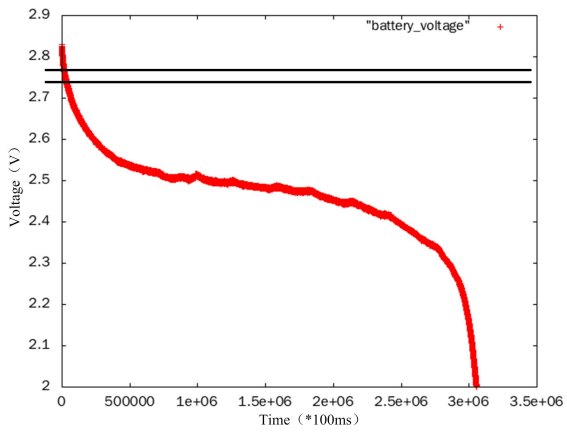
Table4-1 Current consumption of CC2420 module under different working modes

Mode	current consumption
Voltage regulator off mode(OFF)	0.02μA
Power Down mode(PD)	20μA
Idle mode (IDLE)	426μA
Receiving mode	18.8mA
Sending mode	17.4mA

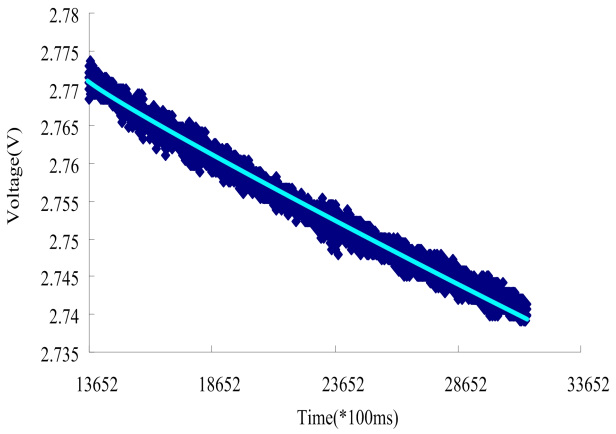
B. Average energy consumption

No matter whether there is NC or not, the energy consumption on the original nodes is the same between NCC and TC. As to decoder, NCC has the Gaussian Elimination operation, while TC does not. Because the Gaussian elimination is operated in CPU, and the energy consumption is too small that it is ignored in the experiment. So the energy consumption of encoders is the chief concern. There are three sources for encoders to consume energy: the packet's sending, the packet's receiving (include overhearing and idle listening) and the nodes's dormancy (including OFF, PD and IDLE). The current consumption of CC2420 module under different working modes can be achieved in [12], as shown in Table.1.

The change of battery voltage will be used to represent the energy consumption of the node in experiment. But, in global view, the discharge curve from 2.82V to 2V of AA battery which is used to supply power to the node is not linear, as shown in Fig.5(a) (In the discharge processing, the node's wireless module is always open, and the packets with voltage information are generated and sent out every 100 ms. So, the energy consumption is proportional to time, that is, the figure can be treated as the relation curve of energy consumption and voltage). In order to make the voltage's change denote the energy consumption clearly, it's necessary to set a linear change interval of voltage. For this reason, a interval from 2.77V to 2.74V is adopted in the experiment, the enlarged is shown in Fig5.(b), and from the figure, we can see that the discharge curve is linear in this interval.



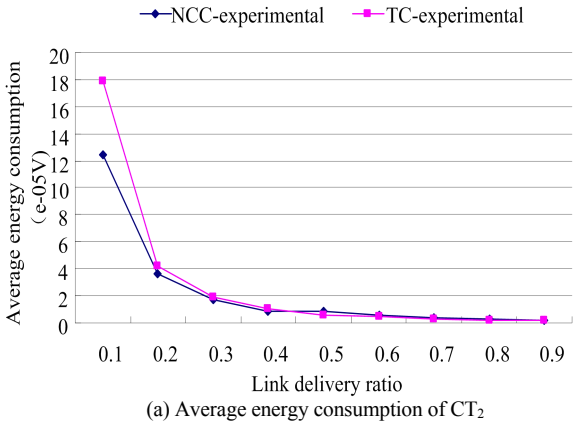
(a) Discharge curve between 2.82V and 2V



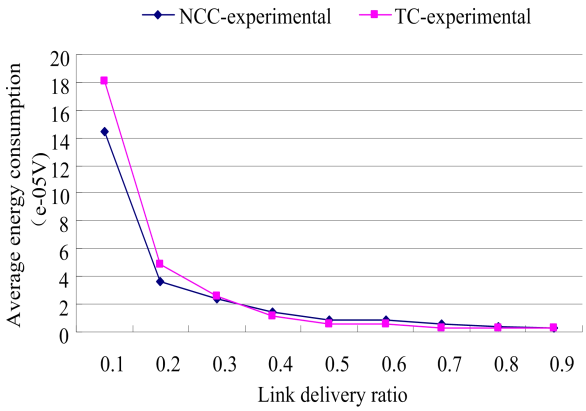
(b) Discharge curve between 2.77V and 2.74V

Fig.5 Discharge curve of the battery

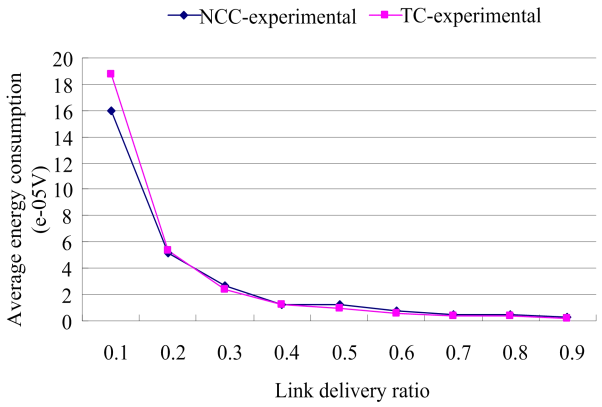
In CT_2, CT_3 and CT_4, the average energy consumption of NCC and TC under different link delivery ratios is shown respectively in Fig.6. For NCC, average energy consumption means the energy consumed by the encoder when the decoder recovers one original packet (represent by the change of voltage). And for TC, the average energy means the energy consumption of the interior nodes when the root receives one original packet successfully.



(a) Average energy consumption of CT_2



(b) Average energy consumption of CT_3



(c) Average energy consumption of CT_4

Fig.6 Average energy consumption comparison of NCC and TC

From the Fig.6, it is clear to see that the average energy consumption of NCC is not always higher than TC under arbitrary link delivery ratio. When the link equality is low, because the abundant link of NC can improve the transmission reliability and make NCC receive more packets than TC. The reliability benefit of NCC can make energy efficient and reduce the average energy consumption greatly. But when the link equality is high enough, the packets NCC receives are fewer than TC, and the energy efficient benefit reduces greatly. The encoders in NCC have to overhear original packets, and open the radio many times to receive packets, then the performance of encoders causes greater energy consumption than TC, so the consumption averaging to every recovered packets enlarge the average energy consumption.

IV. CONCLUSION

Realistic wireless sensor network is the basic platform in this paper, and NC is realized in traditional convergecast sub-tree model. It is important to notice that the converging sub-tree is ubiquitous (The broken line represents one convergecast sub-tree in Fig.1). NC is realized, and experimental analysis is under real network. It is hard to see this work reported in media, so the work in the paper has a certain sense. The realization of NC-based convergecast just adds a network coding layer to traditional convergecast sub-tree model, and the original topology need not to be changed, so the NC can be applied to practical environment. From the results we can see that, when the link equality is low, NCC can improve the collection rate of packets greatly, and has lower average energy consumption than TC. But when the equality is good enough, the performance of NCC is poorer.

The results show that, NC does not benefit traditional convergecast under arbitrary link delivery ratio. However, it is also obvious that there are critical points on collection rate and average energy consumption with different link delivery ratios. So this can be the future work. For example, if it is just needed to open the NC function when the collection rate is below the critical point and turn down the NC function when the collection rate is above the critical point. Just in this way, the performance of the whole network will be improved greatly. Then how to test out the critical point is a valuable issue.

ACKNOWLEDGMENT

This work is supported by the National Natural Science Foundation of China under the grants 61303210 and 61303211, and Zhejiang Provincial Natural Science Foundation of China under grant LQ12F02009.

REFERENCES

[1] Ahlswede R, Cai N, Li S R, et al. Network information flow [J]. IEEE Transactions on Information Theory, 2000, 46(4): 1204-1216.

[2] Li S R, Yeung R W, Cai N. Linear Network Coding [J]. IEEE Transactions on Information Theory, 2003, 49(2): 371-381.

[3] Ho T, Medard M, Shi J, et al. On Randomized Network Coding [C]. 41st Annual Allerton Conference on Communication, Control, and Computing, 2003: 11-20.

[4] Katti S, Rahul H, Hu W, et al. XORs in The Air: Practical Wireless Network Coding [C]. Proceedings of the ACM SIGCOMM, Pisa, italy, September 2006: 243-254.

[5] Hou I-H, Tsai Y-E, Abdelzaher T F, et al. AdapCode: Adaptive Network Coding for Code Updates in Wireless Sensor Networks [C]. IEEE 27th Conference on Computer Communications, 2008: 1517-1527.

[6] Nguyen D, Tran T, Nguyen T, et al. Wireless Broadcast Using Network Coding [J]. IEEE Transactions on Vehicular Technology, 2009, 58(2): 914-925.

[7] Tang Z, Wang H, Hu Q, et al. How Network Coding Benefits Converge-Cast in Wireless Sensor Networks [J]. KSII Transactions on Internet and Information Systems, 2013, 7(5): 1180-1197.

[8] Tang Z, Wang H, Hu Q. Network coding in convergecast of wireless sensor networks: friend or foe? [C]. 24th International Symposium on Personal, Indoor and Mobile Radio Communications: Mobile and Wireless Networks, 2013: 2469-2473.

[9] Samarasinghe K, Voigt T, Mottola L, et al. Network Coding with Limited Overhearing [C]. Proceedings of the 8th European Conference on Wireless Sensor Networks (EWSN2011), 2011.

[10] Voigt T, Roedig U, Landsiedel O, et al. Practical Network Coding in Sensor Networks: Quo Vadis? [C]. Proceedings of the 3rd International Workshop on Networks of Cooperating Objects, 2012.

[11] Keller L, Atsan E, Argyraki K, et al. SenseCode: Network Coding for Reliable Sensor Networks [J]. ACM Trans. Sen. Netw, 2013, 9(2): 1-20.

[12] Chipcon, SmartRF CC2420, 2.4GHz IEEE 802.15.4 / ZigBee-ready RF Transceiver [EB / OL]. http://www.alldatasheet.com/datasheet-pdf/pdf/125399/ETC1/CC2420.html, 2014-3-20.

Text Similarity Calculation Method based on Ontology Model

Tao Chi, Hanshi Wang, Lizhen Liu, Wei Song, Chao Du
Information and Engineering College, Capital Normal University
Beijing 100048, P. R. China

***Abstract*—Chinese text similarity calculation plays an important role in the Chinese information processing field. This paper uses ontology to calculate text similarity. The ontology has been widely used in natural language processing. Its unique concept model on the relationship offers a new processing method for word similarity calculation. Based on the characteristics of ontology and combined with HowNet and TongYiCi CiLin, we design a Hybrid Word Similarity Calculation-HWSC method and present a text similarity calculation method based on ontology model. The experimental results show that the method can make full use of the characteristics of ontology to accurately calculate the similarity between two texts.**

Keywords—Text similarity; Ontology; HowNet; TongYiCi CiLin

I. INTRODUCTION

The similarity calculation between two texts has always been a hot topic in natural language processing. It plays a vital role in many fields of natural language processing. For example automatic question answering, machine translation, information retrieval, automatic abstract, etc. Normally, if two texts have similar meaning, and their structure is similar, we think these two are similar. In addition, this paper uses ontology to calculate similarity between two texts. Ontology has been widely used in the field of Artificial Intelligence. In the early 1990s, Tom Gruber introduced the ontology as a technical term to mean a specification of a conceptualization: An ontology is a description (like a formal specification of a program) of the concepts and relationships that can formally exist for an agent or a community of agents[1].

This paper proposes a text similarity calculation method based on ontology model. Through the ontology model, this method converts two texts to ontology, combined with the Hybrid Word Similarity Calculation method (HWSC), and calculate the similarity between ontology. This method makes use of the structure of ontology effectively to accurately calculate the similarity between two texts. And the Hybrid Word Similarity Calculation method is a hybrid method by overlapping with three kind of word similarity calculation methods, they are word similarity based on Chinese character literal, word similarity based on HowNet and word similarity based on TongYiCi CiLin. The word similarity is embodies in their surface features and semantic expressions. These three methods show different sides of word similarity. In order to get accurate results, the HWSC method sums them up according to weighting factors. Ultimately, this paper designs an experiment to show the results of the HWSC method. We use student's answer and standard answer for short-answer question as experimental data, and calculate the similarity value between them. Then the method automatically gives points for student's answer, uses the points to compare with teachers' manual checking result. The experimental results show that the method can accurately calculate the similarity between two texts.

The remainder of the paper is organized as follows. In Section 2, we introduce some related works. The method of text similarity calculation based on ontology model is proposed in Section 3. Experiments and evaluations are reported in Section 4. We conclude the paper in Section 5.

II. RELATED WORK

Calculate the similarity between two texts, is to find the semantic and syntactic structure similarity between sentences. At present, there are a lot of similarity calculation methods, mainly classified into the following three: word overlapping method[2], the method based on corpus[3] and the method based on linguistics[4]. Word overlapping method computes sentence similarity through the same vocabulary that shared in two sentences. The method based on corpus assembles the words in sentences and regards them as a feature set, and calculates the vector Angle cosine values based on corpus as similar values. The method based on linguistic takes advantage of the semantic relations between words and its grammatical components to confirm the sentence similarity.

The above three methods calculate similarity between two texts from three different angles. This paper combines with the characteristics of these three methods, mixs them, and presents a new method of text similarity calculation.

The work of this paper mainly has two aspects: Firstly, a Hybrid Word Similarity Calculation method is put forward, which combined with words overlapping method and method based on corpus. Secondly, an ontology model is put forward by the use of the method based on linguistic.

III. TEXT SIMILARITY CALCULATION

A. Hybrid Word Similarity Calculation method

Word similarity is a subjectivity concept. According to actual need, the concept should be adjusted and extended in

practice. In this article, word similarity is defined as the coincidence degree between words on their domain knowledge space. There is a quantitative definition of word similarity.

- **Definition:** The similarity between word one and word two has a quantitative calculation result, value space is [0, 1].The more distance between two words meaning, the smaller the value.

According to the definition, we can get the specific text similarity value. All of the method in this paper will calculate text similarity with a certain value between 0 and 1.

Hybrid Word Similarity Calculation method is a mixture of three kind of word similarity calculation methods. They are word similarity based on Chinese character literal[5], word similarity based on HowNet[6] and word similarity based on TongYiCi CiLin[7].

1) Word similarity based on Chinese character literal

The similarity between words is affected by the number of the same morpheme and the weight of the morpheme in each word. In Chinese, to express the concept of a word, the most central part is in the second part of the word. A morpheme which located in the second part of the word plays a greater role than others. Because of this phenomenon, we need to increase the weight of morpheme in second part of the word. There is an algorithm.

$$SimZM(A,B) = \alpha \times \frac{F(A,B)}{2} + \beta \times d \times \frac{W(A,B)}{2} \quad (1)$$

$$F(A,B) = \frac{|SameHZ(A,B)|}{|A|} + \frac{|SameHZ(A,B)|}{|B|} \quad (2)$$

$$W(A,B) = \sum_{i=1}^{|A|} Weight(A,i) + \sum_{j=1}^{|B|} Weight(B,j) \quad (3)$$

$$d = \min\{\frac{|A|}{|B|}, \frac{|B|}{|A|}\}. \quad (4)$$

A and B represent for two words. $SimZM(A,B)$ represent for the similarity of A and B. $F(A,B)$ is the similarity between words in the number of same morpheme. $W(A,B)$ is the similarity between words in the weight of morpheme in each word. $SameHZ(A,B)$ represent for the collection of the same morpheme between A and B, $|A|$ represent for the number of characters in A, $Weight(A,i)$ represent for the i^{th} morpheme weight in A, d is coefficient of position, α and β represent for weight coefficient of the same morpheme similarity and same morpheme position similarity, $\alpha \times \beta = 1$.

The algorithm reflects the literal similarity between words.

2) Word similarity based on HowNet

HowNet[8] is one of the main knowledge bases in Chinese information processing, which is based on semantics. Words are mainly classified into two kinds: one is content words, the other is function words. Function words do not express any actual concept, but each concept of content words is composed of a set of sememe description formulas. They can be divided into four parts: first independent, other independent, relationships and symbols.

We calculate the similarity between first independent sememe according to their distance in the tree which is formed from relationship between up and down. For other independent sememe, we adopt the strategy about the whole divided into parts, matching two groups of sememe according to the similarity from big to small, and regard average value as the similarity of the whole. Formula is as follows:

$$SimZW(S_1,S_2) = \sum_{i=1}^{4} \beta_i \prod_{j=1}^{i} Sim_j(S_1,S_2) \quad (5)$$

$Sim_j(1 \le j \le 4)$ is the similarity between first independent, other independent, relationships and symbols. S_1 and S_2 represent for two words. $\beta_i\,(1 \le i \le 4)$ is an adjustable parameter, which is on behalf of the weight in each sememe description formula, and $\beta_1 + \beta_2 + \beta_3 + \beta_4 = 1$, $\beta_1 \ge \beta_2 \ge \beta_3 \ge \beta_4$ which means from Sim_1 to Sim_4 the effect is decreasing.

Word similarity based on HowNet reflects the semantic similarity between words.

3) Word similarity based on TongYiCi CiLin

TongYiCi CiLin is compiled by Mei JiaJu et al. in 1983. It is a synonym dictionary which includes synonyms for a word and a certain number of similar words. The words are classified by the hierarchical structure of the tree, and enciphered by serial number. In this dictionary, every word has its own, one and only, serial number. We use the number to make sure the semantic distance between words and calculate words' similarity. Furthermore, the serial number has five layer structures. To calculate word similarity, we need to know the layer where the words are different in, and compute by the fallowing formula.

$$SimCL(A,B) = a_i \times \cos(n \times \frac{\pi}{180}) \times (\frac{n-k+1}{n}). \quad (6)$$

a_i is branch coefficient between two words in their corresponding layer, each layer has their own coefficient. $\cos(n \times \frac{\pi}{180})$ is a regulation parameter. Its function is to control the words similarity between [0, 1]. n is the total number of the nodes in a branch layer. $\frac{n-k+1}{n}$ is a controls parameter. k is the distance between two branches.

Word similarity based on TongYiCi CiLin is to solve the words similarity through word meaning.

4) Hybrid Word Similarity Calculation method

The above three methods reflect word similarity in different level. Combining with these methods, this paper proposes a Hybrid Word Similarity Calculation method. First we calculate the value of words similarity through three methods, and then add them together according to proportion as a final result.

$$Sim(A,B) = \alpha \times SimZM(A,B) + \beta \times SimZW(A,B) + \gamma \times SimCL(A,B) \quad (7)$$

α, β, γ are the threshold values of these three word similarity methods. And $\alpha + \beta + \gamma = 1$. $SimZM(A,B)$ is the similarity calculated by word similarity based on Chinese character literal method. $SimZW(A,B)$ is the similarity calculated by word similarity based on HowNet. $SimCL(A,B)$ is the similarity calculated by word similarity based on TongYiCi CiLin.

B. Structure of Ontology Model

To compute text similarity, two problems need to be solved. One is similarity of words in sentences. The other is similarity of semantic structure between sentences. This paper uses Hybrid Word Similarity Calculation method to solve the first problem. And then, an ontology model is put forward to solve the second one.

In Chinese, a word in the sentence is endowed with different parts of speech. Such as adjectives, verbs, nouns, etc[9]. Considered from semantic structure, there is a mutual connection between words. For example, adjectives are usually located in the front of verbs. The connection between words can be reflected by their position in the sentence. This paper proposes an ontology model[10]. It regards the part of speech of words as an abstract ontology concept, and regards the connection between the words as the connection between ontology concepts. Graphic ontology model are as follows.

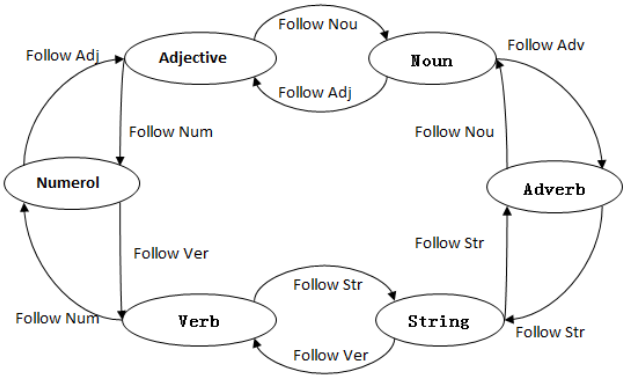


Fig. 1. Schematic Diagram of Ontology Model

In this ontology model, the part of speech is a concept. The location between words is the relationship between ontology concepts. We care about six parts of speech, there are adjectives, verbs, nouns, adverbs, strings and verbs. And each part has connection with others. Such as adjectives follow with nouns, and located in the front of verbs. However, the relationship between words is not confirmed. Figure 1 just shows part of the possible relationships. The exact relationship will be obtained by training.

In the HWSC method, this model will convert two texts to two ontology, and then this method calculates their similarity. To calculate text similarity by ontology model has several advantages.

- Ontology is more structured than text.
- For a text, the number of words and the position relationship is limited. That means the size of ontology is limited. And it can reduce operation cost.
- Ontology is easy to find out a word's information of the text. It can improve computational efficiency.

This paper proposes a similarity calculation algorithm to calculate the similarity between ontology. The main idea is as fallows.

a) Read two texts, and word processing by ICTCLAS [11].

b) According to ontology model convert two texts to ontology.

c) Based on the Hybrid Word Similarity Calculation method, calculate the similarity between individuals in corresponding ontology concepts.

d) Calculate the similarity between ontology concepts.

e) Calculate the similarity between ontology, and output as the result of text similarity.

To calculate text similarity, first of all, we need to text preprocessing. ICTCLAS is a word segmentation system. Its precision of part-of-speech tagging is up to 94.63%. After word segmentation, we need to convert two texts to ontology. According to ontology model, the word with special part of speech, such as adjectives adverbs, will be selected. We use them to build ontology. And then, calculate the similarity between individuals in corresponding ontology concepts by the mothed of HWSC. Then summarize result, calculate the similarity between ontology concepts, thereby calculate the similarity between ontology. As a result, we get a value between 0 and 1 as the similarity between two texts.

For example: there are two texts.

One. Five parameters of hidden markov model are the state number, the number of observations corresponding to each state, the initial state probability distribution, the state transition probability matrix and the observation probability distribution matrix.

Two. Five parameters of hidden markov model are the initial state probability distribution, the number of observations corresponding to each state, the state number and the state transform probability matrix.

We can see that the two texts are the same for the most part, only different in the order of the statement and the second paragraph. After word segmentation and build ontology by ontology model, we can get two ontologies.

TABLE I. INDIVIDUALS OF ONTOLOGY

	Adj	*Verb*	*Noun*	*Str*	*Num*	*Adv*
Ontology one	Null	Transition, Correspond	Matrix, Probability matrix, Probability distribution, Initial state, Observations, Number, State	Null	Null	Null
Ontology two	Null	Transform, Correspond	Probability matrix, Initial state, Number, Observations, Probability distribution, State, Matrix	Null	Null	Null

TABLE 1 shows the individuals of ontology. To calculate the similarity between ontology, we need to get the information of each word, and compare them. Take word "Transition" and word "Transform" for example.

"Transition" in ontology one we can get information like these:

{Name= Transition, Part-of= Verb, Follow-Noun= Probability matrix }

"Transform" in ontology two we can get information like these:

{Name= Transform, Part-of= Verb, Follow-Noun= Probability matrix }

To compare with this information, and calculate by HWSC method, we can get the similarity between individuals. And then summarize result, get the eventual outcome.

When follow the steps of the algorithm to calculate, we get a value of "0.8433" as the text similarity. It is basic conform to the actual.

IV. EXPERIMENTS AND DISCUSSIONS

When we evaluation text similarity, the subjectivity is strong. It's hard to get a result. In order to obtain a relatively objective evaluation result, this article uses student's answer and standard answer for short-answer question as experimental data. These data are come from exams of school. Through similarity calculation method based on ontology model, calculate text similarity between student's answer and standard answer. And then automatically give points for student's answer. Use the points to compare with teachers' manual checking result. Set *A* as student's answer, *B* as standard answer, *G(A,B)* as result of text similarity calculation method based on ontology model, *F(A,B)* as teacher's manual checking result, *Z(A,B)* as the total point. And then set a threshold value *w*, when

$$\left|\frac{G(A,B)-F(A,B)}{Z(A,B)}\right| < w \qquad (8)$$

We think that the calculation result is correct.

In this experiment, we randomly drew two groups of sample experiments for testing. Each experiment contains 1000 experimental data. We set different threshold to conduct several experiments, the results are as follows.

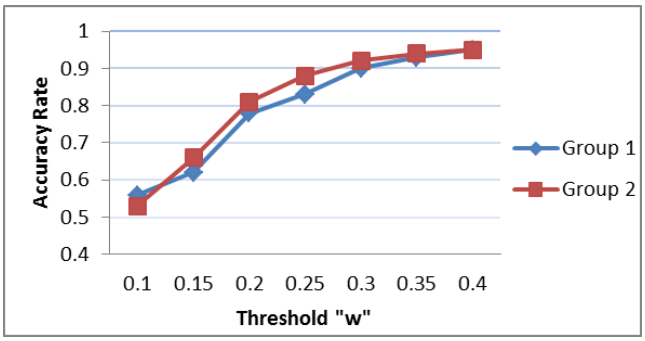


Fig. 2. Experimental Results

From Fig. 2, we see that the similarity calculation method based on ontology model can accurately calculate the similarity between two texts. When *w* set to 0.3, the accuracy can reach above 90%.

V. CONCLUSIONS

Text similarity calculation method based on ontology model is designed for calculating the similarity between texts. It uses ontology model to convert text to ontology, and then calculates the similarity between ontology based on Hybrid Word Similarity Calculation method. This method avoids chaos of data and increases the accuracy of text similarity calculation. In the meantime, it puts forward a new solution to text similarity calculation and helps researchers for further research. By observing the result of the experiment, we found that this method has advantage in comparing similarity between two texts.

ACKNOWLEDGMENT

This work was supported in part by National Science Foundation of China under Grants No. 61303105 and 61402304; the Humanity & Social Science general project of Ministry of Education under Grants No.14YJAZH046; the Beijing Educational Committee Science and Technology Development Planned under Grants No.KM201410028017; Academic Degree Graduate Courses group projects and the Beijing Key Disciplines of Computer Application Technology.

REFERENCES

[1] T. R. Gruber, "Toward principles for the design of ontologies used for knowledge sharing?," International journal of human-computer studies, vol. 43, pp. 907-928, 1995.

[2] D. Metzler, Y. Bernstein, W. B. Croft, A. Moffat, and J. Zobel, "Similarity measures for tracking information flow," in Proceedings of the 14th ACM international conference on Information and knowledge management, 2005, pp. 517-524.

[3] P. Shrestha, "Corpus-based methods for short text similarity," Rencontre des Étudiants Chercheurs en Informatique pour le Traitement automatique des Langues, vol. 2, 2011.

[4] P. Resnik, "Semantic similarity in a taxonomy: An information-based measure and its application to problems of ambiguity in natural language," arXiv preprint arXiv:1105.5444, 2011.

[5] S. Banerjee and T. Pedersen, "Extended gloss overlaps as a measure of semantic relatedness," in IJCAI, 2003, pp. 805-810.

[6] B. Ge, F.-F. Li, S.-L. Guo, and D.-Q. Tang, "Word's semantic similarity computation method based on Hownet," Jisuanji Yingyong Yanjiu, vol. 27, pp. 3329-3333, 2010.

[7] T. Jiu-le and Z. Wei, "Words Similarity Algorithm Based on Tongyici Cilin in Semantic Web Adaptive Learning System [J]," Journal of Jilin University (Information Science Edition), vol. 6, p. 010, 2010.

[8] Q. Liu and S.-J. Li, "The Word Similarity Calculation on<< HowNet>>," in Proceedings of 3rd Conference on Chinese lexicography, 2002.

[9] C.-T. J. Huang, Logical relations in Chinese and the theory of grammar: Taylor & Francis, 1998.

[10] X. Tao, Y. Li, and N. Zhong, "A personalized ontology model for web information gathering," Knowledge and Data Engineering, IEEE Transactions on, vol. 23, pp. 496-511, 2011.

[11] H. P. Zhang and Q. Liu, "ICTCLAS," Institute of Computing Technology, Chinese Academy of Sciences: http://www. ict. ac. cn/freeware/003_ictclas. asp, 2002.

Author Index

A

B

C

D

F

G

H

Y

Z