Task Scheduling and Seedblock Based Fault Tolerance in Cloud

B. Arunkumar
M.E., (Ph.D.), Assistant Professor, Faculty of Engineering,
Dept. of Computer Science and Engineering, Karpagam University,
Eachanari, Coimbatore, Tamil Nadu, India

M. Kesavamoorthi
PG Research Scholar, Faculty of Engineering,
Dept. of Computer Science and Engineering, Karpagam University,
Eachanari, Coimbatore, Tamil Nadu, India.

Abstract
Cloud computing is one of the essential paradigms of concern to the data sharing resources and utilization of data computation in data centers. Resource services in cloud computing is widely used these days. Cloud computing has a main drawback of fault tolerance. Maintaining of the fault tolerance is a necessity in providing availability and reliability of critical services in cloud services. The existing fault tolerant techniques use the Primary Backup (PB) model to backup the tasks of different cloud to allocate resource utilization. It considers only fault tolerant and task scheduling not to improve the resources. This paper proposes a combination of fault tolerance algorithms EIPR and SBA. It ensures a utilization of resources with high performance and efficient scheduling. The proposed EIPR algorithm implements the technique of task replication across multiple clouds. Also the backup recovery process is implemented through SBA algorithm. The purpose to implement the fault tolerance in cloud is to predict the failures, dynamic task scheduling, lower cost of resources and efficient functioning of data centers.

Keywords: tolerance, Primary-Backup(PB), EIPR, SBA, task scheduling.

Introduction
Cloud computing is one of the large scale resource providing with the data are not stored in a local system, stored in different data center through the internet. We use sharing the resources and service in the world wide, different host to store and utilize in different areas use the applications. Most of the applications are used in IT ability services. It also shares and demand the service cloud computing in IT fields [1]. Cloud computing is real time tasks and virtual machine allocation and distributed computing for the data center, multiple host environment to build the tasks and networking web applications and services. Most of IT Industries are need cloud Services with high cost and high demand services. Virtual machine isolation techniques are commonly used in cloud platforms to implement partitioning of resources in physical machines.

Most of the data center resources, deploy the virtual machines to allowing the users to accommodate the cloud data [2]. These virtual machines are using a few of the computational time by allowing the user to use the resources.

Cloud resource utilization increasing any fault occur in the mechanism in the systems to prevent system failure occurrence. When the fault occur, to overcome the fault using many type of mechanism in virtual machines on the data center’s resources [2]. The fault tolerance mechanism mainly used to fault recovery and detect and remove the fault, recovering automatically on the task in cloud computing. The advantage gained by implementing the fault tolerance for cloud computing are load balancing, scheduling the task in virtual machines, lower cost of scheduling and allocation of tasks with better performance criteria, etc. Usually several virtual machines run more than one instance of the application. If one server goes down, another virtual machine is automatically rescheduled the task and the resources are allocated automatically. Automatic fault tolerant techniques handle these kinds of faults very efficiently [3].

An essential and efficient way to achieve fault tolerance is by way of scheduling. It allocates multiple backup copies of tasks on various servers. The main target of the scheduling algorithm is to employ the resources properly and managing the load balance into the resources in particular performance time [4], [5]. Transient faults are handled by the fault-tolerant scheduling algorithm. To ensure optimum performance time of each task, all tasks are assigned to scheduling and the tasks are scheduled in uni-processor system [6]. The cloud efficiency depends on the task scheduling algorithm. Reliable operation, fault tolerance and robustness are based on the backups. The backups are provided based on the backup condition which provides the ratio of backup required. The primary backup approach is one way that the cloud system of
MeikandQiu has proposed a new model called FESTAL [9]. In the existing system, Xiaomin Zhu, XiaoQin, run tasks which are independent, real-time and QoS-aware (QAFT) is developed for heterogeneous clusters to fault tolerance a dynamic fault-tolerant and scheduling recovery. By integrating the QoS-based scheduling with better system require the dynamic fault tolerance options and error recovery. By integrating the QoS-based scheduling with better fault tolerance a dynamic fault-tolerant and scheduling algorithm (QAFT) is developed for heterogeneous clusters to run tasks which are independent, real-time and QoS-aware [8]. In the existing system, Xiaomin Zhu, XiaoQin, MeikandQiu has proposed a new model called FESTAL [9].

Virtualization and elasticity are considered for real-time tasks in clouds in highly efficient fault tolerant scheduling. This is achieved by FESTAL, i.e., Fault-tolerant Elastic scheduling algorithms. It is based on the elastic resource provisioning proposed in the fault-tolerant context. This optimizes the utilization of the resources and at the same time supports the fault tolerance in clouds. In order to increase the accuracy this model takes into account only the scheduling and fault identification. Thus FESTAL is very much tolerant to faults but cannot tolerate failure in multiple hosts. To overcome this issue, we have proposed in this paper, a combination of Enhanced IC-PCP with Replication (EIPR) algorithm, Seed Block Algorithm (SBA) for fault tolerance and efficient scheduling in cloud.

Related Works
In order to achieve improved availability, reliability and elasticity of the computing systems, the fault tolerant scheduling process uses a replication technology based scheduling – schedules multiple copies of tasks to make it more effective. To achieve effective load balancing and reduction in dynamic migration cost in virtual machines, a scheduling strategy is proposed for Virtual Machine resources which is based on the scheduling with variation and historical data. The task scheduling [11] has possible distributed load balancing algorithm which increases the reliability, flexibility and performance by way of Active Clustering. Live Virtual Machine migration is discussed by Katharina Haselhorst, Matthias Schmidt et al. [12]. This high performance live migration built based on the VM, is low in performance while working with multiple hosts.

In this paper our main goal is to provide the multiple hosts supporting the fault tolerance mechanism in the cloud by using dynamic, efficient scheduling of the resources by avoiding the fault in the cloud. Also the fault-tolerant mechanism, FESTAL cannot tolerate the failure of multiple hosts is considered as the main disadvantage. So we used to identify the fault and accuracy to tolerance with backup, recovery and task scheduling process in the cloud. This paper proposes a combined EIPR and SBA algorithm to improve the performance to a higher level, effective and efficient scheduling of cloud. This paper also aims to provide an accurate and effective scheduling to improve the task reproduction which addresses the accuracy in fault tolerance challenges. To maintain a high level of data efficiency, data recovery services are required in cloud computing. SBA is thus used in collecting information from remote locations and it will help in recovering the data in case it is deleted or cloud might be destroyed. The proposed Seed Block Algorithm without using any existing techniques also takes into consideration the security of the backup files which are stored in the remote servers. If for any reason, the cloud is destroyed or the files are deleted then, Seed Block Algorithm can be used to recover the lost files. Seed Block Algorithm executes the recovery process with less time than the current processes. This addresses the time related troubles. Next is the security of the backed up files in the remote server. Existing encryption concepts are not used by the Seed Block Algorithm. If the task crashes all the VMs running on the hosts, they become unavailable for the new tasks [13]. To avoid this situation, SBA uses backup copies maintained in various hosts for fault recovery. This ensures ease of use and reliability of the services. However once a selected host is loaded heavily, it is the sense of duty of the load balancing to move some VMs to lightly loaded hosts.

Problem Formulations
The main fault tolerance problems in cloud computing are fault detection and recovery of task process in the dynamic scheduling. The task scheduling and replication process in the cloud data centers to improve the high performance of migration in virtual machines and low cost working with the multiple hosts. The proposed method is the combination of EIPR and SBA algorithm for task scheduling and replication process in the cloud with efficient and effective performance. When a new task arrival in the host, they are first scheduling the task dynamically. Next to process the task model and scheduling dynamically every new task arrives in the host. Next, to monitoring the load balance of virtual machines in loaded for heavy or light. These complete processes in explaining details in following scenario.

Preprocessing Model
The scheduling model introduces the models, terminology used in this paper. This is a more efficient fault tolerance technique for task level in big and long running applications. For each changes done on the system a check point document is marked down when the task is completed. This will enable during failure of tasks, to restart from the point where the recently checked point occurs. This is more efficient than
starting from the beginning. This kind of fault tolerance technique is very efficient for long running processes in large applications. In short, recent checked point state is used to restart the job when the task fails or the changes done on the system.

Scheduler Model
The most important and the main focal activity in cloud computing is the Scheduling. To obtain maximum profit level and to get the most optimum efficiency in the work load we need to do scheduling. The scheduling algorithm implements the proper resource managing and load management among the resources. This enables to achieve the minimum performance time.

The workflow scheduling systems need to be designed such that they are fault tolerant if any failure happens in the computing environment. The workflow scheduling mainly concerned about the mapping and managing the workflow tasks. These tasks obviously are not under the control of the workflow systems. The proposed model assumes that workflow tasks are processed in order set up based on the control and data dependency. During runtime, if a task fails, it will be resubmitted to any resource that are available or to the same resource for further processing.

Task Model and Scheduling
Task scheduling is the core of the cloud computing system. The proposed algorithm tries to reduce the make span of the jobs with the use of minimal resources. The type of task to be scheduled is dependent on the scheduling algorithm. It provides a) Load balancing of the systems and b) higher executing efficiency. Any failed task will be resubmitted either to the same or a free machine. To improve the flexibility and reliability of the system in cloud, the efficient implementation of the Task scheduling is required. The final outcome of an efficient result to the user is achieved by finding out the complete and best sequence for various tasks to be executed within the set time bound range.

Backup controller
To get the desired result and successful execution, the tasks are replicated and are executed on multiple resources. The first step before starting a task is to create a backup copy. This is done by the backup copy controller. These two copies of the tasks are then delivered to the real-time controller. Real time controller will determine if these two copies of tasks can be completed before the set deadline time constraint range. If this is not possible, then real-time controller requests the resource controller for new additional resources. The task will be rejected if the timing constraint cannot be satisfied even by utilizing these additional new resources.

After completion f the primary task successfully, the result of the success is informed to the backup copy controller. It in turn informs the VM so that the scattered backup copie’s execution is canceled. But if the host failed in the fault tolerant mechanism, backup copy controller will be informed on this failure. Normally the execution will be cancelled. However VM will not be instructed to cancel the execution, but they will be executed based on the fault tolerant schedule already accepted.

Combination of EIPR & SBA
In this paper, they are proposed and used the combination of Seed Block Algorithm (SBA) and Enhanced IC-PCP with Replication (EIPR) algorithm. For fault tolerance and efficient scheduling in cloud, the smart remote data backup is the Seed Block Algorithm (SBA). The scheduling, task and backup recovery process of this system is explained in the system architecture diagram in figure 1.

**Figure 1: System Architecture**

**SBA Algorithm**
The proposed algorithm deals with the simplicity of the backup and recovery process. SBA computation is implemented with concept of XOR operation (exclusive OR). That is, if exists two A and B data files, The operation A+B creates X. If the file A is destroyed or deleted, the file is retrieved by X-OR of file X with file B. Eg: A = XB. Fig 1 describes the system architecture of the main, remote and client servers. Users are assigned with random numbers which are unique [16]. This client id is registered in the main cloud and then a seed block is generated for that client by X-ORing a random number. The seed block is being stored at the remote server corresponding to each of the clients. Initially the file is generated by the client in the main cloud. This will be X-ORed with the seed block and then it will be stored in the remote server. If in the main cloud, in case the file is damaged or deleted, then the original file will be X-ORed with the seed block and used.

**EIPR Algorithm**
The completion probability of the scientific workflow within the user-defined deadline is increased by the proposed IC-PCP algorithm. This IC-PCP algorithm is called Enhanced IC-PCP with Replication algorithm (EIPR algorithm). It provides a significant performance variation, high availability of services by way of task replication [17]. Fig 2 explains a high level
Algorithm steps:
Step 1: Combined provisioning of the cloud resources and task scheduling.
Step 2: Adjusting the Data transfer aware provisioning.
Step 3: Task replication.

Figure 2: (a) Original scheduling enabled by the IC-PCP algorithm. (b) Utilization of an available idle slot for replication of T4 (no extra cost incurred). (c) Allocation of VMs for two extra time units for replication of T2 and T7.

In the EIPR algorithm, the first step is to find out the quantity and type of VMs used for the purpose of execution of workflow and completion of each VM. Also the order and placement of tasks to be determined [17]. The availability of the VMs is affected by the scheduling. Scheduling in turn affects the time of virtual VMs. This implies the close relation between provisioning and scheduling. That is why, if both of these problems are solved, then a more efficient scheduling and provisioning is achieved.

Task Replication
EIPR algorithm aims to reduce such effects by way of utilization of the task replication in the idle slots of provisions or to enable replication of new VMs allocated, if only the replication budget allows. To increase the overall efficiency of the performance, EIPR mainly targets the space replication, which is considered as the main goal. Thus to reduce the faults and increase the fault tolerance, the tasks are replicated only in the different VMs which is entirely opposite to the single VM approach, where same task is scheduled for multiple times in the same VM which might increase the fault tolerance. The proposed process is available in the Algorithm 5 of the supplementary material available online. The task replication process can be called as the semi-active task replication technique for fault tolerance [14]. The main difference is that the tasks are replicated in hosts which are performance-independent rather than failure-independent locations. Also only a single entity manages the replication here.

Experimental Results
Experiments are conducted in Intel core i5 system with 2.6 GHZ speed, with Windows 7 OS platform and using CloudSim 3.0.3 simulator. The CloudSim toolkit models the cloud system components. It supports modeling of datacenters, virtual machines, host, scheduling and resource provisioning policies. Toolkits are useful to implement any hypothesis and evaluate the effectiveness of the concepts prior to the development of the software. Toolkits also enables to reproduce tests multiple times. Four virtual machines using the VM component is created and the properties set are, 512 MB RAM, MIPS as 250, 1000, 250 and 500 respectively. The Cloudlet component is used to create multiple tasks (10 tasks in total), with the property set as Cloudlet length of 10000, 20000, 20000, 100000, 100000, 200000, 100000, 200000, 1000 and 10000 respectively. Also 4 virtual machines are considered with MIPS 1000, 500, 250, 250 with RAM as 512 MB. The experiment is conducted for varying numbers of tasks like 200, 300, 400 and 500 respectively. SBA and EIPR algorithms are implemented for comparison and analysis. After EIPR using the simulation results the task performance and replication of task with load balancing is efficient speed and cost also reduce in that scheduling. The fault recovery of SBA giving good performance and the recovery of fault in multiple host and execute the replication process effectively. To minimize the total cost of the application execution the resources with different resource costs and computation performance are grouped in such a way to minimize the cost of execution for the cloud service providers.

Conclusion and Future Work
Occurrence of faults is one of the most important problem that need to be addressed in cloud computing. In order to improve the reliability and to achieve high performance in cloud computing, failures should be assessed and handled effectively. This paper discussed the combination of the SBA and EIPR algorithms, for implementing the cloud tolerance techniques. The SBA(Seed Block Algorithm) is an efficient way to recover the file with respect to time efficiency. It is very easy to implement and economical in cost. SBA maintains the data integrity to a high level. SBA also reduces the time required to recover the file. EIPR implements the concept of replication of tasks to increase the chance of resource allocation. Experiments using the EIPR concludes that the EIPR increases the resource utilization and reduces the overall total execution time of the processes. In the future, the new faults will be identified with compliance with cloud computing systems and also check load balance and multiple fault recover at the same time in the cloud and also gives effective safe scheduling and load balancing in the cloud.

References


