Preemptive Scheduling of Real Time Services With Resource Allocation Strategy for Cloud Computing

Syeda Sheema Samreen, Shahapur Farhat Kauser
Mtech (CNA)
Dept SCE, S.I.E.T Bijapur India

***Abstract* - This paper presents preemptive scheduling approach to focus on providing a solution for online scheduling problem of real-time tasks, we present a system that uses virtualization technology to allocate resources for the tasks that are reaching its deadline. The primary objective of real time preemptive scheduling is to increase throughput and minimize average response time instead of meeting deadlines. Earlier real-time tasks are scheduled non-preemptively with the objective to maximize the total utility, but the task with higher priority needs to wait until the currently running task gets completed. Hence if the waiting task reaches its deadline gets aborted. That before a task is aborted, it consumes system resources including network bandwidth, storage space and processing power. This leads to affect the overall system performance and response time of a task. Preemptive algorithm is proposed where the task with higher priority are executed first, and the tasks which are reaching its deadline are spoon to virtual machine, this this increases the total system performance.**

I. INTRODUCTION

Cloud computing is a prototype in which resources are permanently stored in servers on the internet and cached provisionally on clients. Large pool of systems is connected in private or public network. Cloud has the extended sized vision of computing as utility with pay-as-use basis. Amazon played a key role in cloud computing development by launching Amazon web service on utility basis in 2006. Before scheduling tasks on cloud computing, the characteristics of the cloud should be taken into account. Some of the characteristics of cloud include

1.1 On-demand self service
1.2 Ubiquitous network access
1.3 Location independent resource pooling
1.4 Rapid elasticity
1.5 Pay per use[1]

Infrastructure as a Service (IaaS) provides the processing, storage, networks, and other fundamental computing resources to users [1] [3]. IaaS users can deploy arbitrary appication, software, operating systems on the infrastructure, which is capable of scaling up and down dynamically. IaaS user sends programs and related data, while the vendor's computer does the computation processing and returns the result. The infrastructure is virtualized, flexible, scalable and manageable to meet user requirements.Cloud service provider gives many resources to the consumer through private cloud or public cloud. A number of users access the same resource, but they exchanged. Each task happen in real time system has timing properties and these properties are measured when scheduling the task on real time system. In real time applications or services, scheduling is the method to access the system to accomplish quality of service [4].

Cloud Computing refers to the use of computing, platform, software, as a service. It's a form of utility computing where the customer need not own the necessary infrastructure and pay for only what they use. Computing resources are delivered as virtual machines [1][3]. While there exist different interpretation and views on cloud computing [2], it is less disputable that being able to effectively exploit the computing resources in the clouds to provide computing service at different quality levels is essential to the success of cloud computing.

Millions of user share cloud resources by submitting their computing task to the cloud system. Scheduling these millions of task is a challenge to cloud computing environment. Resource allocation: The significance of resource allocation in cloud computing is a process to allocate the available resources to the needed cloud application over the internet. Resource allocation starves services if allocation is not managed precisely, There are different resource allocation strategy, virtualization is one[5].

II. PRELIMINARY

In this paper, the arrived real time tasks $\alpha t = \{T1, T2, T3,\ldots\ldots..,Tn\}$ are arranged in a priority order. A ready queue consists of n real time tasks which are scheduled non- preemptively. In this scheduling criterion, the currently executing task cannot be removed until it completes its execution. The highest priority tasks are waiting for a longer duration of time. The task will be aborted from its execution when it misses its deadline. Therefore the response time of the task will be greater and The survey of scheduling of real time services non preemptive considers the profit and penalty time of each task,this is the time utility

function[6,7]. the overall system performance can be degraded. With this problem, we developed an online pre-emptive scheduling with task which are reaching to its deadline are migrated to virtual machine.

III. PREMPTIVE ON-LINE SCHEDULING OF REALTIME SERVICES

In this section, we present a preemptive online scheduling of real time services with task migration algorithm to provide the solution for the task abortion when its misses its deadline and most preferably task with the higher priority is taken for its execution. Our pre-emptive scheduling algorithm works at scheduling points that include the arrival of a higher priority task and the critical point of the current task. Each task has its own priority. The priority of the task will be calculated by its deadline and cost. The critical point is the point in which a task misses its deadline. The detailed algorithm is described in algorithm

***Algorithm*: PRE-EMPTIVE ON-LINE TASK SCHEDULING ALGORITHM**

Step 1: Let {T1,T2…….Tn} be the accepted task in the ready queue
.

Step 2: Arrange the tasks in the ready queue based on the priority

Step 3: Task with the highest priority in the ready queue taken for its execution.

Step 4: If a new task arrives, put the task at the head of the ready queue and sort the task based on the priority.

Step 5: A new task arrives with the highest priority is ready to run and pre-empts the currently executing task comparatively.

Step 6: The pre-empted task re-enters into the ready queue and sorted according to the priority.

Step 7: Whenever a waiting task reaches its deadline, it will be migrated to another virtual machine.

The real time tasks are accepted in the ready queue. For each task, the priority is calculated based on deadline and cost. Sort all the tasks in the ready queue based on the priority and the task with highest priority is selected for execution. Upon the completion of the current task, choose the task from the ready queue which has the highest priority and starts its execution. If a new task arrives, it is first inserted at the head of the ready queue and calculates the priority. Based on the priority of the task, it is compared with all the tasks and inserted accordingly in the queue. This procedure continues until the entire ready queue becomes a list ordered according to their priority. The currently executing task is preempted whenever a new task arrives with the higher priority. The preempted task is put into the ready queue and sorts it according to the priority.
Whenever the currently executing task reaches its deadline, the task is instantly migrated to another virtual machine.

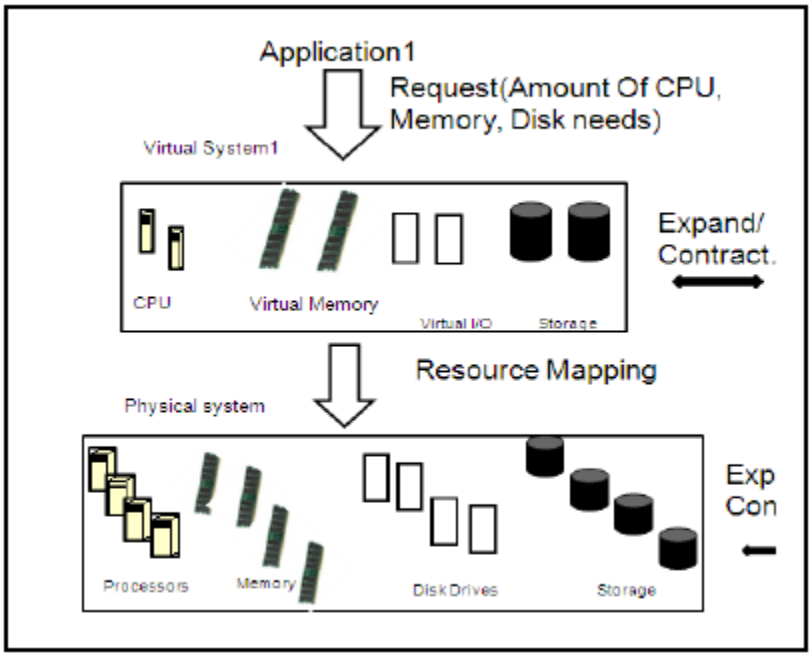

Figure1. Mapping of virtual to physical resources

IV. PREDICTING FUTURE RESOURCE NEEDS

We need to predict the future resource needs of Virtual Machine. As said earlier, our focus is on Internet applications. One solution is to look inside a Virtual Machine, for application level statistics, e.g., by parsing logs of pending requests. Doing so requires modification of the VM[8]

which may not always be possible. Migrating an task to virtual machine may require extra effort.

V. CONCLUSION

Considering the tremendously large scale of the computing resource for cloud computing, it is necessary that cloud scheduler are designed to assist the services to reflect on the large scale of computing resource for cloud computing. In existing approach, the tasks are scheduled non preemptively. Whenever a task enters into the ready queue, it is not rubbed out until it completes its execution. The highest priority task is waiting for vague amount of time. When the time reaches the deadline of the current task, the current active task is immediately discarded and the task with the highest expected gain is selected for execution. Therefore, it affects the overall system performance and response time. In order to address these challenges, this paper is proposed to minimize the response time and to improve the efficiency. The task with higher priority in the ready queue is accepted for its execution. The currently executing task is pre-empted as soon as the task with the higher priority enters into the ready queue. The pre-empted task re-enters into the ready queue and sorted according to their priority. A task will be migrated to new virtual machine, when it misses its deadline.

REFERENCES

[1] M. Malathi, "Cloud Computing Concepts", IEEE,2011
[2] Paul, M., Sanyal, G., "Survey and analysis of optimal scheduling strategies in cloud environment", IEEE, 2012
[3]F. Casati and M. Shan (2001)., Definition, execution, analysis and optimization of composite e-service, IEEE Data Engineering
[4] R.Santhosh and T.Ravichandran," Non-Preemptive on-line scheduling of real-time services with task migration for cloud computing"
[5]Dr R.Sridaran Dean."Different Resource Allocation Strategy in Cloud Computing"
[6] R. K. Clark. Scheduling dependent real-time activities. PhD thesis, Carnegie Mellon University, 1990.
[7]shaou Gang quan,Shangping Ren"Scheduling of online realtime services"2010
[8] Zhen Xiao, Senior Member, IEEE, Weijia Song, and Qi Chen "Dyanamic Resource allocation Strategy in Cloud Computing"