FAULT-TOLERANCE SUPPORT FOR DYNAMIC OPTIMAL PROPORTIONAL-SHARE IN A SELF-ORGANIZING CLOUD SYSTEM

N.P. Ponnuviji

Student, Department of Computer Science and Engineering, SRM University, Chennai, (India)

ABSTRACT

The increasing popularity of Cloud Computing as an attractive alternative to classic information processing systems has increased the importance of its correct and continuous operation even in the presence of faulty components. But fault tolerance is still a crucial issue in SOC. By utilizing VM technology, the performance and fault isolation of cloud resources can be provisioned on demand in a fine grained, multiplexed manner. The design of the DOPS resource allocation method, utilizes the proportional share model. The key idea is to redistribute available resources among running tasks dynamically, where these tasks could use up the maximum capacity of each resource in node. To locate for qualified nodes in the SOC environment, a fully decentralized range query protocol called P-G CAN has been tailored for DOPS. Pointer Gossiping is responsible for making sure every node in the system eventually knows important information about every other node's state including those that are unreachable. The sensitivity analysis conducted on the model, violates certain assumptions that creates more impact on the dynamic optimal resource allocation of multiple attributes due to faulty components. Hence an analysis will be made in a VM environment, aiming to find a node already present in the network and identify a zone that can be split. Thereby, updating the routing tables of nodes neighbouring the newly split zone. During this step, faulty nodes will be identified and re-routing will be done either by merging zones or managing with the neighbour node, all performed dynamically by the SOC system.

***Keywords**:Dynamic Optimal Proportional Share (DOPS), Pointer-Gossiping Content Addressable Network (PG-CAN),Self –Organizing Cloud (SOC),) Sun Grid Engine Cluster (SGE), Virtual Machine (VM).*

I INTRODUCTION

Cloud computing is sharing of resources on a large scale that leads to cost effective and location independent. Based on the utility and consumption of computing resources, the jobs will be allocated into the corresponding resources and transmitted through a cloud. A cloud involves deploying groups of remote servers and software networks that allow centralized data storage and online access to computer services or resources. It is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications and services). The increasing demand for flexibility in obtaining and releasing computing resources in a cost-effective manner has resulted in a wide adoption of the Cloud Computing paradigm. Fault tolerance computing is one that can continue to correctly perform its task in the presence of hardware failures, network failures. The faults on the resources may occur due to network fault, physical fault, media fault, processor fault, process fault, service expiry fault, etc., A new architecture called self-organizing cloud (SOC) can connect a large number of desktop computers on the Internet by a P2P network. Through multiple VM instances constructed by SOC idle resources will be allocated to tasks waiting for execution. Even though optimization of task's resources allocation under user's budget is dealt under SOC, the process of handling fault-tolerance problem is still pending. In this paper, we provide a fault-tolerance support for DOPS based PG-CAN system, that dynamically analyse the weights and cost of resources and identify the optimal resource and find whether it could be split up into zones, thereby,providing an optimal route by eliminating unqualified nodes in the network.

II EXISTING SYSTEM

In the existing system, the transmission in the cloud takes place after identifying the valid resources based on the resources requirements and budget and are named as qualified nodes. When suddenly due to a hardware failure or network failure when a resource or node becomes unqualified (i.e., not able to perform transmission), the cloud server makes the transmission pending. The problem is identified manually by rectifying Convex Optimization Problem. Hence in this scenario, to locate qualified nodes in the SOC environment, a fully decentralized range query protocol, named pointer-gossiping CAN (PG-CAN) has been tailored for DOPS.

2.1 Disadvantages

- Identification of fault-tolerance is not handled in both centralized and decentralized environment.
- The SOC does not handle the fault-tolerance support in the DOPS based VM environment.
- Manual analysis of rectifying Convex Optimization Problem only is dealt.
- The sensitivity analysis of how violation would impact the optimal resource allocation is not handled.

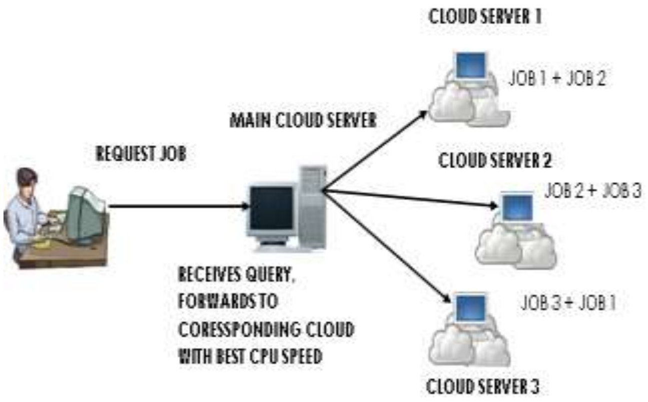


Figure -1Architecture Diagram of Existing System

- Fails to identify the Unqualified nodes in the SOC system.

2.2 Resource Allocation in Cloud

Resource allocation is an integral, evolving part of many data centre management problems such as virtual machine placement in data centres, network virtualization, and multi-path network routing. Since the problems are inherently NP-hard, most existing systems use custom-designed heuristics to find a suitable solution. However, such heuristics are often rigid, making it difficult to extend them as requirements change.

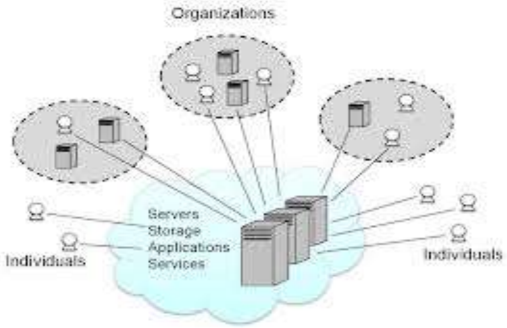
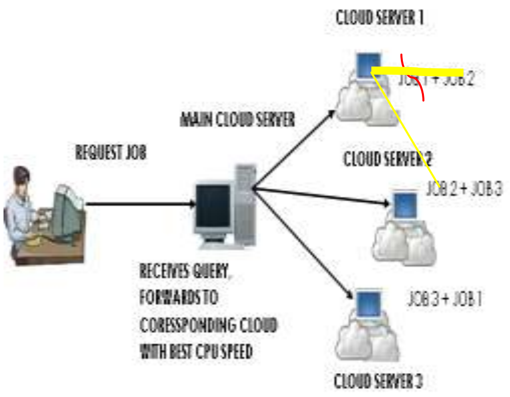


Figure-2 Resource Allocation in Cloud

Resource provisioning for cloud services in a comprehensive way is crucial to any resource allocation model. Any model should consider both computationalresources and network resources to accurately represent and serve practical needs. Another aspect that should be considered while provisioning resources is energy consumption. This aspect is getting more attention from industry and governments parties. Calls of support for the green clouds are gaining momentum.

III PROPOSED SYSTEM

In the proposed system, novel cloud architecture, namely self-organizing cloud (SOC), can connect a large number of desktop computers on the Internet by a P2P network. Each participating computer acts as both a resource provider and a resource consumer. They operate autonomously for locating nodes with more abundant resource or unique services in the network to offload some of their tasks; meanwhile they could construct multiple VM instances for executing tasks submitted from others whenever they have idle resources. We focus on two key issues in the design of SOC: 1) the multi attribute range query problem in a fully decentralized environment for locating a qualified node to satisfy a user task's resource demand with bounded delay and 2) how to optimize a task's execution time by determining the optimal shares of the multi attribute resources to allocate to the tasks with various QoS constraints, such as the expected execution time and limited budget.



When JOB1 becomes unqualified, the SOC dynamically, switches to another optimal JOB or Resource and continues the transmission

Figure-3 Architecture Diagram of Proposed System

The Proposed System comprises the following modules:

A. Creation of Jobs

Initially, the user creates the job that he wants to execute. The user can specify the priority by which the tasks have to be executed. The user can also specify the server name in which the task has to be scheduled.

B. Resource Allocation Module

Once the job is created, the manager updates the job based on the user specified priority that to be executed on the server. The optimal resource vector is also computed in this module for the processes. Based on this optimal resource vector, the resources are allocated for the jobs that are created by the user. The resources along with their weights will be assigned and allocated at the first stage.

C. Optimal Routing Module

Based on the allocated resources, the DOPS identifies the optimal resources and finds whether it could be split up into zones. It will analyse the weights assigned for each resources and its cost using the Interior –Point Method.

To overcome, the Convex-Optimization problem (i.e., maximize the usage of resources and minimize the cost involved), Interior –Point method will be applied. Now an optimal route will be identified (can be one or more routes).

D. Faulty Node Identification Module

Once the transmission is started, the resources present in the optimal routes, may suddenly become unqualified, due to the following reasons-

i) Overload in the resource.

ii) Network Fault.

iii) Processor Fault.
iv) Media Fault, etc.,

E. **Pointer Gossiping CAN**

A resource discovery protocol, namely pointer-gossiping CAN (PG-CAN) is used to find the qualified nodes. Like traditional CAN, each node under PG-CAN is responsible for a unique multi-dimensional range zone randomly selected when it joins the overlay. Every node will periodically propagate the state-update messages about its available resource vector to the duty node whose zone encloses this vector. The query message will be routed to the duty node.

F. **Re-routing Module**

This module we implement the Sun Grid Engine (SGE) Cluster that identifies the faulty node or unqualified node. This engine collects the details of the load balance status of main cloud server and also the remaining proxy cloud servers spread along the split-up zones. Once the faulty node is identified, the SGE removes it from the cloud server using the PG-CAN, then identifies the next new optimal path in the VM environment, and performs the transmission along the new path dynamically. It also allocates the process to the corresponding jobs located in the server dynamically, by doing the required updating in the routing table of the resources in the cloud.This module undergoes a set of sub-modules given below:-

F-i). Node-Joining Module

This module entirely changes the optimal route, by removing the faulty node from the zone and adding the new node to the zone for further transmission. It updates its own routing table and also the routing table of main server and proxy cloud server.

F-ii) Node Departing Module

This module removes the relevant information about the faulty node. It also removes the faulty node's information from main cloud server and proxy cloud server. The routing tables are updated at regular intervals based on the transmission. Information related to the load balance of existing nodes, network traffic and process allocation of jobs is dealt at regular intervals.

G. ***Task Scheduling***

In the task scheduling process, each task is allocated a resource using pointer relay algorithm. The current node sends the query message to another duty node; upon receiving such a message, the remote duty node will also perform the process of task scheduling. Each process on completing its execution will be notified by the user. This task scheduling process is based on cloud activator protocol. This protocol allocates the resource to process whenever it's being interrupted by other processes during its execution. When a process is executing on a server, when no interrupt occurs, the process completes by its optimal resource allocation strategies. In case, when interrupt occurs during the process execution, the protocol activator which uses the algorithm of resource query. This maintains the random selection policy.

Upon receiving the query result, the requesting node will randomly choose one out of them as the final resource node for executing the submitted task. With this random selection policy, we can effectively mitigate the decision conflict among different tasks.

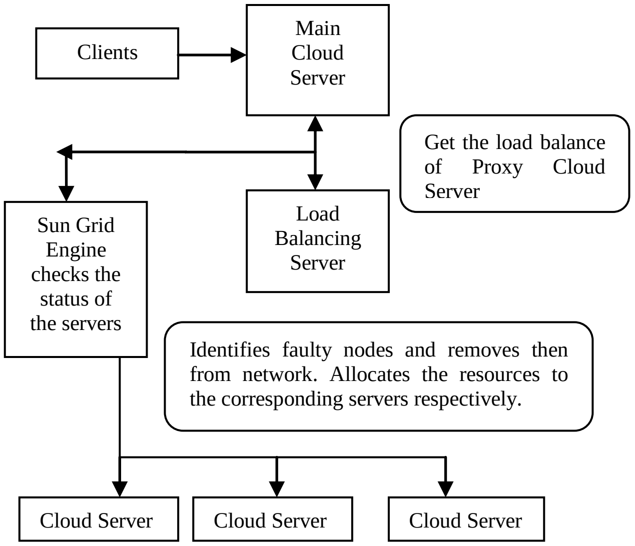

Clients
Main Cloud Server
Get the load balance of Proxy Cloud Server
Sun Grid Engine checks the status of the servers
Load Balancing Server
Identifies faulty nodes and removes then from network. Allocates the resources to the corresponding servers respectively.
Cloud Server
Cloud Server
Cloud Server

Figure-4 Flow Diagram of Proposed System

H. Advantages

- Identifiesthe faulty or unqualified nodes during transmission dynamically in both centralized and decentralized environment.
- The SGE (Sun Grid Engine) cluster handles the fault-tolerance support in the DOPS based VM environment.
- Brute-force strategies and Interior-point method is applied in rectifying Convex Optimization Problem.
- Contrary to existing solutions which often generate bulky messages per request, our protocol produces only one lightweight query message per task on the Content Addressable Network (CAN) using the Pointer Gossiping technique.

IV RELATED WORK

Shen Di et.al [7] explored integrating volunteer computing into cloud architectures using a self-organizing cloud to reap the huge potential of untapped commodity computing power over the internet. Dr. Lakshmi Prasad et.al [8] has analysed and proposed various ways to handle fault-tolerance. The occurrence of various faults and its types during various scenarios are dealt. Zhang et.al [4] proposed a Byzantine fault tolerance framework for building reliable system in voluntary resource cloud infrastructure. The paper analysis failures happened on primary level will decrease the overall performance of a BFT group. When the primary is faulty, a primary updating procedure will be triggered in the Request Execution phase. Patra et.al [5] approached fault taxonomy and need of fault tolerance in cloud computing. Various proposed models for fault tolerance are discussed and compared on the basis of Metrics for fault tolerance in cloud. But faces more challenges which need some concern for every framework or model.Chandrakala et.al [6] proposed a load balancing algorithm for virtual machine. The algorithm checks the CPU utilization depending on the request. The intermediate node is used to monitor the load of each node.

REFERENCES

[1] Heng Tao Shen, Yan Feng Shu and Bei Yu, “Efficient Semantic-based Content Search in P2P network”, IEEE Trans. Knowledge and Data Eng., vol.16, no.7, pp.813-826, July 2004.

[2] Michael Feldman, Kevin Lai, Li Zhang, “The Proportional-Share Allocation Market for Computational Resources”, IEEE Trans. Parallel and Distributed Systems, vol.20, no.8, pp 1075-1088, Aug.2009.

[3] J.Wang, S.Wu, H.Gao, J.Li and B.Ooi, “Indexing Multidimensional Data in a Cloud System”, Proc. ACM Int’l Conf. Management of Data (SIGMOD’10), pp.591-602, 2010.

[4] Yilei Zhang, ZibinZhengand and Michael R. Lyu, “BFT Cloud: A Byzantine Fault Tolerance Framework for Voluntary- Resource Cloud Computing”, 4th International Conference on Cloud Computing, IEEE, 2011.

[5] Prasenjit Kumar Patra, Harshpreet Singh and Gurpreet Singh, “Fault Tolerance Techniques and Comparative Implementation in Cloud Computing”, International Journal of Computer Applications, vol.64, no.14, February 2013.

[6] N. Chandrakala and Dr. P. Sivaprakasam, “Analysis of Fault Tolerance Approaches in Dynamic Cloud Computing”, International Journal of Advanced Research in Computer Science and Software Engineering, ISSN: 2277 128X, vol.3, Issue 2, February 2013.

[7] Shen Di and Cho-Li Wang, “Dynamic Optimization of Multi-Attribute Resource Allocation in Self- Organizing Clouds”, IEEE Transactions on Parallel and Distributed Systems, vol.24, No.3, March 2013.

[8] Dr. Lakshmi Prasad Saikia, YumnamLanglen Devi, “Fault-Tolerance Techniques and Algorithms in Cloud Computing”, International Journal of Computer Science and Communication Networks, vol 4(1), 01-08, ISSN: 2249-5789.

[9] dm-ioband: http://sourceforge.net/apps/trac/ioband, 2012.

[10] J.E. Smith and R. Nair, “Virtual Machines: Versatile Platforms for Systems and Processes”. Morgan Kaufmann, 2005.

[11] D. Krishna, “Dynamic Resource Allocation using Virtual Machines for Cloud Computing Environment”, International Journal of Ethics in Engineering and Management Education”, ISSN:2348-4748, vol.1, Issue 12, December 2014.

[12]Eugen Feller, Louis Rilling, Christine Morin, “Snooze: A Scalable and Autonomic Virtual Machine Management Framework for Private Clouds”, HAL Id: hal-00651542,https://hal.inria.fr/hal- 00651542, December 2011.