ISSN : 2347 - 8446 (Online)
ISSN : 2347 - 9817 (Print)

A State of Art Scheduling Algorithms in Cloud Environment

[I]V R Bithiah Blessie, [II]A.Stanislas, [III]Dr L.Arockiam
[I]M.Phil Scholar, [II]PhD Scholar, [III]Associate Professor
[I,II,III]Dept. of Computer Science, St. Joseph College Trichy, Tamilnadu, India

Abstract
The Cloud Computing is one of the fast improving technologies. It is a new technology which is achieved by distributed, parallel and grid computing. Cloud Computing offers end user as an "a pay as go model". Cloud computing provides different types of resources like software, platform and infrastructure as a service through internet. The end consumers request for available services depends upon their desired Quality of Service. One of the most challenging problems is scheduling. Some of the scheduling techniques are task scheduling, job scheduling, workflow scheduling. In that task scheduling is the significant task in cloud computing environment because based upon time the end user has to pay for the resources. This paper surveyed various types of Scheduling techniques which are used in the cloud computing.

Keywords
Cloud Computing, Quality of Service, Scheduling, Task Scheduling, Job Scheduling.

I. Introduction

Cloud Computing is internet connected mode of supercomputing. It is a large-scale distributed computing model, which depends on the economic size of the operator that is virtualized, abstract and dynamic[1].

A Cloud is a parallel and distributed computing system consisting of a collection of inter-connected and virtualized computers that are dynamically provisioned and presented as one or more unified computing resources based on service-level agreements (SLA) established through negotiation between the service provider and customers [2]. The National Institute of Standards and Technology (NIST) [3, 4] characterizes cloud computing as " cloud computing as a pay-per-use model for enabling available, convenient, on-demand network access to a shared pool of configurable computing resource that can be rapidly provisioned and released with minimal management effort or service provider interaction".

Fig.1 [4] presents an overview of the NIST cloud computing reference architecture, which identifies the most important components, their activities and functions in cloud computing. The diagram describes a basic, high-level architecture and is proposed to assist the understanding of the requirements, uses, characteristics and standards of cloud computing.

The Cloud model consists of five necessary characteristics, three service models, and four deployment models. The five necessary characteristics are on-demand self-service, broad network access, resource pooling, rapid elasticity, and measured service. The three service models are namely Software as a Service, Platform as a Service, and Infrastructure as a Service. The four deployment models are namely public cloud, private cloud, hybrid cloud, community cloud [3].

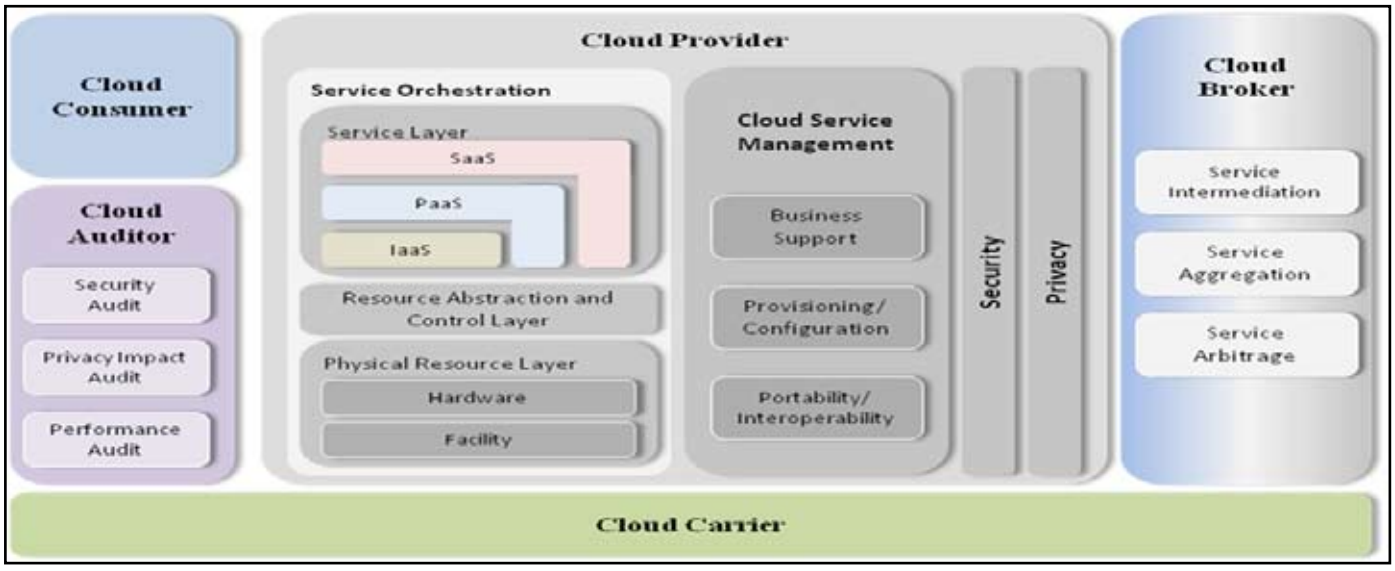

Fig.1: The Conceptual Reference Model.

Quality of Service (QOS) [1],[5] is a parameter reflecting the performance of the Internet. In the cloud computing, QOS is a standard of user's satisfaction to the services. For example, some require more memory to store data, and some others may need more CPU time to compute complex task, etc. Select the completion time, bandwidth, cost, distance, reliability as a QOS parameters when the task implemented in the virtual machine resources.

- Completion time: Tasks need to be completed within a minimum time for the real time requirements of users.
- Bandwidth: The bandwidth requirements need to be considered priority when users require a higher communication

© 2014, IJARCST All Rights Reserved

bandwidth, such as multimedia streaming needs.
- Cost: Cloud computing is paid according to the user demand. The cost is a factor which users concern.
- Distance: Users want tasks to be run in the near resources.
- Reliability: For the users long-running tasks, cloud need to provide stable and reliable performance, such as cloud storage service.

II. Scalability

Scalability [6] is the ability of a system, function or model that describes its capability to bring off and achieve under an enlarged or expanding workload. A scheme that scales well will be able to sustain or even increase its tier of public presentation or competence when tested by superior equipped demands. There are two types of scalability.
- Scaling up
- Scaling down

Scaling up: It is also called vertical scaling. It is the capacity to enhance the capacity of existing hardware or software by adding resources [7],[8].
Scaling down: It is also called horizontal scaling. It is the ability to connect multiple hardware or software entities in a single unit [9]. The key terms which are associated in scalability are listed below [10].
- Load balancing
- Workload
- Scheduling
- Resource allocation
- Quality of service
- Service Level Agreement

III. Scheduling

A Scheduling [11] is a set of rules that determine the jobs to be executed at a particular time. It is the process of deciding how to commit resources between varieties of possible tasks. The main aims of scheduling algorithms are to reduce resource malnourishment and make sure the fairness among the parties utilizing the resources.

A. Scheduling Process

Scheduling process in the cloud can be widespread into three stages namely [11],

1. Resource discovering and filtering:-

The datacenter broker discovers and collects status information of the resources present in the network system related to them.

2. Resource selection:-

Based on certain parameters of task and resource target resource is selected. This is determining stage.

3. Task Submission:-

Task is submitted to the selected resource.

B. Types of Scheduling

Scheduling can be mainly divided into three types [12],[13],[14]. They are,
1. Long term scheduling
2. Medium term scheduling
3. Short term scheduling

1. Long Term Scheduling:

The long term scheduler decides which task is to be admitted to the system for execution and when, and which one should be exit.

2. Medium Term Scheduling:

The medium term scheduler decides when the processes are to be suspended and resumed.

3. Short Tern Scheduling:

The short term scheduler is also called a dispatcher, which decides which of the ready process can have the resources and for how long.

C. Types of Scheduling Algorithms.

The scheduling types of algorithms [15] can be illustrated in the Fig 2.

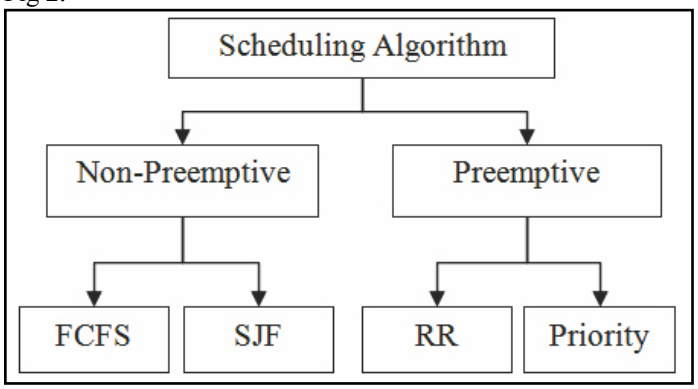

Fig. 2: Scheduling Algorithm.

1. Non-Preemptive Scheduling Algorithm:

In the Non preemptive algorithms, it is determined that once a process is entered into the running state, then it is not detached from the processor until it has finished its service time. It has two types of algorithm [14,15].

i. FCFS

FCFS is a First-Come-First-Served algorithm. It is also called as First-In –First-Out. It is the simplest scheduling algorithm, in which first arrived task is to be processed first. It is better for longer jobs [14,15].

ii. SJF

SJF is a Shortest Job First Algorithm. This chooses the task with shortest processing time first. So the long tasks are always executed after shorter tasks [14,15].

2. Preemptive Scheduling Algorithm:

In the preemptive algorithms, it is determined that if a task is presently using the processor and a new task with a highest priority enters into the processor. Then the task with the highest priority should forever be the one presently using the processor [14,15].

i. Round Robin (RR)

Round robin is one of the simplest scheduling algorithms, which assigns time slices to each task in equal portions. It handles all tasks without priority [14],[15].

ii. Priority Based.

In this priority based scheduling algorithm, each task is assigned a priority. The task at the beginning of the list with the highest

© All Rights Reserved, IJARCST 2014

ISSN : 2347 - 8446 (Online)
ISSN : 2347 - 9817 (Print)

priority is selected first [14,15].

IV. Existing Scheduling Algorithms.

1. Optimistic Job Scheduling Algorithm:

Shalmali, et al [16] presented an optimistic job scheduling system for cloud computing. In this paper they design an application for scheduling the multiple request. They use job scheduling algorithm and static load balancing approach for developed their application. The multiple requests are handled by the non-preemptive priority queuing model. Multiple users can request the different resources then they check the completion of processes and the scheduled processes states are true then the next scheduled request is performed. If the scheduled process state is false then the schedule request in waiting state. They perform two operations namely,
1. Uploading a file. 2. Downloading a file.

They use three folders for their experiment. When the upload or download process is performed the size of the folders is verified for load balancing. More than client sends the request at the same time the less priority is discarded in the queue then the new packet is inserted in the queue. When current users' activity completes then other high priority scheduled users task will be performed. Finally they analyze this method is suitable for user and service provider that means it provide the quality of service to the user and maximum profit for service provider.

2. Cost-Based Task Scheduling Algorithm:

Selvarani, et al [17] presented an improved cost-based scheduling algorithm for building competent mapping of tasks to existing resources in cloud. In this paper they schedule the task based on the cost of resources and computation performance. The available resources are grouped to the processing capability. The coarse-grained tasks are processed in the selected resources, so that the Computation-Communication ratio is reduced. Their objective is to optimize of makespan and optimize of cost. These steps are used to arrange the tasks according to their priority levels. Received tasks priority are calculated using this formula "$L_k = \sum_{i=1}^{K} R_{i,k} \text{ X } C_{i,k}/P_k$".
The tasks are sorted based on the priority and it will be stored in the high, medium and low priority levels. When the new task is arrived then the priority is calculate and this will be placed in the list. After arranging the task grouping and scheduling is performed. The total length of all tasks is smaller than or equal to resource Machine Instructions per second and the tasks are ungrouped then the current total length is calculated. If the total processing requirements of a task group is not equal to zero then the new tasks group is created then assign the unique ID to the newly created task group and it will be inserted into the task group list. Insert the allocated resource ID into the target resource list. After the execution of the task-groups by the assigned resources send them back to the target resource list. The implemented ABC algorithm performance is better than the original ABC scheduling algorithm.

3. Priority based job scheduling Algorithm:

Shamsollah, et al [18] proposed a new Scheduling algorithm based on the multi-criteria and multi-decision scheduling algorithm. This Scheduling algorithm consists of three level of scheduling: object level, attribute level and alternate level. The object Level is the scheduling level, the resource level is the attribute level and the job level is the alternative level.
In the analysis of proposed algorithm discusses about three important issues. They are complexity, consistency and finish time (makespan). The complexity of the proposed algorithm is mostly due to computing priority vectors of comparison matrixes. Consistency indicates that each of comparison matrixes has a logically reasonable value. The consistency of the proposed algorithm is mostly depends on the decision makers. Finish time (makespan) of the proposed algorithm is mostly focuses on priority of jobs. In this algorithm priority can be put by job resource ratio. Then priority vector can be correlated with every queue. This algorithm has advanced throughput and less finish time.

4. Compromised Time-Cost Scheduling Algorithm:

Ke, et al [19] proposed a novel based time cost scheduling algorithm which considers the uniqueness of cloud computing to contain instance- intensive cost-constrained workflows by compromising finishing time and cost with user contribution enabled on the fly. The imitation has established that CTC (Compromised-Time-Cost) Algorithm.
This algorithm needs to consider the following aspects: a) Background load: in order to estimate the execution time more accurately , background load is consider when calculating the task execution time on each specific server. b) Dynamic adjustment: in order to adapt to the dynamic change of load, after tasks are distributed to a server, the server may reschedule the tasks when encountering heavy load. c) Checking and rescheduling: in order to deal with the execution failures, uncompleted tasks will be rescheduled with a high priority for the next round. This algorithm (CTC) can achieve lesser cost than others while gathering the user-indicated deadline or decrease the mean finishing time then others within the user execution cost. The tool used for simulation is SwinDe W-C (Swinburne Decentralized Workflow for Cloud).

5. Reliable Scheduling Algorithm:

Arash, et al [20] proposed a reliable scheduling algorithm in cloud computing environment. This RSDC Algorithm is based on the PPDC (Processor-set Partitioning and Data Distribution) algorithm. The PPDD algorithm discuss about a general load balancing and scheduling with multiple loads originating from multiple processors in a single-level tree network. PPDD algorithm proposed just the running time of processors has been considered while the RSDC algorithm proposed that request and acknowledge time is considered.
In this RSDC algorithm, major job is separated to sub jobs. In order to stabile the jobs, the request and acknowledge time are deliberate individually. The scheduling of every job is complete by scheming the request and acknowledges time in the form of shared job. According to the observations that the variation rate of RDSC algorithm is much lower than the variation rate of PPDD algorithm. From this RDSC is much more efficient than PPDD algorithm. So that efficiently of the system is increased.

6. Priority based Service Scheduling

Dakshayini, et al [21] has proposed priority based service scheduling for cloud. The proposed architecture consists of two levels, Cloud service Provider Level (SPL) and User Level(UL). SPL provides a set of services to the user with appropriate communication between several components of the cloud. Different components of the cloud are request control manager, service manager in association with resource usage accounting manager. The user level provides

© 2014, IJARCST All Rights Reserved

secured access point between service provider and user.
The new scheduling algorithm is based on the admission control and priority scheme. In this algorithm priority is assign to every admitted queue. Access of each queue is determined by calculating tolerable hindrance and service cost. Benefit of this algorithm is that this policy with the future cloud architecture has accomplished very high (99%) service achievement rate with definite QoS. As this policy provide the highest preference for vastly rewarded user service-requests, on the whole servicing cost for the cloud also increases.

7. Max-Min Scheduling Algorithm using Petri Net:

El-Sayed et al [22] has proposed a new algorithm based on crash of RASA algorithm. The improved Max-min algorithm is based on the predictable finishing time instead of total time as a assortment basis. Petri nets are used to model the simultaneous performance of distributed systems. Experimental outcome shows availability of load balance in tiny cloud computing atmosphere and total little makespan in large scale dispersed system Max-min demonstrates achieving schedules with equivalent lower makespan rather that RASA and original Max-min.

8. Job Scheduling Algorithm based on Berger Model:

Baomin, et al [23] has proposed a new algorithm for scheduling process to establish dual fairness constraint. Based on the berger model this paper proposed job scheduling algorithm for first time. The first constraint is to categorize user tasks by QoS inclination, and set up the general expectation function in agreement with the arrangement of tasks to hold down the fairness of the resources in selection process.
The second constraint is to describe resource equality justice function to judge the fairness of the resource allocation. According to constraint the algorithm always allocate tasks on the optimal resources in order to persuade the QoS requirement of user and avoid considering a long task for execution. Research outcome of this algorithm shows successful execution of the user tasks and obvious better performance.

V. Comparison Among The Scheduling Algorithms

Scheduling Algorithm	Scheduling Method	Scheduling Parameters	Scheduling Factors	Findings	Environment
Optimistic Differentiated Job Scheduling System for Cloud Computing	Dependency mode	Quality of service, Maximum Profit	Single job with multiple user	1. The QoS needs of the cloud computing user and the highest income of the cloud computing service provider are accomplished.	Cloud Environment
Improved Cost-Based Algorithm For Task Scheduling In Cloud Computing	Batch mode	Cost, Performance	Unscheduled task group	1. Measures both resource cost and computation performance. 2. Improves computation ratio.	Cloud Environment
A Priority based job scheduling Algorithm in cloud computing	Dependency mode	Priority to each queue	An array of job queue	1. Less finish time.	Cloud Environment
A Compromised – Time – Cost Scheduling Algorithm.	Batch mode	Cost and time	An array of workflow instances	1. It is used to reduce cost and time	Cloud Environment
RSDC(Reliable Scheduling Distributed In Cloud Computing)	Batch mode	Processing time	Grouped Task	1. It is used to reduce processing time. 2. It is efficient for load balancing	Cloud Environment
An Optimal Model for Priority based Service Scheduling Policy for Cloud Computing Environment	Batch mode	Quality of Service, service request time.	An array of workflow instances	1. High QoS 2.High throughput	Cloud Environment
Extended Max-Min Scheduling Using Petri Net and Load balancing	Batch mode	Load balancing, finishing time.	Grouped task.	1. It is used to efficient load balancing. 2. Petri Net is used to remove the limitations of Max- min algorithm.	Cloud Environment
Job Scheduling Algorithm based on Berger Model in Cloud Environment	Batch mode	QoS Fairness constraint completion time.	Bandwidth of tasks	1. Improving task execution time and get better performance.	Cloud Environment

© All Rights Reserved, IJARCST 2014

ISSN : 2347 - 8446 (Online)
ISSN : 2347 - 9817 (Print)

VI. Conclusion

Scheduling is one of the most important key issues in the management of application execution in cloud environment. In this paper we have analyzed and surveyed various existing algorithms in cloud computing and tabulated various parameters. Existing algorithms gives high throughput and cost effective. They do not consider reliability and availability. So there is a need to implement scheduling algorithm that improve reliability and availability in cloud computing environment.

References

[1] *GAN Guo-ning, HUANG Ting-lei, GAO Shuai, "Genetic Simulated Annealing Algorithm For Task Scheduling Based on Cloud Computing Environment", IEEE, 2010, pp: 60 – 63.*

[2] *R. Buyya, C. S. Yeo, S. Venugopal, J.Broberg and I.Brandic, "Cloud computing and emerging IT platforms: vision, hype, and reality for delivering computing as the 5th utility", Future generation system, 2009, pp: 599-616.*

[3] *P.Mell and T.Grance, "The NIST Definition of Cloud Computing", National Institute of Standards and Technology, Information Technology Laboratory, Technical Report version 15 ,2009.*

[4] *http://www.kurzweilai.net/nist-issues-government-cloud computing-roadmap-and-architecture.*

[5] *Linan Zhu, Qingshui Li, and Lingna He, "Study on Cloud Computing Resource Scheduling Strategy Based on the Ant Colony Optimization Algorithm", IJCSI .International Journal of Computer Science Issues, Vol.9, Issue 5, No 2, Sep-2012, pp 54-57.*

[6] *Scalability: http://www.investopedia.com/terms/s/scalability.asp*

[7] *vertical-scalability: http://www.techopedia.com/definition/15323/*

[8] *Vertical Scalability: http://searchcio.techtarget.com*

[9] *horizontal scalability- http://searchcio.techtarget.com*

[10] *Somasundharam, Prabha, Arumugam, "Scalability issues in cloud computing", Advanced Computing (ICoAc), 2012, pp 1-5*

[11] *Sonal Dubey, Sanjay Agrawal., "QoS Driven Task Scheduling in Cloud Computing", International Journal of Computer Applications Technology and Research Volume 2– Issue 5, 595 - 600, 2013, pp 595-600*

[12] *Types of Scheduling: http://www.cim.mcgill.ca/~franco/OpSys-304-427/lecture-notes/node38.html.*

[13] *Types of Scheduling: http://en.wikipedia.org/wiki/Scheduling_%28computing%29.*

[14] *Types of Scheduling: http://orzhovgilden.site11.com/projects/CPU.html.*

[15] *Types of scheduling Algorithms: http://webdocs.cs.ualberta.ca/~tony/C379/Notes/PDF/05.4.pdf.*

[16] *Shalmali Ambike, Dipti Bhansali, Jaee Kshirsagar, Juhi Bansiwal " An Optimistic Differentiated Job Scheduling System for Cloud Computing" International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622 www.ijera.com Vol. 2, Issue 2,Mar-Apr 2012, pp.1212-1214.*

[17] *Mrs.S.Selvarani, Dr.G.Sudha Sadhasivam, "Improved cost-based algorithm for task scheduling in Cloud computing", IEEE 2010, pp: 1-5.*

[18] *Shamsollah Ghanbari, Mohamed Othman "A Priority based Job Scheduling Algorithm in Cloud Computing" International Conference on Advances Science and Contemporary Engineering, 2012 (ICASCE 2012, pp:147-152..*

[19] *Ke Liu, hai Jin, Jinjun Chen, Xiao Liu, Dong Yuan, Yun Yang, "A Compromised – Time – Cost Scheduling Algorithm in SwinDe W-C for Instance-Intensive Cost-Constrained Workflows on Cloud Computing Platform" International Journal of High Performance Computing Applications ,Volume 24 Issue 4, November 2010, ACM, pp:445-456.*

[20] *Arash Ghorbannia Delavar,Mahdi Javanmard , Mehrdad Barzegar Shabestari and Marjan Khosravi Talebi "RSDC (Reliable Scheduling Distributed In Cloud Computing)" International Journal of Computer Science, Engineering and Applications (IJCSEA) Vol.2, No.3, June 2012, pp:1-16.*

[21] *Dr. M. Dakshayini, Dr. H. S. Guruprasad "An Optimal Model for Priority based Service Scheduling Policy for Cloud Computing Environment" International Journal of Computer Applications (0975 – 8887) Volume 32– No.9, October 201, pp:23-29.*

[22] *El-Sayed T. El-kenawy, Ali Ibraheem El-Desoky, Mohamed F. Al-rahamawy "Extended Max-Min Scheduling Using Petri Net and Load Balancing" International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2, Issue-4, September 2012.*

[23] *Baomin Xu,Chunyan Zhao,Enzhao Hua,Bin Hu. —Job scheduling algorithm based on Berger Model in Cloud Environment‖ Advance in Engineering Software, ELSEVIER, 2011, pp.419-425.*

Biograhpy

Bithiah Blessie V R is doing MPhil research in the Department of Computer Science, St. Joseph's College (Autonomous), Tiruchirappalli, Tamil Nadu, India. She has attended International and National Conferences, Seminar and Workshops. Also presented papers in National Conferences and Seminars. She is presently working on Scalability issues in Cloud Computing Her area of research is Cloud Computing.

A. Stanislas is doing his Ph.D. in Computer Science in St. Joseph's College (Autonomous), Tiruchirappalli, Tamil Nadu, India. Prior he received his MCA at Loyala College, Chennai. He worked as Lecturer in St. Xavier's College, Dumka, Jharkhand. He has attended many national and international conferences and workshops, presented papers and published a few papers. He has also delivered a few guest lectures on Green and Cloud Computing in Seminars. His Research area includes Networking, Green Computing and Cloud Computing.

© 2014, IJARCST All Rights Reserved

Dr. Arockiam. L is working as Associate Professor in the Department of Computer Science, St. Joseph's College (Autonomous), Tiruchirappalli, Tamil Nadu, India. He has 25 years of experience in teaching and 17 years of experience in research. He has published more than 187 research articles in the International / National Conferences and Journals. He has also presented 2 research articles in the Software Measurement European Forum in Rome. He has chaired many technical sessions and delivered invited talks in National and International Conferences. He has authored a book on "Success through Soft Skills". His research interests are: Big Data,Cloud Computing, Software Measurement, Cognitive Aspects in Programming, Data Mining and Mobile Networks. He has been awarded "Best Research Publications in Science" for 2010, 2011, & 2012, "Best Teacher Award" for 2012-13, 2013 -14 and ASDF Global Awards for "Best Academic Researcher" from ASDF, Pondicherry for the academic year 2012-13.

© All Rights Reserved, IJARCST 2014