CLASSIFICATION OF RESOURCE MANAGEMENT AND PRICING MODELS IN CLOUD COMPUTING

¹DILIP KUMAR S M, ²NAIDILA I PATIL

¹Dept. of Computer Sc. Engg., University Visvesvarya College of Engg., Bangalore, India 560 001 ²Dept. of Computer Sc. Engg., Acharya Institute of Technology, Bangalore, India 560 107 E-mail: dilipkumarsm@gmail.com, sadashiv@acharya.ac.in

Abstract- The wide acceptance of Cloud Computing by small and medium enterprises to large enterprises for achieving computing service has been the latest trend. With such a move enormous demand for the resources arises in different dimensions. Due to the nature of cloud, dynamic allocation of resources to provide desired services among different users in a cost effective way becomes very challenging and essential among the cloud providers. They incorporate different resource management techniques to meet the QoS specified by the users. To have a clear insight on the resource allocation methods used in cloud environment and to come up with better strategies, in this paper we provide some of the previous resource management techniques and analyze them. These approaches are classified rigorously from different perspectives and analyzed along with their merits and demerits. Further, we present different pricing models used during resource allocation. This paper could bring an intuition among researchers in developing new optimal resource management and pricing methods such that cloud computing potential can be reaped to the fullest extent.

Keywords- Cloud Computing, Resource management, Pricing model.

I. INTRODUCTION

Cloud computing the popularly used model is a distributed system consisting of a collection of interconnected and virtualized computers that are

dynamically provisioned based on SLA. It provides a pool of computing resources which the users can access based on pay-as-you use concept. The basic principles of cloud computing is to shift the computing done from the local computer into the network. Small medium and large enterprises procure resources which include server, network, storage and application services without huge investment on its purchase and maintenance, but use it for getting their work done easily. Managing resources in such large dynamic cloud environments is a very complex and challenging task due to the dynamic nature of the cloud. In view of this, efficient resource management techniques are the prerequisite in order to gain the real benefit of cloud model. Resource management includes scheduling the applications, allocation, control and coordination of the resources among the virtual machines (VMs). Resources can be requested on-demand or can be reserved which guarantees the availability of resources which are instances of metered services. Economic models for such resource

management (RM) techniques should provide cost effective services to consumers along with good profit margin for the service providers. This has motivated the survey on literature and to discuss several resource management and pricing techniques used at various levels during the service of the request. Different perspectives towards the classification of resource management and pricing techniques are presented based on their relevance to several factors.

II. CLASSIFICATION OF RESOURCE MANAGEMENT TECHNIQUES AND PRICING MODELS

In this section, we present the classification of the existing RM and pricing techniques based on different criteria and are explored in tabular form. A. Classification based on resource management techniques used during resource negotiation Managing resources in large dynamic cloud environments is a very complex and challenging task as requests for the type and amount of resources is not known in advance most of the time [4]-[6]. Table I presents the classification of RM techniques used during resource negotiation.

TABLE I CLASSIFICATION BASED ON RESOURCE MANAGEMENT TECHNIQUES USED FOR RESOURCE NEGOTIATION.

	Resource Allocation Method/Technique	Efficiency	Platform Used		Application Supported	Merits/Demerits
for Provisioning- Based Resource and Application	interfaces or resource	performance	simulations o		Large-scale scientific applications	Superior performance for the application at a realistic price
Computational	Flexible Advanced	failure and reallocation rate	Distributed metascheduler		Diverse synthetic workloads	It increase occu- pancy and thereby increase the obtain- able utilization of the system
Dynamic Grid	based with myopi	Prevents congestion & collapse of gric with improved long- term profits	framework	y		Hedging using op- tions improves re- source allocation

International Journal of Advances In Computer Science and Cloud Computing, ISSN: 2321-4058

TABLE II
CLASSIFICATION BASED ON RESOURCE SCHEDULING TECHNIQUES USED FOR RESOURCE
MANAGEMENT.

		MANAOL				
Literature Work	Resource Allocation	Efficiency	Platform	Scalability/QoS	Application	Merits/Demerits
	Method/Technique		Used		Supported	
Cloud Computing	Hybrid genetic algorithm	Improved resource	Experiment	Algorithm	Any application	Improved genetic
Resource Scheduling	combined with knapsack	utilization and saves	done in a	supports	that can be	Algorithm calculate:
Policy Based on Genetic	problem. Pre-migration	energy cost	simulated	scalability to	specified in three	more precisely
Algorithm with Multiple	strategy		environment	any extent		with reasonable
Fitness [33]						scheduling policy
A Cooperative Game	Nash Bargaining Solution	Provides a Pareto	lt is a	Optimal	High-performance	Accounts for
Framework for QoS	based game theoretic	optimal solution to	centralized	solution to	computing	fairness
Guided Job Allocation	solution	the QoS objective,	strategy	the QoS	applications	to users, but is less
Schemes in Grids [34]		relatively low	4			fault-tolerant
		overhead and good				
		performance				
A Hierarchical Approach		Satisfies	Trace driven	Provides	Transactional	Unifying framework
for the Resource			simulation			supports capacity
Management of Very		availability		distributed	applications	allocation and load
Large Cloud Platforms	adaptive machine	guarantees and		resource		balancing
[35]	0 11	minimize energy	1	management		
		costs.				
Dynamic Resource	Live VM migration	Improves overall	Trace	With higher	Combines different	Achieves overload
Allocation using Virtual		utilization of server,			types of workloads	
Ma- chines for Cloud		pre- vent overload in	simulation		to improve the	0
	0	the system		is distributed		
Environment [36]		effectively			of server resources	
A Budget Constrained		Meet users	Grid			Automatic
Scheduling of		budget constraint and execution time	testbed			handling of overload & under
Workflow Applications on Utility Grids using		minimization.				utilization during
Genetic Algorithms	con-verge			users QoS		job execution is
[43]	U			constraint		not addressed

are done using strategies such as Just-in-Time Bidding, Flexible Advanced Reservations, Two Phase Contracts, Second Chance Substitute Providers, Advertisement and query based interfaces. Platforms such as Distributed meta-scheduler, Matlab and Discrete time simulators are considered for conducting the experiments. These method provides improved resource utilization and ensures availability of resources followed by achieving the desired QoS for the consumers.

B. Classification based on resource scheduling techniques used for resource management

Resources are scheduled based on cost constraints, time constraints and QoS for which different scheduling algorithms are considered. Table II presents the classification based on resource scheduling techniques used for RM. Max-Min and Min-Min, Gang scheduling, Priority based scheduling, genetic algorithm combined with knapsack problem, aggressive and conservative backfill algorithms are used for scheduling the resources. They meet the QoS within the budget constraints and in the given time from the consumer's perspective.

C. Classification based on techniques used in federated clouds

In view of accommodating all the requests in the cloud with minimal rejections, federated clouds are

used where in cloud providers give their unused capacities during low demand periods and take spare capacity during their peak time in order to maximize their profits. Table III provides the details for techniques used in federated clouds. Game theory, Semi-Markov Decision Process, push-pull, live virtual machine migration are the techniques used in federated cloud.

VM sharing decision provides good resource utilization followed by optimal execution efficiency and maximized profit for the providers.

D. Classification based on resource management techniques used for different types of workload. Ad hoc parallel data processing are one of the killer applications for IaaS clouds.

Table IV provides the details of techniques used for different types of workload. Static and dynamic allocation of resources for data processing are supported, wherein dynamic allocation provides better utilization of heterogeneous resources.

Consolidation of workloads, parametric estimation, feedback control model, non-cooperative game theory and adaptive mechanism are used to schedule VM which improves minimizes migration overheads and maximize mixed workload performance with guaranteed optimal benefits within time constraints.

E. Classification based on load balancing and fault management techniques used in resource management

Unpredicted bursty request in cloud could arise at any time for which the cloud should be able to handle. It is addressed by extending the capacity of a local cloud, finding least loaded system or a general mobile cloud computing system which consists of multiple cloud domains each with its own set of resources. Table V provides the details for load balancing and fault management related work to RM. Semi-Markov Decision, Gossiping protocols, pushcheckpointing-recovery and trace-replay pull, methods are used for load balancing and fault management. Fair resource allocation among sites/applications is obtained by dynamic adaptation in response to the load changes and any faulty situation in the cloud.

F. Different pricing and penalty models

Based on the resource allocation methods different pricing models are considered in the literature. Table VI gives the details of different pricing and penalty model. These models dynamically adjust the price due to the nature of cloud. Static pricing will either burden the consumer or might lead to revenue loss to the provider, therefore dynamic pricing is an efficient way of pricing. Based on demand, availability of resources, time and QoS constraints pricing models are set. Some pricing models attempts to make use of the discounts available via spot instances but there is no certainty of the resource availability and its price. Static and dynamic penalty functions are also considered in case there are agreement breaches which are obtained based on Win Price, Substitute Price or Bid Difference.

TABLE III

CLASSIFICATION BASED ON RESOURCE MANAGEMENT TECHNIQUES USED IN FEDERATED CLOUDS.

Literature Work	Resource Allocation		Platform	Scalability/QoS		Merits/Demerits
	Method/Technique		Used		Supported	
A Novel Economic	Game theory, dynamic	Profit obtained from	Matlab	Supports sala-	All cloud	Long-term revenue
Sharing Model in a	programming approach	repeated game derives		bility	applications	maximization based on
Federation of Selfish	along with grim trigger	a higher revenue using				the future workloads
Cloud Providers [37]	strategy is used	a simple grim trigger				uncertainty
		punishment strategy				
	Service request decision			Supports scala-	All cloud	Minimal no. of ser-
	making process is mod-		Matlab	bility	applications	vice rejections which
	eled as a Semi-Markov	and the users				degrade user satisfac-
cation in Mobile Cloud	Decision Process					tion level
Networks [38]						
Dynamic Optimization		Achieves maximized			All cloud	Protocol produces one
of Multiattribute RA in		resource utilization		ability through		lightweight query mes-
Self-Organizing Clouds		and delivers optimal		self organizing		sage/ task and sup-
		execution efficiency		clouds		ports high adaptabil-
	query protocol					ity in dynamic node-
Efficient Resource	Hierarchical framework	Costefficiency over	A discrete	Supports	All cloud	Distributed intra-cloud
Mapping Framework	for inter-domain	a large number of	event	scalability	applications	resource mapping
over Networked Clouds	F F F F F F F F F F F F F F F F F F F	Virtualetwork request	java-based			approach allows for
via Iterated Loca	in a networked cloud	instances and	simulator			effi- cient and
Search based Request	environment with an	networked cloud sizes				balanced al- location
Partitioning[40]	iterated local search	with minimum				of cloud re- sources
Decentralized Meta-	Meta-broker that pro-		Cloudsim	Supports	Massive	Provides effective
brokers for Inter-Cloud	vides an autonomous	multiple brokers into a		scalability as it	amount of	performance levels
Modeling brokering	orchestrator	single aggregated view	1	connects across	services	with meta-brokering
coordinators for inter-	-	identical to distributed		clouds	oriented	solution
operable RM [41]		resource managers			application	

TABLE IV

CLASIFICATION BASED ON RESOURCE MANAGEMENT TECHNIQUES FOR DIFFERENT TYPES OF WORK LOAD.

Literature Work	Resource Allocation	Efficiency	Platform	Scalability/QoS	Application	Merits/Demerits
	Method/Technique		Used		Supported	
Autonomic Placement	Consolidation of work-	Maximizes mixed	Discrete event	Achieves QoS	Heterogeneous	Maximizes
of Mixed Batch &	loads on the same	workload	based		set of mixed	performance and
Transactional Workloads	physical hardware and	performance	simulator		workloads	provides service
[42]	Leveages visualization					differentiation
	control mechanisms					based on goals.
Exploiting Dynamic RA	Task are represented as	Improves utilization	Eucalyptus	Automatically	Parallel data	Automatic handling
for Efficient Parallel	DAG and uses feedback	and reduce the		allocate/deallocate	processing	of overload and
Data Processing [43]	based monitoring data	processing cost		VMs in execution	applications	underutilization
Automatic VM	Epidemic parameter	Speeding up the	Private cloud	Scalability of	Analytic and	Speed up of the
Allocation for scientific	estimation algorithm	execution time for	and A*STAR	parameter es-	scientific	execution time
application [44]	and multiple loop	scientific		timation module in	applications	with prediction
	decomposition	applications		epidemic forecast		accuracy
	algorithm.					
Resource Provisioning	Multi-input-multi- output	Guarante ed	Linux clusters	Provides required	Real world	Maximizes application
with Budget	feedback control	optimal		QoS with help of	adaptive applicatio	QoS, within time
Constraints for	model with	application		feedback control	ns	constraint and
Adaptive Applications	reinforcement	benefit within		model		budget limit
in Cloud Environments	learning	the time constraint				
[45]		constraint				

International Journal of Advances In Computer Science and Cloud Computing, ISSN: 2321-4058

TABLE V

CLASSIFICATION BASED ON LOAD BALANCING AND FAULT MANAGEMENT TECHNIQUES USED IN RESOURCE MANAGEMENT.

	Resource Allocation Method/Technique	Efficiency	Platform Used	Scalability/QoS	Application Supported	Merits/Demerits
via Asynchronous Replication & State Synchronization [48]	, ,	1		Supports scala- bility and re- quired QoS	Kernel-intensive workloads and I/O intensive workloads	Provides reduced migration downtime and network
Provisioning to Extend the Capacity of Loca	ing strategy with check- pointing in conservative and aggressive selective	,	Cloudsim	Supports scalability	High-performance Applicatios requiring large no. of resources over short periods or time	takes into
A Gossip Protocol for Dynamic RM in Large Cloud Environments [50]	push-pull paradigm & uses CYCLON ar overlay protocol in a	Protocol continuously executes on dynamic, local input and does not require global synchronization	event simulator		1	Fair resource allocation, dynamically adapts the allocation to load changes

TABLE VI PRICING AND PENALTY MODEL

Literature Work	Pricing Model	Description	Efficiency	Dynamic Pricing
Strategies for Computational Economies	Available capacity Win/loss ratio Time based	Pricing based on projected provider capacity, win/loss ratio ,time since the provider last won an auction	substitute strategies provides revenue	All models support dynamic pricing
An End-to-End Framework for Provisioning-Based Resource and Application Management [32]			offerings in order to accommodate users with	
Efficient Resource Mapping Framework over Networked Clouds-Iterated Local Search based Request Partitioning [40]	like availability of resources	based on Iterated Local Search		the resource with average utilization
A Mechanism of Specifying and Determining Pricing in Utility Computing Environ- ments [46]	based on Exponential Weighted Moving Average	ply & demand curves over	applications are guaranteed	Captures dynamic demand on resources and adapts pricing to dynamic demand
Commodities: A Novel Financial Economic Model	with	depreciation, quality of ser-	constraints from providers perspectives	Cloud pricing problem has not included providers cost of maintenance and other costs

CONCLUSION

Cloud computing is the popularly used utility paradigm for providing services to various end users. Resources are delivered through reservation plan or can be requested on the fly and the pricing model is set accordingly. The benefit of cloud computing can be achieved to maximum extent through efficient resource management. This paper focuses on the various resource management and pricing techniques along with their efficiency and constraints. Many of these focus on managing the resources by considering only a few parameters. A still better approach for resource management will help in diverse usage of cloud computing. The classification made here would lead for developing new resource management techniques and pricing model in future work.

REFERENCES

- S.Son, K. M. Sim, "An adaptive tradeoff algorithm for multi-issue SLA negotiation", 121, pp. 32-41, Dec. 2010.
- [2] M. Mazzucco, M. Dumas, "Reserved or On-Demand Instances A Revenue Maximization Model for Cloud Providers", IEEE Intl. Conf. on Cloud Computing, pp. 428-235, Jul. 2011.
- [3] Borja Sotomayor, Rubn S. Montero, Ignacio M. Llorente, Ian Foster, "Virtual Infrastructure Management in Private and Hybrid Clouds", IEEE Internet Computing, 13(5), pp. 14-22, Sep. 2009.
- [4] K. M. Sim, B. Shi, "Concurrent Negotiation and Coordination for Grid Resource Coallocation", IEEE Transaction on Systems, Man, and Cybernetics, Part B, Cybernetics, 40(3), pp. 753-766, May 2010.
- [5] K. M. Sim, "Grid Resource Negotiation: Survey and New Direction", IEEE Transaction on System, Man,and

International Journal of Advances In Computer Science and Cloud Computing, ISSN: 2321-4058

Cybernetics, Part C, Application and Reviews, 40(3), pp. 245-257, May 2010.

- [6] T. Goiri, F. Julia, J.Oriol Fito, M.Macias, J. Guitart, "Supporting CPUbased Guarantees in Cloud SLA's via Resource Level", Future Generation Computer System, 28(8), pp. 1295-1302, Oct. 2012.
- [7] A.W. Mualem, D. G. Feitelson, "Utilization, Predictability,Workloads and User Runtime Estimates in Scheduling the IBM SP2 with Backfilling", IEEE Tr. on Parallel and Distributed Systems, 12(6), Jun. 2001.
- [8] K. Bubendorfer, "Fine Grained Resource Reservation in Open Grid Economies", Proc. 2nd IEEE Intl. Conf. on E-Science and Grid Computing, pp. 1-18, Dec. 2006.
- [9] K. Bubendorfer, Macro A. Netto, R. Buyya, "SLA-based advance reservations with flexible and adaptive time QoS parameters", Proc. 5th Intl. Conf. on Service-Oriented Computing, pp. 119-131, Jan. 2007.
- [10] I. Foster, C. Kesselman, C. Lee, R. Lindell, K. Nahrstedt, A.Roy, "A Distributed Resource Management Architecture That Supports Advance Reservation and Co-Allocation", Proc. Intl. Workshop on Quality of Service, pp. 27-36, Jun. 1999.
- [11] S Sindhu and Saswati Mukherjee, "Efficient Task Scheduling Algorithms for Cloud Computing Environment", High Performance Architecture and Grid Computing, 169, pp. 79-83, Jul. 2011.
- [12] Gao Ming, Hao Li, "An Improved Algorithm Based on Max-Min for Cloud Task Scheduling", Recent Advances in Computer Science and Information Engineering, IJCA, 50, pp. 217-233, May 2012.
- [13] Selvarani S. and Sadhasivam G.S., "Improved cost-based algorithm for task scheduling in cloud computing", IEEE Intl. Conf. on Computational Intelligence and Computing Research, pp. 1-5, Dec. 2010.
- [14] Guiyi Wei, A. V. Vasilakos, Yao Zheng, Naixue Xiong, "A Game- Theoretic Method of Fair Resource Allocation for Cloud Computing Services", The Jr. of Supercomputing, 54(2), pp. 252-269, Nov. 2010.
- [15] Zhenping Lu, Xiangming Wen and Yong Sun, "A game theory based resource sharing scheme in cloud computing environment", IEEE World Congress on Information and Communication Technologies, pp. 1097-1102, Nov. 2012.
- [16] M. Mattessa, C. Vecchiola, R. Buyya, "Managing Peak Loads by Leasing Cloud Infrastructure Services from a Spot Market", 12th IEEE Intl. Conf. on High Performance Computing and Communication, pp. 180-188, Sep. 2010.
- [17] T. Goiri, J. Guitart, J. Torres, "Economic Model of a Cloud Provider Operating in a Federated Cloud", Information Systems Frontiers, 12(4), pp. 827-843, Apr. 2012.
- [18] B. Rochwerger, et. al., "The Reservoir Model and Architecture for Open Federated Cloud Computing", IBM Jr. of Research and Development, 53, no. 4, pp. 535-545, Jul. 2009.
- [19] A. Au Young, B. N. Chun, A. C. Snoeren, A. Vahdat, "Resource Allocation in Federated Distributed Computing Infrastructures", Proc. 1st Workshop on Operating System and Architectural Support for the On-Demand IT Infrastructure, Feb. 2004.
- [20] C. Adam, R. Stadler, "Service Middleware for Self-Managing Large Scale Systems", IEEE Tr. on Network and Service Management, 4(3), pp. 50-64, Dec. 2007.
- [21] Ronnie Chaiken, Bob Jenkins, Bill Ramsey, Darren Shakib, Simon Weaver, Jingren Zhou, "SCOPE: Easy and Efficient Parallel Processing of Massive Data Sets", Very Large Data Bases Endowment Inc, 1(2), pp. 1265-1276, Aug. 2008.

- [22] Gurmeet Singh, C. Kesselman, E. Deelman, "An End-to-End Framework for Provisioning-Based Resource and Application Management", IEEE Systems Jr., 3(1), pp. 25-48, Mar. 2009.
- [23] He Huang, Liqiang Wang, "P and P: A Combined Push-Pull Model for Resource Monitoring in Cloud Computing Environment", IEEE Intl. Conf. on Cloud Computing, pp. 260-267, July 2010.
- [24] Sangho Yi, Kondo, D. and Andrzejak A., "Reducing Costs of Spot Instances via Checkpointing in the Amazon Elastic Compute Cloud", IEEE Intl. Conf. on Cloud Computing, pp. 236-243, Jul. 2010.
- [25] Xuan Li, "Pricing and peak aware scheduling algorithm for cloud computing", IEEE PES Innovative Smart Grid Technologies, pp. 1-7, Jan. 2012.
- [26] Hao Li, Jianhui Liu and Guo Tang, "A Pricing Algorithm for Cloud Computing Resources", IEEE Intl. Conf. on Network Computing and Information Security, pp. 69-73, May 2011.
- [27] Tsakalozos K., Kllapi H., Sitaridi E., Roussopoulos M., Paparas D. and Delis A., "Flexible use of cloud resources through profit maximization and price discrimination", IEEE Intl. Conf. on Data Engineering, pp. 75-86, Apr. 2011.
- [28] X. Bai, D. Marinescu, L. Boloni, H. Jay Siegel, R. Daley, I. Wang, "A Macroeconomic Model for Resource Allocation in Large-Scale Distributed System", Jr. of Parallel and Distributed Computing, 68(2), pp. 182-199, Feb. 2008.
- [29] K. Chard, "Drive: A Distributed Economic Meta-Scheduler for the Federation of Grid and Cloud Systems", Victoria University of Wellington, 2011.
- [30] Kyle Chard, and Kris Bubendorfer, "High Performance Resource Allocation Strategies for Computational Economies", IEEE Tr. on Parallel and Distributed Systems, 24(1), pp. 72-84, Jan. 2013.
- [31] Amir Danak, Shie Mannor, "Efficient Bidding in Dynamic Grid Markets", IEEE Tr. on Parallel and Distributed Systems, 22(9), pp. 1483-1496, Sep. 2011.
- [32] Gurmeet Singh, Carl Kesselman, Ewa Deelman, "Adaptive Pricing for Resource reservations in Shared Environments", IEEE/ACM Intl. Conf. on Grid Computing, pp. 74-86, Sep. 2007.
- [33] Zhen Xiao, Weijia Song, and Qi Chen, "Dynamic Resource Allocation using Virtual Machines for Cloud Computing Environment", IEEE Tr. On Parallel and Distributed Systems, pp. 1 - 11, Jun. 2012.
- [34] Rilky Subrata, Albert Y. Zomaya, Bjorn Landfeldt, "A Cooperative Game Framework for QoS Guided Job Allocation Schemes in Grids", IEEE Tr. on Computers, 57(10), pp. 1413-1422, Oct. 2008.
- [35] Addis Bernardetta, Ardagna Danilo, P. Barbara, Squillante Mark S, Zhang Li, "A Hierarchical Approach for the Resource Management of Very Large Cloud Platforms", IEEE Tr. on Dependable and Secure Computing, 10(5), pp 253-272, Oct. 2013.
- [36] Yazir Y.O., Matthews C., Farahbod R., Neville S., Guitouni A., Ganti S. and Coady, Y., "Dynamic Resource Allocation in Computing Clouds Using Distributed Multiple Criteria Decision Analysis", IEEE Intl. Conf. on Cloud Computing, pp. 91-98, Jul. 2010.
- [37] Nancy Samaan, "A Novel Economic Sharing Model in a Federation of Selfish Cloud Providers", IEEE Tr. on Parallel and Distributed Systems, pp. 1-11, Jan. 2013.
