



Enhancing the High Performance Resource Allocation and Data Sharing in Cloud Computing

Ramya Krishnasamy¹ and P.G.Kathiravan²

M.Tech II Year Student, Department of Information Technology, V.S.B Engineering College, Karur, Tamilnadu, India¹

Assistant Professor, Department of Information Technology, V.S.B Engineering College, Karur, Tamilnadu, India²

ABSTRACT: Utility computing is used for academic research. In fifth utility model the recent success of commercial cloud providers, computation and storage is being realized. The computational economies are performs major role to proposed as an efficient means of resource allocation. It has been limited due to lack of performance and high overheads. To achieve a high performance resource allocation, addresses the performance limitations and regeneration theory by defining the strategies to reduce the failure, reallocation rate, increase occupancy and increase the utilization of the system. These strategies include overbooking, advanced reservation, just-in-time bidding and using substitute providers for service delivery. These proposed strategies have been evaluated with respect to grid and cloud environments.

KEYWORDS- Utility computing, Resource allocation, Grid computing, Cloud computing

I. INTRODUCTION

Cloud computing has successfully applied economic (utility) models for computation. Consumers can now outsource computation, storage and pay only for the resources used. The computation market could be realized by a high-performance confederation architecture that spans both Grid and cloud computing providers. Computational economies are used to allocating resources in both centralized and decentralized computing systems. Adoption of economies in production systems has been limited due to poor performance, high latency, and high overheads. Overheads in computational economies can be explained, for example, in a competitive economy resources are typically “reserved” by m participants for the duration of a negotiation. In more cases, there are only n “winning” participants, therefore all other m-n reservations are necessary wasted for the duration of that negotiation. This type of scenario is clearly evident in auction or tender market. The suggested application of two general principles for address these inefficiencies: first, to avoid commitment of resources, and second, to avoid repeating negotiation and allocation processes. To filtered these principles into five high-performance resource utilization strategies, such as overbooking, advanced reservation, just-in-time (JIT) bidding, progressive contracts, and using substitute providers. These resource allocation strategies can be employed either through allocation protocols or by participants used to increase resource occupancy and therefore optimize overall utilization. These techniques are used in the context of a market-based cloud or Grid using the DRIVE Meta scheduler. Each strategy has been implemented in DRIVE for analyzing the workloads. These analysis focuses on two different areas: first the strategies are evaluated with respect to increased occupancy, allocation, and system utilization to compare their value in a range of economic and non economic scenarios. Second, due to the economic motivation of this work, these strategies are evaluated with respect to the system and service provider revenue to determine the effect of economic policies on the strategies themselves and also the effect of these strategies on revenue.

II. RELATED WORKS

The computational market was the future market that enabled users to invite for compute time on a shared departmental machine. These machines use the distributed computational economies, such as Spawn and Enterprise, modern brokers, metaschedulers and distributed architectures such as Nimrod/G, DRIVE and SORMA. DRIVE is the

metascheduler and is designed around the idea of “infrastructure free” secure cooperative markets. Overbooking has been previously used in computational domains to increase utilization and profit, to compensate for poorly estimated task duration. Backfilling is combined with overbooking to increase provider profit. The overbooking decisions are based on SLA risk assessment it can be generated from job execution time distributions.

In Globus Architecture for Reservation and Allocation (GARA) is one of the first projects to define a basic advanced reservation to support Quality of Service reservations over different resources. Catalina and Sun Grid Engine are the other schedulers have used to support advanced reservations. Reservation aware schedulers also used to improve system utilization due to the additional tractability specified by some consumers.

In Resource allocation, various concepts are looked at last minute bidding and sniping. These concepts are mainly used in online auction. The typical last minute biddings are 1. Shrink bidders and 2. Incremental bidding. Just-in-time bidding was first proposed to reducing the effect of auction latency in distributed auctions.

III. DRIVE ARCHITECTURE

The DRIVE (Distributed Resource Infrastructure for a Virtual Economy) architecture is used for efficient resource allocation. It is the first distributed economic meta scheduler designed to allocate workload in distributed and federated computing environments. DRIVE abstracts an economic market which allows any economic protocol to be used. The core meta scheduling services are hosted on participating resource providers as a condition of joining the Virtual Organization (VO). This DRIVE architecture minimizes the need for committed infrastructure and distributes management functionality across participants which provide security guarantees insure environments.

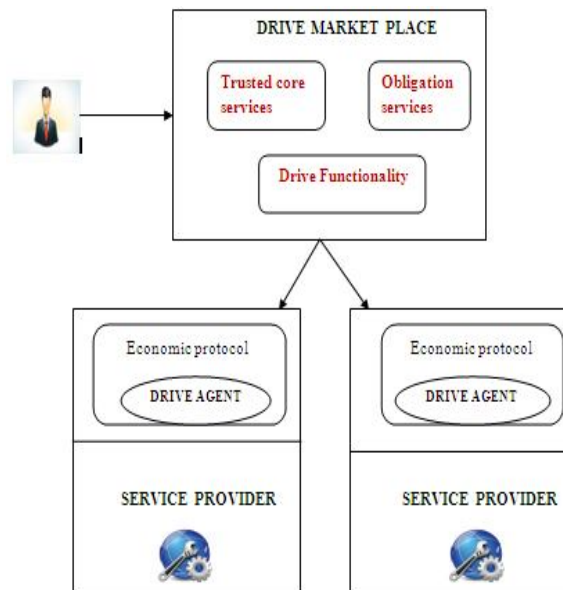


Figure3.1 DRIVE ARCHITECTUTE



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol.2, Special Issue 1, March 2014

Proceedings of International Conference On Global Innovations In Computing Technology (ICGICT'14)

Organized by

Department of CSE, JayShriram Group of Institutions, Tirupur, Tamilnadu, India on 6th & 7th March 2014

DRIVE architecture (fig.3.1) contains three types of components, 1.Service provider: each resource provider or service provider is represented by a DRIVE Agent. DRIVE agent which implements these standard functionalities such as, reservations, policies, valuation, and plug-ins for choosing the economic protocol, 2.DRIVE agent: it uses various policies and pricing functions to price the goods, 3.DRIVE market place: it includes a number of independent services and mechanisms to provide common functionality including resource discovery, allocation, security, Virtual Organization(VO) management, and contract management.

IV. HIGH UTILIZATION STRATEGIE

The economic negotiation presents a number of opportunities to implement utilization for improving policies and strategies before, during, and after negotiation. The following high-performance strategies are defined according to a traditional auction model.

4.1 Overbooking

Overbooking is to provide substantial utilization, profit advantages and over estimated task duration. Overbooking is a common technique used in yield management and it can be seen in many commercial domains such as, air travel and bandwidth reservation. Airlines can be routinely overbooking the aircraft to maximize occupancy by ensuring they have the maximum number of passengers on a flight.

Overbooking policies are carefully created and it is generally based on historical data. It acknowledges the possibility of an unusually large proportion of customers showing up, and includes clauses in their contracts to find passengers to later flights and compensate passengers financially. For example, an auction generally has a single winner, and multiple m losers. The winner gains the eventual agreement, there is no such compensation for the m losers of the auction process, any resources r put away during the auction will decrease the net usage of the system by mr.

4.2 Just-in-time Bidding

During the negotiation that resource state may change and therefore cancel a provider's valuation. There are two ways to minimize the effect of latency:

1. Cutting down the duration of the auction. The problem of this approach is the minimal time for providers to discover the auction and to calculate their bids.

2. Bid as late as possible. The main advantage of this approach is that providers can calculate their bids with the most up to date resource state and the resources are reserved for a shorter time. The primary problem of this approach is time sensitivity.

4.3 Advanced Reservations

Advanced reservation is commonly used in distributed system which provides performance predicate, to assemble resource requirements, and provide quality of service (QoS) guarantees. In Grid and cloud systems to evolve the task of planning job requirements is becoming more complex, requires the coordination of interdependent jobs in order to achieve larger goals. These tasks are available at particular resources in order to run efficiently.

4.4 Substitute Providers

The substitute provider technique can reduce the allocation losers generated from overbooking and therefore increase utilization of the system. The damaging aspect of this approach is the potential for increased consumer cost, as the substitute price (SP) is, greater than the privilege winning price. This cost can be offset through compensation provided by the violating party. The potential impact on auction protocols, all participants must be aware of the auction.



V. EXPERIMENTAL FUNCTIONS

The rating functions are primarily based upon local information known by the provider and aim to incorporate the perceived risk of a deal. The penalty function provide a formal way of categorizing agreement breaches and punishing responsible parties.

5.1 Pricing Functions

The bid prices are determined based on a combination of current conditions, projected conditions. In the following equations, P_{unit} is the price per job unit and b denotes a bid in the specified range. Job units (JU) are defined as the product of CPU utilization and duration ($J_{units}=J_{utilization} \times J_{duration}$). The Random and Constant pricing functions are baseline functions. The other pricing functions attempt to simulate some aspect of a dynamic supply and demand model, based on these functions the prices are adjusted dynamically to reflect the projected capacity or to factor in previous successes.

- Random: unit price is determined irrespective of any other factors

$$P_{unit} = \text{Random}(0, B)$$

- Constant: unit price is the same for every request

$$P_{unit} = c, c \in (0, B)$$

- Available capacity: the unit price is calculated based on projected provider capacity at the time when the job would execute. $U_{provider}$ is the projected utilization of the provider, U_{job} is the utilization of the requested job and $C_{provider}$ is the total capacity of the provider

$$P_{unit} = \frac{U_{provider} + U_{job}}{C_{provider}} \times B$$

- Win/loss ratio: the unit price is based on the previous win/loss ratio seen by the individual provider. R is the specified ratio, W is the number of wins since time t_0 , and L is the number of losses since time t_0

$$P_{unit} = (RW - L) \times \frac{B}{R} + \frac{B}{2}$$

- Time based: the unit price is based on the time since the provider last won an auction. The unit price decrements every T_{period} seconds, $T_{lastwin}$ is the time since the last allocation. $T_{lastwin}$ is set to 0 at time t_0

$$P_{unit} = B - \left(\frac{T_{lastwin}}{T_{period}} \right)$$

5.2 Penalty Functions

Penalty functions divided into two distinct penalty types: Constant penalties are fixed penalties that are statically defined irrespective of any other factors, whereas Dynamic penalties are based on a non-static variable designed to reflect the value of a violation.

- Constant: constant penalty defined statically irrespective of the job requirements or bid price



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol.2, Special Issue 1, March 2014

Proceedings of International Conference On Global Innovations In Computing Technology (ICGICT'14)

Organized by

Department of CSE, JayShriram Group of Institutions, Tirupur, Tamilnadu, India on 6th & 7th March 2014

$$P_{\text{defaulter}}=c, c \in \mathbb{R}_{\geq 0}$$

- Job units: an α penalty based on the requirements of the job in units. J_{units} is the number of units in a job, and c is a constant penalty per unit

$$P_{\text{defaulter}}=J_{\text{unit}} \times c, c \in \mathbb{R}_{\geq 0}$$

- Win price (WP): an α penalty based on the winning bid (pre-substitutes). $\text{Price}_{\text{win}}$ is the price to be paid by the winning bidder

$$P_{\text{defaulter}}=\text{Price}_{\text{win}}$$

- Substitute price: an α penalty based on the substitute bid. $\text{Price}_{\text{substitute}}$ is the price to be paid by the substitute winning bidder

$$P_{\text{defaulter}}=\text{Price}_{\text{substitutes}}$$

- Bid difference (BD): a β penalty defined as the difference between the original win price and the substitute price

$$P_{\text{provider}}=\text{Price}_{\text{substitute}}-\text{Price}_{\text{win}}$$

VI. EVALUATION

In the evaluation, each of the strategies is evaluated empirically with respect to allocation of resources occupancy, utilization, and economic implications. The occupancy can be defined as the number of contracts satisfied and utilization as the amount of a host's resource capacity that is used. These strategies are evaluated under different workload scenarios on a testbed deployment.

6.1 Synthetic Workloads

Several synthetic workloads that simulate different workload conditions. Every synthetic workload is derived from a production Grid trace obtained from AuverGrid, the small sized Grid in the Enabling Grids for E-science in Europe (EGEE) project which uses Large hadron collider Computing Grid (LCG) middleware.

The AuverGrid workload is characteristic of a traditional batch workload model, in which jobs arrived infrequently and are on average long running. There are two ways to use this data: 1) a sample can be taken over a fixed period of time to simulate a workload characterized by long duration batch processing 2) a synthetic high-performance workload can be generated to reflect a modern fine grained dynamic workload by increasing the throughput while maintaining the general workload model. The dynamic usage model is designed to more accurately represent interactive usage of distributed systems as seen in Software-as-a-Service (SaaS) requests, workflows, and smaller scale adhoc personal use on commercial clouds.

6.1.1 Batch Model

The batch workload is generated from a two day sample of the complete AuverGrid trace. The two days AuverGrid chosen include the busiest day (number of jobs) in the trace. Batch model represents favorable auction conditions as the ratio of auction latency to inter arrival time is large. Job duration is on average over 1hour per job, and jobs use approximately 80percent of a single processor.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol.2, Special Issue 1, March 2014

Proceedings of International Conference On Global Innovations In Computing Technology (ICGICT'14)

Organized by

Department of CSE, JayShriram Group of Institutions, Tirupur, Tamilnadu, India on 6th & 7th March 2014

6.1.2 Dynamic Model

The cloud usage is evolving from a traditional batch model to a more dynamic on demand model. The modern usage is therefore characterized by extensible, short duration, adhoc, and interactive usage. The modern usage presents a potentially worst case scenario for auction performance. The allocation performance without considering the duration of the jobs. The workload is used to represent increasing overall utilization and are based on the utilization limit of the testbed.

6.2 Testbed Configuration

The testbed is configured with 20 virtualized computers and it distributed over 10 machines grid. These are connected by a gigabit ethernet network. The 20 virtualized providers having 512 MB of memory allocated to the hosting container.

6.3 Strategy Evaluation

The strategy evaluation is used to evaluate with respect to the number of auctions completed, contracts created and overall system utilization. These strategies are described: Overbooking (O), Second chance substitutes (S) and Advanced reservations (R). In addition, Guaranteed (G) strategy also implemented to compare other strategies. The main difference between these strategy combinations is related to the options.

- G: The providers bid based upon the expected utilization, the bid is never beyond their allocated capacity. If the bids are guarantee, the providers cannot reject the final contracts and no opportunities to use the second chance substitute providers.
- O: The providers bid based on the actual utilization. Providers can bid beyond capacity they may select to accept or reject contracts based on the capacity at the time of contract creation.
- S+O: The provider bid is similar for the above strategy. The losing providers may be substituted with a second chance substitute providers at the time of contract creation.
- R+O: The providers bid based on projected utilization at the time of execution. Here, no second chance substitutes are considered in this situation.
- R+S+O: The providers bid based on projected utilization at the time of execution. The contract cannot be satisfy reservation window. The losing providers may be substituted with a second chance provider.

6.4 Economic Analysis

Economic analysis evaluates the implications of the high utilization strategies on provider revenue and also the impact of economic policies on the high utilization strategies. The economic analysis includes the various high utilization strategies such as, overbooking, advanced reservations, second chance substitute providers.

VII. CONCLUSIONS

The economic resource allocation provides a well studied and efficient means of scalable decentralized allocation it has been stereotyped as a low performance solution due to the resource commitment overhead and latency in the allocation process. These high usage techniques proposed in this paper are designed to reduce the impact of these factors to increase occupancy and improve system utilization. The high usage strategies have each been implemented in the DRIVE metascheduler and evaluated using a series of batch and interactive workloads. The increase in allocation rate was shown to be up to 286 percent for the dynamic workloads and 109 percent for the batch model. In future using regeneration theory and click based event to increase performance, reduce overheads and quickly sharing the resources from cloud server.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol.2, Special Issue 1, March 2014

Proceedings of International Conference On Global Innovations In Computing Technology (ICGICT'14)

Organized by

Department of CSE, JayShriram Group of Institutions, Tirupur, Tamilnadu, India on 6th & 7th March 2014

REFERENCES

- [1] K. Chard and K. Bubendorfer, "Using Secure Auctions to Build A Distributed Meta-Scheduler for the Grid," Market Oriented Grid and Utility Computing, series Wiley Series on Parallel and Distributed Computing, R. Buyya and K. Bubendorfer, eds., pp. 569-588, Wiley, 2009.
- [2] K. Chard, K. Bubendorfer, and P. Komisarczuk, "High Occupancy Resource Allocation for Grid and Cloud Systems, a Study With Drive," Proc. 19th ACM Int'l Symp. High Performance Distributed Computing (HPDC '10), pp. 73-84, 2010.
- [3] R. Buyya, D. Abramson, and J. Giddy, "Nimrod/g: An Architecture for a Resource Management and Scheduling System in a Global Computational Grid," Proc. Fourth Int'l Conf. High Performance Computing in Asia-Pacific Region (HPC Asia '00), pp. 283-289, 2000.
- [4] D. Neumann, J. Stoßer, A. Anandasivam, and N. Borissov, "Sorma - Building an Open Grid Market for Grid Resource Allocation," Proc. Fourth Int'l Workshop Grid Economics and Business Models (GECON '07), pp. 194-200, 2007.
- [5] R. Buyya, R. Ranjan, and R.N. Calheiros, "Intercloud: Utility- Oriented Federation of Cloud Computing Environments for Scaling of Application Services," Proc. 10th Int'l Conf. Algorithms and Architectures for Parallel Processing, p. 20, 2010.
- [6] G. Birkenheuer, A. Brinkmann, and H. Karl, "The Gain of Overbooking," Proc. 14th Int'l Workshop Job Scheduling Strategies for Parallel Proc. (JSSPP), pp. 80-100, 2009.
- [7] M.A. Netto, K. Bubendorfer, and R. Buyya, "Sla-Based Advance Reservations with Flexible and Adaptive Time Qos Parameters," Proc. Fifth Int'l Conf. Service-Oriented Computing (ICSOC '07), pp. 119-131, 2007.
- [8] A.E. Roth and A. Ockenfels, "Last-Minute Bidding And the Rules for Ending Second-Price Auctions: Evidence from Ebay and Amazon Auctions on the Internet," Am. Economic Rev., vol. 92, no. 4, pp. 1093-1103, 2002.
- [9] P. Bajari and A. Hortacsu, "Economic Insights from Internet Auctions," J. Economic Literature, vol. 42, pp. 457-486, 2004.
- [10] K. Bubendorfer, "Fine Grained Resource Reservation in Open Grid Economies," Proc. IEEE Second Int'l Conf. e-Science and Grid Computing (E-SCIENCE '06), p. 81, 2006.
- [11] C. Castillo, G.N. Rouskas, and K. Harfoush, "Efficient Resource Management Using Advance Reservations for Heterogeneous Grids," Proc. IEEE 22nd Int'l Symp. Parallel and Distributed Proc.(IPDPS '08), pp. 1-12, Apr. 2008.
- [12] K. Chard and K. Bubendorfer, "A Distributed Economic Meta-Scheduler for the Grid," Proc. IEEE Eighth Int'l Symp. Cluster Computing and the Grid (CCGRID '08), pp. 542-547, 2008.

ABOUT THE AUTHORS



MS.K.Ramya received the B.Tech (Information Technology) from Karpaga Vinayaga College of Engineering and Technology in 2012 and pursuing M.Tech (Information Technology) in VSB Engineering College, Anna University, Chennai. Areas of interest includes Distributed Systems and Cloud Computing.



Mr.P.G.Kathiravan received M.Tech (Information Technology) from K.S.Rangasamy College of Technology (Autonomous) in 2012 Anna University, Chennai and B.Tech (Information Technology) from The Kavery Engineering College in 2010, Anna University, Chennai. He worked as Lecturer in the Department of Computer Science and Engineering & Information Technology, Government College of Technology, Coimbatore in the year (2012-13).He currently working as Assistant Professor in the Department of Information Technology in V.S.B Engineering College, Karur .He published and presented various International & National papers and attended various International & National Workshops. His area of interest includes Computer Networks, Cloud Computing, and Grid Computing.