Scalability Performance of the IDMH Heuristic in IaaS Clouds

L Priyanka[1] S Sireesha[2] S Palani[3]
[1,2]Student [3]Assistant Professor
[1,2,3]Department of Computer Applications
[1,2,3]Sri Venkateswara College of Engineering and Technology, Chittoor, India

Abstract— Cloud Computing, it's a dream for the IT Industry, where it is used to store large amount of data in cloud and it is easily retrieved over the internet without the interaction of the user. Cloud Computing is a model for convenient, enabling ubiquitous, and for the on demand sharing pool of resources it may be networks, servers, storage, and applications, it also provides service interaction. As we know that SaaS, PaaS, IaaS are the models of the cloud. Usually a cloud can be managed by a single organization or specific community or general public or it may be a composition of two or more cloud infrastructure. Without requiring the human interaction with each service provider instead of this it automatically has computing capabilities for a consumer such as server time and network storage. In cloud deployment option stakeholders in business and academia are increasingly exploring for their critical applications. Cloud Computing Computational resources on demand in response to workload changes and also make it possible to flexibly procure and scale.

Keywords: Cloud, A*Bases Search, Direct move Heuristic, Searching, Interactive Direct move Heuristic

I. Introduction

Cloud computing is one of the latest technology which are used to store data and it mainly used by the IT Industry, the advantage is we can easily retrieve the data over the internet without the interaction of user. The following are the essential characteristics of the cloud

On Demand Self Service: It explains that without the interaction of human, a consumer can unilaterally provision computing capabilities, such as server time and network storage, as needed automatically, with each service provider.

Broad Network Access: Here we understood that in cloud we have the capabilities that are available over the internet and accessed through standard mechanisms that promote use by mobile phones, tablets, laptops, and workstations.

Resource Pooling: In Cloud we can assign and reassign according to the consumer demand where resources are pooled to serve multiple consumers using a multi-tenant model, with different virtual and physical resources dynamically.

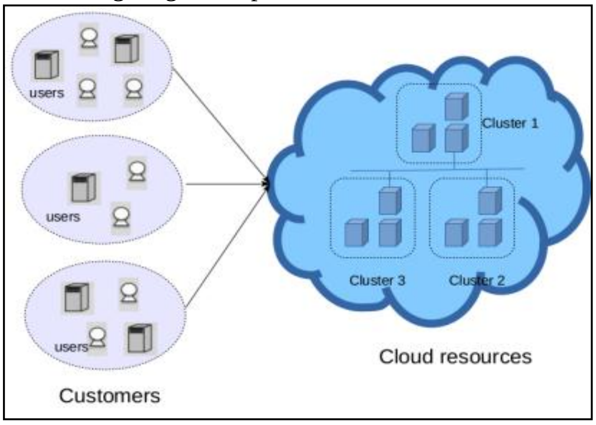
Additionally we have another two characteristics which are essential that are Rapid Elasticity and Measured Service.

A. Models:

In cloud computing models are differentiated into two categories:

1) Service models: It describes about SaaS, Paas, Iaas
2) Deployment models: It describes about the Private Cloud, Public Cloud, Hybrid Cloud and Community Cloud.

The following diagram explains the architecture of cloud



Customers' VMs are not tied to any specific physical machine (PM) in the IaaS cloud platform. PMs of the cloud are grouped into clusters to provide High Performance Computing and/or High Throughput Computing. In one hand, customers pay for cluster computing service for business continuity and service availability reason. On the other hand, service providers offer cluster computing solutions to increase their reputation indicators.

II. Literature Survey

Consolidating multiple underutilized servers into a fewer number of non-dedicated servers can respond to the common problem of server sprawl in data centres. The consolidation can be modelled as a variant of the bin packing problem (items to be packed are the VMs and bins are the PMs) in order to help resource management considering that until 2008, server consolidation exercise was primarily a manual process. A manual process of this nature is time consuming and depends on the subjective assessment of the decision maker. Particularly, in it is studied the server consolidation problem considering technically deprecated PMs into technically superior PMs, but this problem could also be studied as a VMP problem. Experimental tests are performed considering a maximum number of 1000 PMs for mapping into 620 PMs, solving extremely large instances of the problem in a reasonable amount of time, i.e. 0.188 seconds against 6262.1 seconds (1.74 hours) for the optimal solution. The proposed algorithm obtained solutions with a 1.5 ratio more PMs than the optimal solution for largest problem instances, i.e. 620 PMs obtained against 411 PMs from optimal solution.

If a VM is considered to be a three dimensional object (CPU, memory and storage) then the problem of placing the VMs over the PMs looks similar to the three dimensional bin packing problem, but they are not exactly the same. In 3D bin packing problem, a set of three dimensional objects (generally cuboids) are required to be placed inside three dimensional containers (also cuboids). The aim is to pack as many objects in the containers as possible, so that the number of containers required is minimized. While packing objects into a given container, two objects can be placed side by side or one on top of the other, but if we consider VMs as the objects then placing VMs side by side or one on top of the other is not a valid operation. This is because once a resource is utilized or occupied by a VM, it cannot be reused by any other VM. The VMP problem is actually similar to Vector Packing Problem which is also a NP-Hard problem.

VMs and PMs are represented as a d-dimensional vector where each dimension corresponds to one type of resource (e.g. CPU, memory and storage) and for each PM should be balanced along different dimensions. Operational power and thermal dissipation are also studied. For each objective an efficiency metric was defined. For the resolution of the VMP formulation, authors proposed a Genetic Algorithm (GA). Genetic operators are: selection and crossover, where a random process do not perform efficiently, so the selection and crossover is based on each of the three efficiency metrics resulting on better solutions, mutation, where groups of VMs are randomly selected and eliminated.

In are studied two payment plans in cloud computing markets: reservation and on-demand plans (Amazon EC2 and Go Grid). Considering that prices of resources in reservation plans are generally cheaper than in on-demand plans, in advance reservation should be applied as maximum as possible. On-demand plans only should be used when the reservation does not fully meet predicted workload. The proposed algorithm makes a decision to host a certain number of VMs on appropriate cloud providers obtaining an optimal solution from a stochastic integer programming (SIP) formulation considering two stages of decision making: first stage defines the number of VMs provisioned in reservation phase, while the second stage defines the number of VMs allocated in both utilization and on-demand phases.

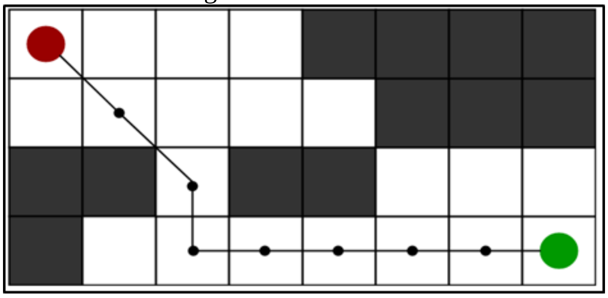
In, it is defined a new high-availability property for a VM; when a VM is marked as k-resilient, it should be guaranteed that it can be relocated to a non-failed host without relocating other VMs as long as there are up to kPMs failures. Anti-location, anti-co-location and resources constraints are also proposed for the VMP formulation, considering the optimization of the problem in its dual space (dual optimization problem). The transformed formulation enable a CP engine to compute a solution in reasonable time on practical cases.

When a client request comes, the infrastructure service provider needs to identify a PM to place the client VM so that the following objectives can be achieved: the SLA associated with the VM is satisfied, and the performance metrics of SLAs associated with the VMs previously running on the target physical node is least affected. The proposed prediction contains the following initiation steps: collect the performance data of a given application during its previous executions and treat it as foreground performance metrics, collect the performance data of candidate physical nodes and treat the data as background performance metrics, and perform Canonical Correlation Analysis (CCA) on the performance data collected in the previous two steps and produce the canonical weight vectors and scores for the most correlated patterns.

III. PROPOSED ALGORITHMS

A. *A*Based Search Algorithm:*

A* algorithm is used in finding the paths and graph traversals, it also considered as one of the most popular technique. It is used to approximate the shortest path in real life like in games, maps mostly in the place where there are many hindrances. **A*** algorithm are unique when compared to the other conventional algorithms.



1) *Algorithm:*

1) procedure EXPAND()(Input: a state S Output: a list of states Open)
2) for each cluster i = 1..m do
3) for each customer j = 1..n, so that j / ∈ i do
4) Generate a new state S0 from S so that j ∈ i
5) if S0 was not expanded AND constraints are respected then
6) Append S0 to Open
7) end if
8) end for
9) end for
10) end procedure

B. *Direct Move Heuristic:*

1) *Algorithm:*

1) procedure DMH()(Input: S,Sg Output: a set of moves stored in P)
2) hcmax ←the maximum number of heavy customers per cluster from Sg
3) for j =1 to n do
4) x←find cluster(j,S);
5) if S[x][j]
   6= Sg[x][j] then
6) i←find cluster(j,Sg);
7) if the assignment of j to i in S is valid then
8) move j from x to i in S.
9) store this move in P
10) end if
11) end if
12) end for return P
13) end procedure

IV. RESULT AND ANALYSIS

The following describes this journal in the form of screenshots

Fig. 3: Homepage

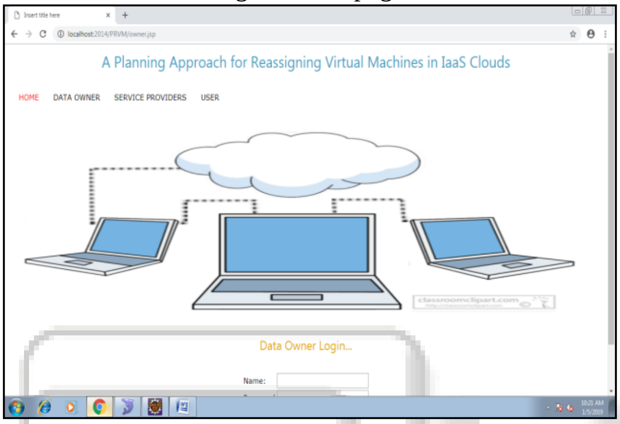


Fig. 4: Data Owner Login

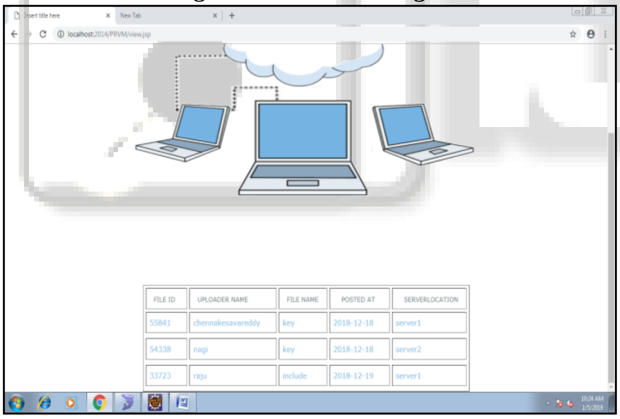


Fig. 5: View Files

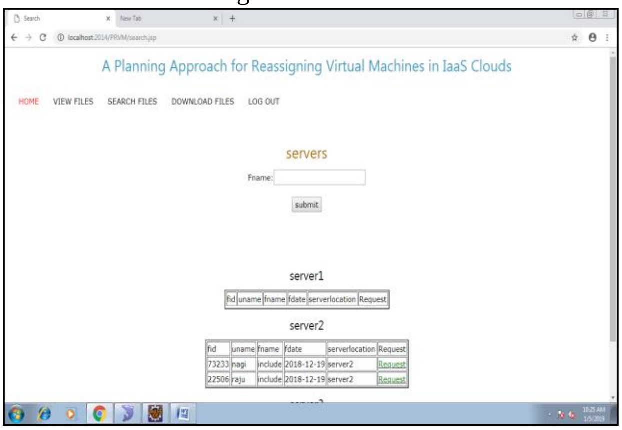


Fig. 6: Search Files

Fig. 7: Download Files

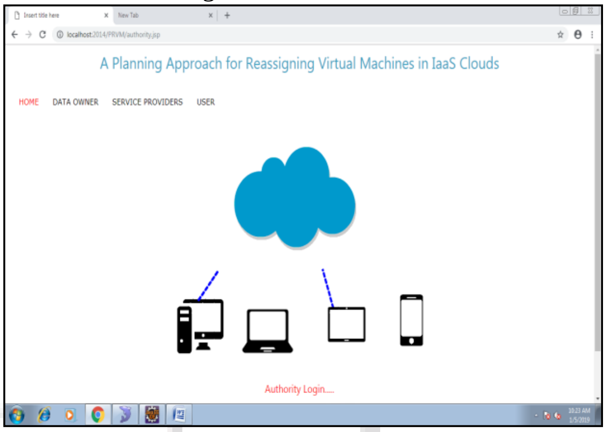


Fig. 8: Search Provider Home Page

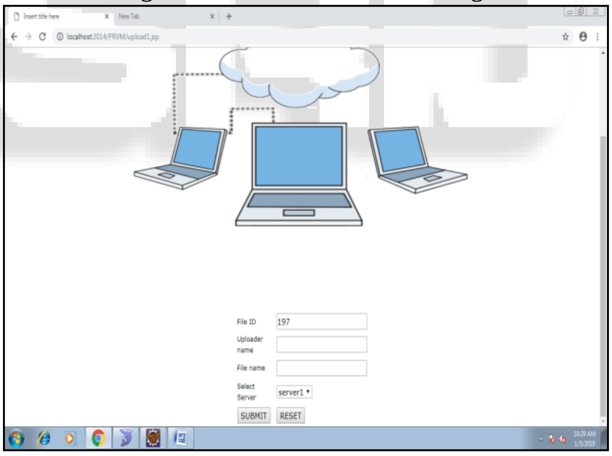


Fig. 9: Search Provider Upload Files

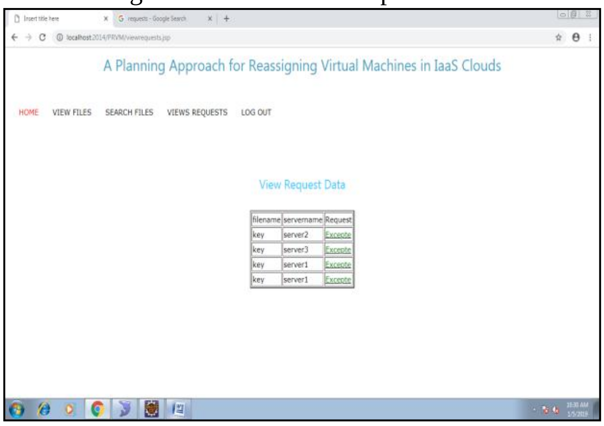


Fig. 10: View Request

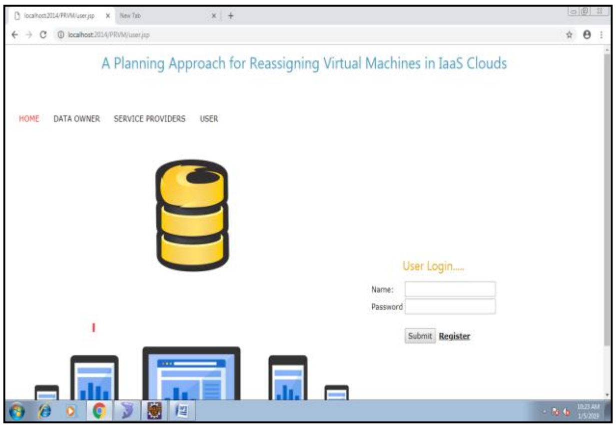


Fig. 11: User Page

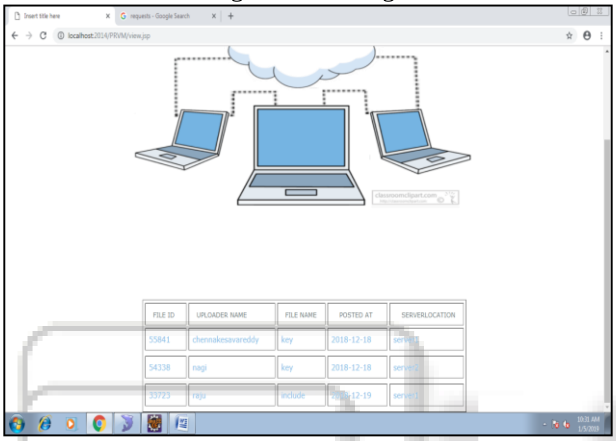


Fig. 12: User View Files

V. Conclusion

In this paper, we proposed a planning approach to the problem of reassigning virtual machines in IaaS cloud platforms. The planning problem, which we addressed in this paper, could be a new real-life problem from IT sector to assess AI algorithms instead of using standard benchmarks from game theory. We have used the well-known A* planning algorithm to solve this problem. However, due to its space limitation, we have proposed two heuristics, called DMH and IDMH. The experimental study shows that IDMH has a good performance in terms of success rate and solution quality. In addition, we have found that the presence of additional resources allow the scalability of the IDMH to large sized problem instances with up to 800 VMs. Our future research work will include the integration of the proposed model/algorithms into an IaaS cloud platform to automate the reassignment task. Furthermore, we will focus on the investigation of the proposed planning approach using other algorithms for further optimization of the objective function.

References

[1] Michael Armbrustetal.: A view of cloud computing, Communications of the ACM Vol.53, No. 4, pp. 50–58, 2010.
[2] Mell, P. and Grance, T.: The NIST definition of cloud computing, 2011.
[3] Irena Bojanova, Jia Zhang, and Jeffrey M. Voas: Cloud Computing, IT Professional, Vol. 15, No. 2, pp. 12–14, 2013.
[4] Funmilade Faniyi and Rami Bahsoon: A Systematic Review of Service Level Management in the Cloud , ACM Computing Surveys Vol.48, No. 3, pp. 1–27, 2016.
[5] Amelie Chi Zhou, Bingsheng He, and Cheng Liu: Monetary Cost Optimizations for Hosting Workflow-as-a-Service in IaaS Clouds, IEEE Transactions on Cloud Computing, Vol. 4, No. 1, pp. 34-48, 2016.

