International Journal of Research (IJR)
e-ISSN: 2348-6848, p- ISSN: 2348-795X Volume 2, Issue 12, December 2015
Available at http://internationaljournalofresearch.org

A Guideline for Virtual Machine Migration Algorithm in Cloud Computing Environment

Vijaya Raju Motru[1]; Dr. Penmetsa V Krishna Raja[2]; Ashok Kote[3]; Gudikandhula Narasimha Rao[4] & Rajesh Duvvuru[5]

[1] Dept. of Information Technology, UshaRama College of Engineering & Technology, Vijayawada
[2,3] Dept. of Computer Science & Engineering ,
[4]Dept. of Geo-Engineering & Centre for Remote Sensing
[2] Sri Vatsavai Krishnam Raju College of Engineering& Technology, Bhimavaram
[3] Nova College of Engineering & Technology, Vijayawada
[4,5] Andhra University College of Engineering (A), Visakhapatnam, Andhra Pradesh, India.
Email: { vijayaraju.m, drpvkraja, kote.ashok } @ gmail.com[1,2,3], { narasimha.geo, drajesh.sch } @cea.auvsp.edu.in[4,5]

Abstract— *Live migration of virtual machines (VMs) is widely used for managing cloud computing platforms. However, live migration causes performance interference on cloud services running on migrated VMs or other VMs collocating with the services during or after migration. Consolidation of multiple applications on a single Physical Machine (PM) within a cloud data centre can increase utilization, minimize energy consumption, and reduce operational costs. However, these benefits come at the cost of increasing the complexity of the scheduling problem. In this paper, we present a topology-aware resource management framework. As part of this framework, we introduce a PlaceMent scheduler that provides and maintains durable allocations with low maintenance costs for data centres with dynamic workloads. We focus on workloads featuring both short-lived batch jobs and latency-sensitive services such as interactive web applications. The scheduler assigns resources to Virtual Machines and maintains packing efficiency while taking into account migration costs. According to our experimental results, we reveal the trade-offs of each migration policy and implementation that are not just related to downtime and migration time and present guidelines for selecting appropriate policies and implementations.*

Index Terms- Cloud Computing; Peer to peer;Virtual Machine consolidation; Physical Machine; Resource management

1. INTRODUCTION

To use live migration for dynamic replacement of VMs, cloud administrators must take interference of live migration performance in to consideration [1]. Since migration consumes computing resources, live migration significantly interferes with of the performance of migrating and collocating VMs [2]. In this paper, we call this performance interference on migrating VMs migration interference, and that on collocating VMs collocation interference. Cloud providers offer an infrastructure to be shared by multiple applications, which is usually expensive and needs to be wisely utilized [3], [4]. Utilization can be improved by running an appropriate mix of application workloads on each individual machine, which is known as consolidation. While consolidation can be used to increase utilization, it also increases the complexity of the scheduling problem [5]. Security is important issue in networks as well as cloud computing [6], because cloud live migration is there in big data systems. So we need to secure cloud policies and cloud mitigations. As a part of this one data maintenance is important issue [7].

A degree of sub-optimal application placement is inevitable due to load changes and the fact that it would be impractically expensive to completely remap every component of every running application across all of the available servers each time a load change occurred [8]. Consequently, there is a need for a scheduler that can respond rapidly to changes in demand, producing efficient and durable packing in a way that accounts for the heterogeneity of the cloud's workloads, imposes low costs of maintaining the packing efficiency, and can scale up to tens of thousands of servers per data centre. After completion of cloud migration then we have to classify the data by using clustering techniques [9]. Every cloud should be maintained by each customer information with their family details also (Healthy customers and unhealthy customers) [10].

We propose a new P2P consolidation framework. Some of this framework's basic functionality has

International Journal of Research (IJR)
e-ISSN: 2348-6848, p- ISSN: 2348-795X Volume 2, Issue 12, December 2015
Available at http://internationaljournalofresearch.org

previously been verified in prototype form. The proposed framework is a general computational model for cooperatively optimizing a global system objective through local interactions and computations in a multi-agent system over a semi-random connectivity. We also introduce a scheduling heuristic designed to provide and maintain durable packing with low maintenance costs for a data centre with a dynamic workload. A scheduler based on this heuristic is shown to achieve such durable packing in a way that avoids costly reconfigurations, and to offer cheap migration plans to maintain packing efficiency [11]. The main contributions of this paper are:

• A formulation of the VM consolidation problem as a distributed optimization problem.
• A topology-aware resource management framework for VM consolidation.
• A heuristic algorithm for VM consolidation that factors in the risks of resource contention, packing efficiency, migration costs, and migration locality to produce durable consolidations and offer cheap migration plans to maintain packing efficiency and reduce resource stranding.
• An in-depth simulation-based evaluation of the system behaviour under different settings and configurations.

2. Requirements for Proposed System

When scheduling VMs to run different services or batch jobs, the scheduler must meet several requirements and it faces a number of challenges in meeting them. A summary of these challenges are:

2.1. Resource contention caused by consolidation: Co-locating different applications can cause performance variability or degradation due to resource contention when resources are being shared [12]. The scheduler should therefore identify complementary workloads and place them together to improve packing efficiency and minimize resource contention.

2.2 Job heterogeneity: A data centre will be required to run different types of applications. In broad terms, two classes of application can be distinguished: long-running interactive services and batch jobs, which perform a specific computation and then finish. Batch jobs that are run in cloud data centres are usually shorter and less latency-sensitive than interactive services, involve constant resource utilization, and do not usually require careful scheduling.

2.3. Migration cost: VM migration is a widely used technique for achieving consolidation once the decision on which jobs to consolidate has been made. However, migrations are often costly. Particularly important costs to consider include the cost of double resource utilization during the migration, the costs of SLA violations caused by migration downtime, the cost of network traffic, and the potential network contention issues that may arise during the migration [13].

2.4. Topological constraints: The scheduler should consider the network topology to avoid high migration costs due to network traffic, contentions, or redundant configurations. Most existing works on scheduling treat the data centre as an unstructured pool of resources, but real data centres Virtual LANs (VLANs), Access Control Lists (ACLs), broadcast domains, and load balancers that impose constraints and create barriers that reduce the scope for agility in migration.

2.5. Risk of load change and contention: The scheduler should factor the risk of change and contention into its decision function so as to avoid frequent migrations and produce durable decisions.

2.6. Computation time: The scheduler should produce a solution with in an acceptable time-frame, and before the solution becomes disparaged due to load changes.

3. Cloud Migration Architecture

Cloud migration consists different procedure for solving the problems of customers. Basically live migration is one of the important issue in cloud computing.

3.1 Live Migration

One of the extremely powerful tools of virtualization is VM migration. We can categorize the migration capability into two major types: the one is called as non-live (offline) migration and the other is called as live (online) migration. In non-live migration process, the status of virtual machine is suspended and users face service interruption during migration. On the other hand, the live migration process will keep the running states of virtual machine and it will not lose the status of VMs during migration process. Our discussion will mainly focus on live migration instead of non-live migration because it is out of scope of our research [14].

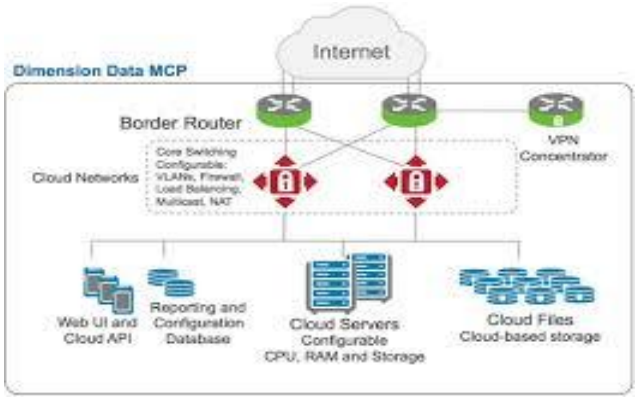

Figure.1. Cloud Data Centre Architecture.

International Journal of Research
International Journal of Research (IJR)
e-ISSN: 2348-6848, p- ISSN: 2348-795X Volume 2, Issue 12, December 2015
Available at http://internationaljournalofresearch.org

At present, most popular hypervisors (Xen, KVM, VMWare, etc) are using the pre copy approach with different ways, but the basic concept of implementation in all is same. In pre copy, first iteratively copies the entire contents of memory from source VM to destination, and then switches the execution states at targeted host. During the process of copying memory contents, the VM remains responsive and progressive at source [15]. The main advantage of using pre copy is the reliability and robustness, because the migration process can be reverted at source host if migration fails [16].
The migration process of pre copy approach is shown in Figure 1. In this figure we can see that the VM from preparation phase to pre copy round N is a live at source. The total number of pre copy rounds are based on the modified/dirty page rate of memory at source host. During pre copy rounds already transmitted memory pages might require retransmission because same memory pages might be modified/dirty after having been transferred to destination. After completion of round N, the VM will stop the execution and copy the remaining dirty pages and VM`s processor states and then VM resumes and restarts at destination.

Cloud Migration Process:

- Awareness is positively correlated with the intent to adopt cloud computing.
- Cost effectiveness is positively correlated with the intent to adopt and use cloud computing.
- . Risk is negatively correlated with the intent to adopt and use cloud computing.
- Data security is negatively correlated with the intent to adopt and use cloud computing.
- Availability of good ICT infrastructure is positively correlated with the intent to adopt and use cloud computing.
- Relative advantage is positively correlated with the intent to adopt cloud computing.
- Compatibility is positively correlated with the intent to adopt cloud computing.
- Complexity is negatively correlated with the intent to adopt cloud computing.
- Observability is positively correlated with the intent to adopt cloud computing.
- Trial ability is positively correlated with the intent to adopt cloud computing.
- Results demonstrable is positively correlated with the intent to adopt cloud computing.
- Ease of use is positively correlated with the intent to adopt and use cloud computing.
- Usefulness is positively correlated with the intent to adopt and use cloud computing.
- Socio-cultural factors are negatively correlated with the intent to adopt cloud computing.
- Age of the university will moderate the intent to adopt cloud computing

4. RESULTS AND DISCUSSION

The evaluation of the proposed P2P management framework and the associated heuristics was performed by simulating a large scale data centre using Peer Sim. A variable load was applied to the data centre due to VM churn and changing VM demand. The simulated data centre consists of 65 536 PMs interconnected in a multi-rooted tree topology, with 1 core router at the core level, 2 clusters at the aggregator level, 32 groups at the third level, and 1024 physical machines in each group [17].

A. Migration interference

Live migration interferes with the migrating VM's workload during migration. The Migrating VM should stop its workload during downtime, which results in serious performance degradation and SLA violations. In addition to the downtime, throughput degradation of the migrating VM is also not negligible when the VM is not in downtime.

B. Collocation interference

As with migration interference, the migration process also deprives computational resources from collocating VMs. One point to be noted is that live migration degrades a collocating VM's throughput even when the migrating VM is not running on the host.

C. Interference Metrics

To quantitatively analyze migration noise, we use the following metrics.

• **Original metrics**: We use downtime and migration time as the original metrics, which are often used to evaluate live migration. Downtime is the length of stop-and-copy and migration time is the term from when the migration process starts its execution until completion.

• **Regular migration interference**: The merge value of the interference occurring while the migrating VM and collocating VMs are collocating on the same host. This metrics indicates the usual performance interference caused during migration.

• **Total migration interference**: The integration of interference appears during migration. This metric indicates the total amount of interference considering the original metrics and our metrics of migration interference.

• **Relocation interference:** Migration interference appears on a host even where migrating VM is not running on. We discuss the interference caused while the migrating VM is running on another host during and after migration.

• Performance parameters

The aim of the evaluation was to answer the following questions: How well are the applications packed? How does each of the core function variables (risk, efficiency, migration locality, migration cost, and imbalance) affect

International Journal of Research (IJR)
e-ISSN: 2348-6848, p- ISSN: 2348-795X Volume 2, Issue 12, December 2015
Available at http://internationaljournalofresearch.org

the stability and reliability of consolidation. It deals how effectively the scheduler minimizes the migration costs. The following are the measurements for performance.

1. Firstly find out average resource utilization.
2. Count number of active servers at time t in a net work.
3. Find out average number of active servers. Average active servers (t)= $\sum_{j=1}^{t} n_{active}(t) / t$.
4. Calculate total number of pending requests: The number of VM requests that are not scheduled within the assigned time frame.
5. Average imbalance rate: This parameter is a representation of resource stranding in the data centre.
6. Generate total number of data transfer in Giga Bytes.
7. Batch job data transfer (GB): The volume of data transferred due to the cold migration of the batch jobs.
8. Service data transfer (GB): The volume of data transferred due to the live migration of the services.
9. Under load data transfer (GB): The volume of data transferred to free up the resources in an under loaded PM, either to open up space for larger jobs or to be put into a power saving mode.
10. Overload data transfer (GB): The volume of data transferred to resolve an overload event.
11. Identify total number of migrations.
12. Identify number of batch job migrations.
13. Observe number of service migrations.
14. Notice overload triggers, under load triggers and un resolved triggers.
15. Stop the process.

We also evaluated the system's performance in cases featuring different ratios of batch jobs to services. The greater the proportion of batch jobs, the greater the volatility of the workload and thus the greater the need for frequent scheduling.

4.1. Efficiency

The performance metrics impacted by changing the weighting of the efficiency factor. Considering efficiency of placement while selecting servers increased the average utilization and reduced the active number of servers required to serve a specific load [18]. Interestingly, the main impact of introducing efficiency was not due to improvements in utilization but to a reduction in the number of under load triggers, which reduced the volume of data transfer required to resolve the under load state in figure 2. Efficient packing also reduced the number of unresolved triggers by up to 70%.

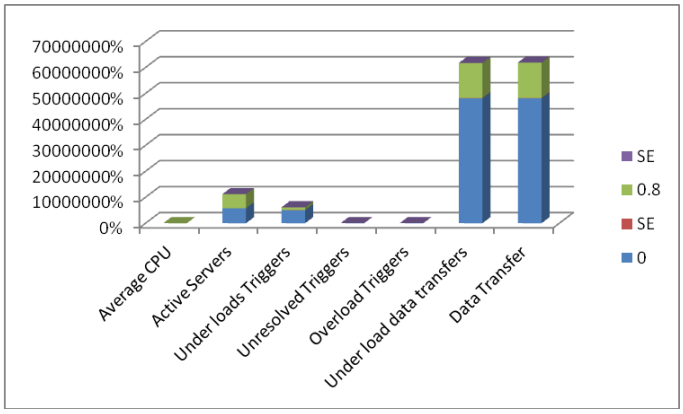

Figure.2. Impact of the efficiency factor on packing efficiency.

4.2. Risk

We evaluated the impact of considering the risk of load variability and resource contention on possible future overloads, migrations and data transfer. Figure.3 shows the numbers of overload triggers and overload data transfers as well as the average number of active servers during the simulation runtime for different risk weightings[19]. Accounting for risk in the score function reduced the number of overload triggers by up to 35% and the volume of data transfer due to offload by up to 33%. However, this came at the cost of a 2% increase in the average number of active servers.

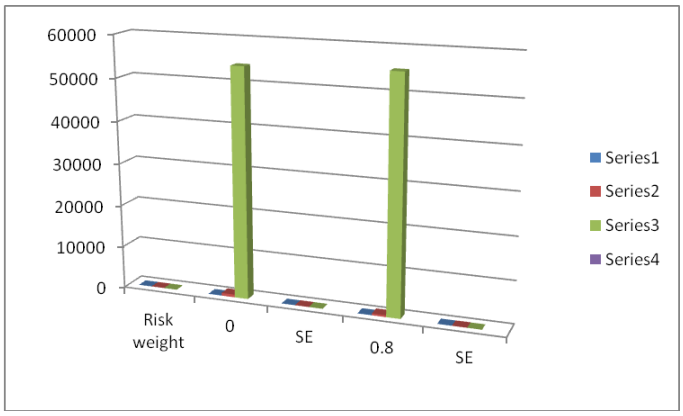

Figure.3. Impact of the risk factor on packing efficiency.

4.3. Imbalance

We examined the consequences of resource stranding and the impact of accounting for the imbalance factor as a way of mitigating the negative effects of stranding. Take the imbalance factor into account improved resource utilization and reduced the number of pending requests. Moreover figure.4 shows placing

International Journal of Research (IJR)
e-ISSN: 2348-6848, p- ISSN: 2348-795X Volume 2, Issue 12, December 2015
Available at http://internationaljournalofresearch.org

complementary workloads together reduced the probability of overload and thus the number of overload triggers [20].

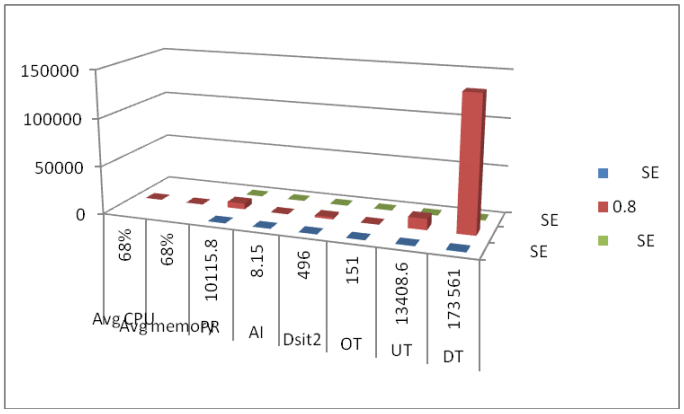

Figure.4. Impact of the imbalance factor on packing efficiency

4.4. Migration Cost

Figure.5 describes the low impact of the migration cost variable is due to the fact that the overall migration cost is mainly dependent on the number of migrations that are performed rather than the volume of data transferred during a single migration. The low impact of the migration cost variable is due to the fact that the overall migration cost is mainly dependent on the number of migrations that are performed rather than the volume of data transferred during a single migration.

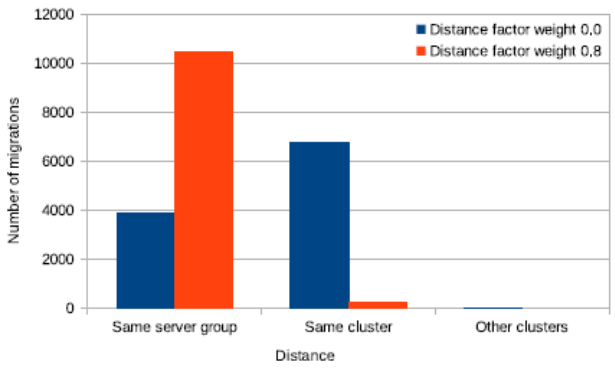

Figure.5. Numbers of migrations performed within groups, between groups, and between clusters with different distance factor weightings.

4.5. Locality Factor

It describes increasing the weighting of the migration locality factor increases the proportion of local migrations within a server group relative to those of migrations between groups or clusters. However figure.6 describes slight increment of the migration cost in terms of data transfer and also the number of sub-optimal state triggers due to the associated trade-off with risk and migration cost.

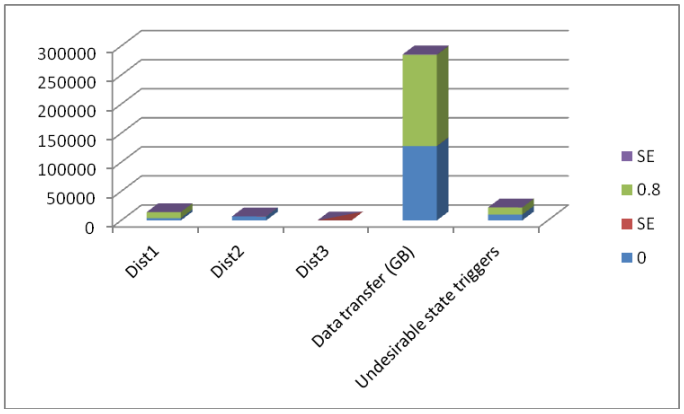

Figure.6. Impact of the locality factor on packing efficiency.

4. CONCLUSION

This research evaluates a decentralized resource management framework for large-scale cloud infrastructures. The P2P structure of the framework provides parallelization, a high degree of concurrency and provides reasonable scalability as the number of PMs and VMs increases. Moreover, the P2P architecture's random overlay allows the system to create a logical dynamic connectivity among a large pool of resources and reduce the negative impacts of static partitioning, which lead to low utilization. It also accounts for the physical proximity of the neighbours when building the logical overlay, thereby reducing the costs of network transit and reconfiguration. We discussed the interference of live migration caused by various migration policies and implementations. However, when throughput-critical services are running on a VM, all migration methods can be candidates depending on the requirements on migrating VMs.

5. REFERENCES

[1] Ostermann, Simon, Alexandria Iosup, Nezih Yigitbasi, Radu Prodan, Thomas Fahringer, and Dick Epema, "A performance analysis of EC2 cloud computing services for scientific computing," in Cloud Computing, pp. 115-131. Springer Berlin Heidelberg, 2010.

[2] Akoush, Sherif, Ripduman Sohan, Andrew Rice, Andrew W. Moore, and Andy Hopper, "Predicting the performance of virtual machine migration," in Modeling, Analysis & Simulation of Computer and Telecommunication Systems (MASCOTS), 2010 IEEE International Symposium on, pp. 37-46. IEEE, 2010.

[3] Svärd, Petter, B. Hudzia, S. Walsh, Johan Tordsson, and Erik Elmroth, "The Noble Art of Live VM Migration- Principles and performance of pre copy, post copy and hybrid migration of demanding workloads," in Technical report, 2014. Tech Report UMINF 14.10.

[4] Clark, Christopher, Keir Fraser, Steven Hand, Jacob Gorm Hansen, Eric Jul, Christian Limpach, Ian Pratt, and Andrew Warfield, "Live migration of virtual machines," in Proceedings of the 2nd conference on Symposium on Networked Systems Design & Implementation-Volume 2, pp. 273-286. USENIX Association, 2005.

[5] Hirofuchi, Takahiro, Hidemoto Nakada, Satoshi Itoh, and Satoshi Sekiguchi, "Reactive consolidation of virtual machines enabled by postcopy live migration" in Proceedings of the 5th international workshop on Virtualization technologies in distributed computing, pp. 11-18. ACM, 2011.

[6] B. Balaji Bhanu , Dr. P. Srinivasulu, Gudikandhula N Rao, "Secure Group Key Communication in Sensor Networks" In International Journal of Advanced Computer Engineering and Architecture, Vol. 2 No. 1 (January- June,2012) ISSN: 2248-9452.

[7] Rajesh Duvvuru, Sunil Kumar Singh, G. Narasimha Rao, Ashok Kote, B.Bala Krishna and M.Vijaya Raju . "Scheme for Assigning Security Automatically for RealTime Wireless Nodes via ARSA". Title: Heterogeneous Networking for Quality, Reliability, Security and Robust ness, QSHINE 2013, SPRINGER, LNICST 115, pp. 185-196, 2013. © Institute for Computer Sciences, Social Informatics and Telecommunications Engineering 2013. ISBN- 978-3-642-37948-2.

[8] C. Clark, K. Fraser, S. Hand, J. G. Hansen, E. Jul, C. Limpach, I. Pratt, and A. Warfield, "Live migration of virtual machines," in Proceedings of the 2nd Conference on Symposium on Networked Systems Design and Implementation (NSDI '05), 2005, pp. 273–286.

[9] G. Narasimha Rao, R. Ramesh, D. Rajesh, D. Chandra sekhar."An Automated Advanced Clustering Algorithm For Text Classification". In International Journal of Computer Science and Technology, vol 3,issue 2-4, June, 2012, eISSN : 0976 - 8491,pISSN : 2229 – 4333.

[10] Rao, Gudikandhula Narasimha, and P. Jagdeeswar Rao. "A Clustering Analysis for Heart Failure Alert System Using RFID and GPS." ICT and Critical Infrastructure: Proceedings of the 48th Annual Convention of Computer Society of India-Vol I. Springer International Publishing, 2014 [11] R. Angles, C. Gutierrez, Survey of graph database models, ACM Comput. Surv. 40 (2008) 1–39.

[12] H. Nguyen, Z. Shen, X. Gu, S. Subbiah, and J. Wilkes, "Agile: Elastic distributed resource scaling for infrastructure-as-a-service," in Proceedings of the 10th International Conference on Autonomic Computing (ICAC '13), 2013, pp. 69–82.

[13] A. Kivity, Y. Kamay, D. Laor, U. Lublin, and A. Liguori, "kvm: the linux virtual machine monitor," in Linux Symposium, 2007.

[14] D. Pham, "Data Clustering Using the Bee Algorithm," Proceedings of the 40th CIRP International Manufacturing Systems Seminar, pp. 233–358, Liverpool: CIRP, 2007.

[15] A. Jain, "Data clustering: a review," ACM computing surveys (CSUR), 31(3), pp. 264–323, 1999.

[16] D. Karaboga, and C. Ozturk, "A novel clustering approach: Artificial Bee Colony (ABC) algorithm," Applied Soft Computing, 11(1), pp. 652–657, 2011.

[17] A. Greenberg, J. Hamilton, D.A. Maltz, P. Patel, The cost of a cloud: research problems in data center networks, ACM SIGCOMM Comput. Commun. Rev. 39 (1) (2008) 68–73.

[18] B. Speitkamp, M. Bichler, A mathematical programming approach for server consolidation problems in virtualized data centers, IEEE Trans. Serv. Comput. 3 (4) (2010) 266–278.

[19] J.L. Berral García, R. Gavaldà Mestre, J. Torres Viñals, et al. An integer linear programming representation for data-center power-aware management, 2010.

[20] P. Svärd, B. Hudzia, S. Walsh, J. Tordsson, E. Elmroth, Principles and performance characteristics of algorithms for live VM migration, ACM SIGOPS Oper. Syst. Rev. 49 (1) (2015) 142–155.

Dr. PENMETSA VAMSI KRISHNA RAJA:

He is presently working as a principal in Sri Vatsavai Krishnam Raju College of Engineering & Technology Bhimavaram. He did his PhD from JNTU, Kakinada. He received his M.Tech (CST) from Andhra University,Visakhapatnam, Andhra Pradesh, India. He has authored more than 20 relevant publications in journals and conferences related to these areas. His research areas include Computer Networks, Network Security, Cloud Computing, Big Data, Data Mining and Software Engineering.

Vijaya Raju Motru received the Bachelor degree in computer science and engineering from university of JNTUH, hyderabad in 2007, and the Master's degree in computer Scinece and engineering from the University of ANU, Guntur, India, in 2010. Currently, he as been working as assistant professor in the department of Information Tecnology. His research areas include, Cloud Computing, Adhoc Networks, Big Data and Network Security.