

# Efficient VM Migration Policy in Cloud Computing Environment



Annie Pathania, Kiranbir Kaur and Prabhsimran Singh

**Abstract** Cloud computing provides resources on shared basis but resources do get exhausted, as more and more resource-dependent tasks are being executed on the cloud. This eventually leads to distortion and one possible solution to overcome this problem is migration. In this paper, we perform VM migration in an energy-efficient manner for which we calculate the load factor on all the individual servers. If the load exceeds the assigned threshold value, then that server is considered as the overloaded host, after this the random selection of the VMs is done from the under-loaded hosts and then the machine with less migration time and more utilization will be migrated to the destination host. Further, we compare our purposed technique with already established techniques. The comparison results in the form of Migration Time, Utilization, and Energy Consumption shows that the proposed technique performs better than the existing one.

**Keywords** Migration · Load balancing · Utilization · Energy consumption · VM sorting

## 1 Introduction

Cloud computing is a research area which deals with on-demand availability of computer resources. Almost all IT industries need flexible cloud computing platform with millions of physical hosts and devices. It is also the computing over Internet with no requirement of the infrastructure at the user's site. The main technology behind cloud computing is virtualization with the help of which the cloud providers offer the on-demand services to the users. These services are Pay-Per-Use services which

---

A. Pathania · K. Kaur · P. Singh (✉)

Department of Computer Engineering & Technology, Guru Nanak Dev University, Amritsar, India  
e-mail: [prabh\\_singh32@yahoo.com](mailto:prabh_singh32@yahoo.com)

A. Pathania  
e-mail: [apathania89@gmail.com](mailto:apathania89@gmail.com)

K. Kaur  
e-mail: [kiran.dcse@gndu.ac.in](mailto:kiran.dcse@gndu.ac.in)

© Springer Nature Singapore Pte Ltd. 2020

M. Dutta et al. (eds.), *Proceedings of International Conference on IoT Inclusive Life (ICIIL 2019)*, NITTTR Chandigarh, India, Lecture Notes in Networks and Systems 116, [https://doi.org/10.1007/978-981-15-3020-3\\_36](https://doi.org/10.1007/978-981-15-3020-3_36)

are flexible, reliable, and available. Instead of purchasing the high-cost infrastructure and maintaining it, organizations can make use of cloud platform. There exists pool of resources associated with cloud and these resources are labeled as local or global pool resources [1]. Local pool resources belong to same cluster and can be accessed by clients belonging to same cluster without authorization. While the global pool resources belong to distinct clusters, these resources are outside the premises of client. In order to access these resources proper authorization is required. The cost encountered while accessing such resources is high [2], and in order to tackle this issue, optimum allocation policies are worked upon where global pool resources are accessed only if local resources are unavailable.

However, reliability is a major concern in cloud computing since the resources and mass requirements may vary. The required resources for the desired period of time may not serve the purpose of client and hence reliability is an issue that is considered in this research. One of the possible mechanisms to enhance reliability is to make use of load balancing strategies [3]. Several mechanisms are prescribed for balancing the load on virtual machines (VMs) and though cloud is considered to be containing virtually infinite resources but still load asserted by various clients makes it fail in one way or the other [4].

This has led to the creation of large-scale data centers which contain large number of nodes which consumes enormous amount of electricity hence leading to environmental degradation. This inefficient usage of resources is the main reason for this high energy. The handling and managing of the over-provisioned resources result toward the high cost. Another problem is the servers, even the ideal servers consume up to 70–75% of the power. Hence keeping the servers underutilized is highly inefficient. High-energy consumption by the infrastructure leads to carbon dioxide emission which leads to greenhouse effects.

With the help of the virtualization, multiple virtual machines can be created on a single physical server which improves the utilization of the resources. This also leads to increased Return on Investments (ROI). Switching the ideal nodes to sleep or hibernate mode can help to achieve reduction in energy consumption. Live migration is one technique, using which the VMs can be consolidated to least number of nodes based upon their requirement. With application's increased demand that leads to rise in the resource utilization can cause performance degradation in VMs. If the application's resource requirements are not met, then the application faces increased time outs, failures, response times, etc. Hence the providers focus on the reduced energy consumption which is a part of Service Level Agreement (SLA).

This paper pays stress on determining accurate load on servers and then initiates migration depending upon the migration time and the utilization of the virtual machine. The load balancing strategies considered for evaluation are broker dependent. The enhancement to load balancing strategies leads to migration strategies. Rest of the paper is organized as follows: Sect. 2 discusses earlier work that focuses primarily on load balancing strategies. Section 3 provides in depth the research methodology followed. Section 4 provides performance analysis and results. Finally, we make concluding remarks in Sect. 5.

## 2 Review of Literature

This section is categorized into two subsections. In the first subsection, we discuss the various work done in load balancing strategies. While the second subsection discusses about migration.

- (a) **Load Balancing Strategies:** In this subsection, we discuss the various contributions made in the field of load balancing in cloud environment.

Nathuji and Schwan [5] were the first to consider the issue of power management of virtualized data centers. Their proposed technique divided the resources into local and global level. The operating system's power management strategies were supported by local level. While the global manager gets the information from the local managers regarding the current resource allocation and then applies its policies to decide if the VM placement needs to be considered.

Pinheiro et al. [6] worked on minimization of power consumption which serves multiple web applications. In their technique the idle nodes were turned off. The load of the nodes was monitored periodically to take the decision to turn off the idle nodes. This check helps to minimize the overall power consumption. However, their technique suffered from a major shortcoming as the algorithm used to run on master node, it created the problem of single point failure.

Chase et al. [7] worked on the problem of homogeneous nature of resources in Internet hosting centers. Since determining the resource demand of the application and allocating the resources is a major challenge. Therefore, they applied a technique where the services used to bid for quantity of resources. The system serves the request for each service by maintaining an active set of servers. By switching the ideal servers to power saving modes the energy consumption was reduced. The web workload was targeted by the system which includes noise in the data. This problem was addressed by applying "flip-flop" filters.

Kusic et al. [8] used heterogeneous environment to define the problem of the power management and addressed it using Limited Lookahead Control (LLC). To estimate the future requests and the future state, Kalam filters are applied to perform the reallocations. Due to model's complexity the execution time was more.

Zhu et al. [9] worked on the issue of automated resource allocation and capacity planning. They considered different time scale ranging from hours to seconds. This places workload onto the groups of servers, reallocates the VMs, allocates resources to the VMs, and applied an approach of setting up the fixed utilization threshold value, which was not that efficient.

Berral et al. [10] used the machine learning technique to optimize energy consumption and SLA. Their proposed approach was designed for the High-Performance Computing (HPC) where concept of deadline constraint occurs making it unsuitable for the mixed workload.

Keahey and Deshpande [11] considered the migration of the multiple virtual machines from multiple source hosts to multiple destination hosts taking network contention into consideration.

Beloglazov and Buyya [12] proposed an algorithm which is used to reduce energy consumption while maintaining high-level SLA. Their approach was based upon a novel adaptive heuristics for dynamic consolidation of VMs which is used to analyze the historical resource usage of the VM.

(b) **Migration:** Since, the literature discussed in the above subsection mainly focused on load balancing strategies that ultimately lead to migration strategies. Generally, load balancing strategies are categorized as the following [13–16]:

- Time Quantum-Based Load balancer
- Weighted Round Robin Load balancing
- Least Connection-based Load balancing
- Weighted Least Connection
- Agent-based Load balancing approach.

The migration is initiated if any of the above load balancing strategies fail. The migration strategies are commonly used for fault tolerance and increasing reliability. Migration is required if deteriorated virtual machines or servers are detected [17, 18]. The migration mechanism could be data oriented or virtual machine oriented [19, 20].

**Data Migration** [21, 22]: Data migration mechanism shifts the data from current server or virtual machine to next virtual machine. The data centers in the cloud act as resource provider and virtual machine mechanism divides the data centers into a set of virtual machines. Time-critical applications often require that virtual machine to complete the assigned task within prescribed deadline. This task execution requires data and if virtual machine is not executing task within deadline then data migration is also required. Moreover, data migration plays a critical role in reactive fault tolerance.

**VM Migration** [23, 24]: VM migration is required incase virtual machines become unusable. The VM migration mechanism could be live or offline. In offline migration, source and destination virtual machines must be switched off while performing transformation. This leads to increase in downtime and migration time. In live migration source and destination both are online in nature. This means that downtime and migration time shows great deviation as compared to offline deviation. Live VM migration, however, suffers from overhead problem of maintaining database and storage resources all the time. A possible solution to overcome this problem is to use hybrid migration that uses advantages of both offline and online migration.

**Core Migration:** The core level migration is within the local pool of virtual machines. The virtual machines are divided into cores as per requirements of the client. The dynamic requirements cause virtual machines to split into parts where each distinct part is known as core. Core migration mechanism initiates if one part of virtual machine fails then other core within same virtual machine is selected for migration. The core level migration is divided into static and dynamic parts. Static cores are fixed that means core division is fixed to specific quantity and in dynamic core division, core partitioning depends upon the requirements of the clients. The reliability in both the cases is high but overhead in dynamic cores is more as compared to static core migration.

So taking inspiration from above studies, migration mechanism in the proposed system is dynamic and online facility of virtual machines is used for decreasing downtime and migration time. Through this research we try to fulfill two main objectives: (a) To reduce the overall energy consumption in the cloud environment. (b) To increase the overall efficiency of the cloud system.

### 3 Research Methodology

Multiple hosts are considered during migration process. Physical machine selected as a host for migration does not perform any migration. In other words, this host is a static host that is optimal enough without deterioration affects. The simulation setup for the proposed mechanism is described in Table 1.

The simulation mechanism first of all initializes the Cloudsim. Once the Cloudsim is initialized, datacenter is created. The datacenters provide resources to the virtual machines. The virtual machines are given proportionate resources corresponding to the datacenter. The broker initialized for the simulation determines the optimal VM and assign cloudlets to the VM. The VM capacity is the critical parameter determining load over the host machine. As the load over the host increases, host may deteriorate. This requires a special host machine that is used for hosting the migrated task. This machine has high memory capacity for holding the task list that is received from deteriorated machines.

The entire implementation of the proposed work is done using Cloudsim 3.0. Host-based migration is considered in the proposed system for achieving optimality of result. With the decreasing migration time and the increasing utilization, Energy consumption is decreased to some extent and if the expected migration time of any VM is less as compared to the other VMs, it will be migrated to the destination host. The flowchart of the proposed work is given in Fig. 1.

The proposed mechanism follows the broker aware monitoring that allows the load to be balanced in case threshold load is exceeded. In case overloaded host is detected, host-level migration is initiated. The overloaded host list is maintained to prevent the load allocation on the overloaded host. The starvation problem and reliability are resolved using the proposed mechanism. This is demonstrated using the result and performance analysis.

**Table 1** Simulation setup for the proposed mechanism

Nodes/software	Configuration
5 Hosts	I3 processor, with Varying virtual machines with each host
1 Physical Machine	Used for migration
Virtual Machines (VMs)	Varying distinct VMs
Netbeans with Cloudsim	Version 8.2 and 3.0

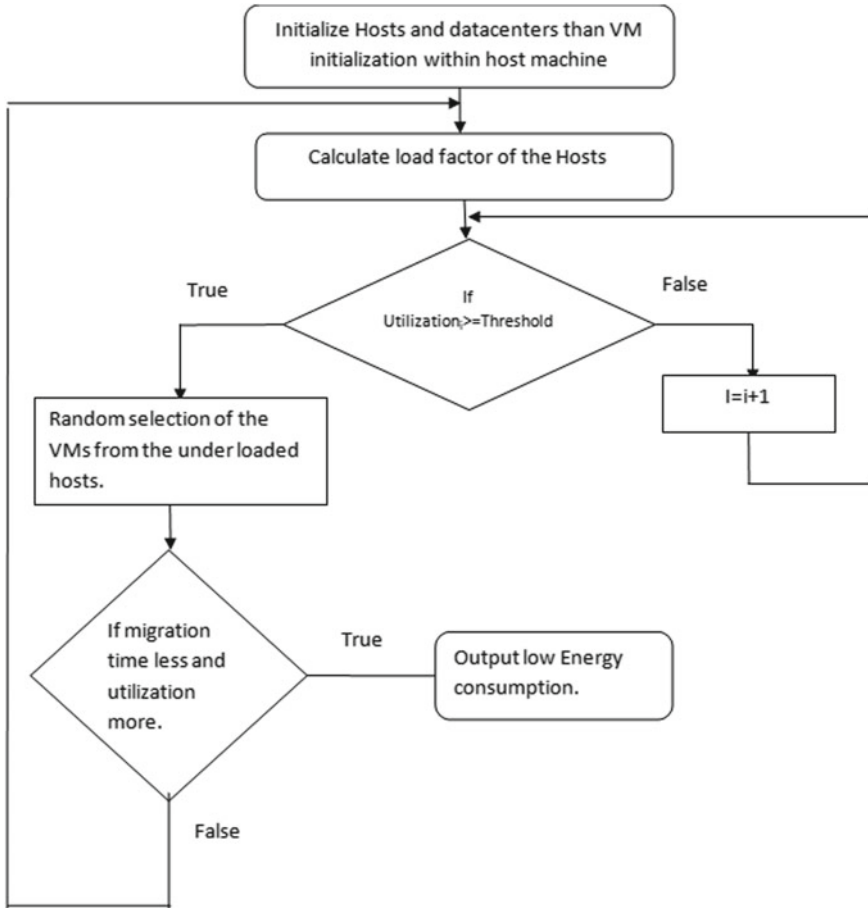


Fig. 1 Flowchart of the proposed work

### 4 Performance and Result

The performance and result analysis indicate the performance betterment in terms of downtime and migration time. The downtime indicates amount of time during which virtual machine does not perform any operation. Migration time indicates the amount of time that is required for transferring the memory and processing elements toward the destination host. Migration time and downtime are being optimized using the proposed mechanism.

To accomplish the task of comparing the efficiency of our algorithm, we have made use of several metrics. These metrics include Energy Consumption by the servers; another metric is VM migration time and the last metric is Utilization of the VMs which is considered high. With the combination of the decreased VM migration

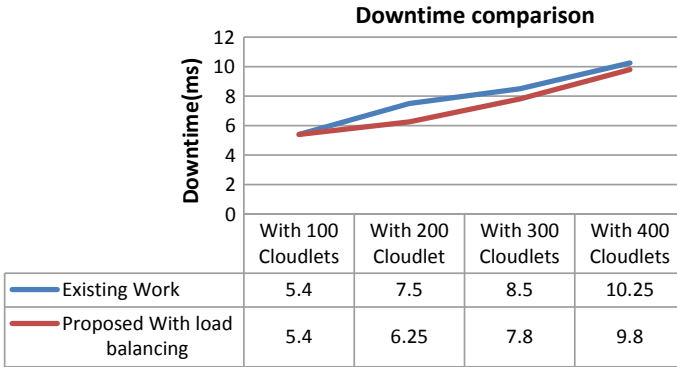


Fig. 2 Downtime of existing and proposed mechanism

time and the increased utilization of the VM, Energy Consumption is decreased. The result of downtime comparison is shown in Fig. 2.

It is observed that at lower cloudlet count, the value of downtime does not show much variation but at high cloudlet count downtime shows significant deviation. The proposed mechanism reduces downtime as the load on the host increases. The migration time is compared against the existing migration time which is much greater at higher loads as compared to proposed mechanism. This is demonstrated through Fig. 3.

The migration time at significantly low load is similar for both existing and proposed mechanism. At higher load the values show deviation toward the proposed system. The Score factor is given in Eq. 1, where f1 is the migration time (factor 1) and f2 is the utilization (factor 2).  $\alpha$  is the weight of the various VMs.

$$SF = \alpha f1 + (1 - \alpha)f2 \tag{1}$$

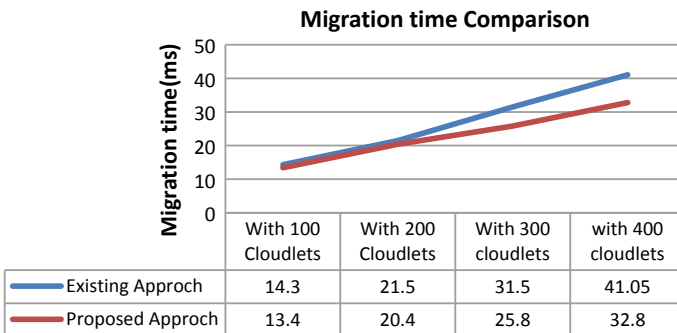


Fig. 3 Migration time comparison of existing and proposed system

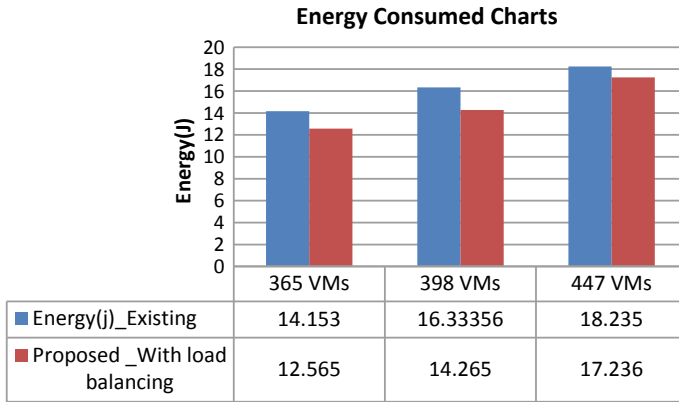


Fig. 4 Comparison of energy consumption

Table 2 Values of various parameters

Number of cloudlets	VM_Migration time (ms)	Utilization	Energy consumption (J)
100	10.235	53.247	2.3
200	15.365	56.784	6.75
300	19.525	52.486	6.24
400	23.365	54.172	7.51

The performance and result are observed in terms of energy consumed which are also optimized through the proposed system. For optimization, this factor must be reduced. By following our proposed approach energy consumption is significantly reduced. The plots in terms of energy consumed are given in Fig. 4. The results show that energy consumed is reduced significantly using our proposed approach.

Through this study, we also tried to established migration time, utilization and energy consumption. The results are shown in Table 2, while they are presented as a correlation plot in Fig. 5. Migration time and energy consumption share a strong positive association of 87%, utilization and energy consumption share a moderate positive association of 38%, while migration time and utilization share a weak negative association of 7%.

## 5 Conclusions

Nowadays cloud computing is providing a cheap alternative to organizations to carry out their operations. However, the problem of load balancing creates a major hindrance at performance level. The aim of this paper was to work on the principle of load balancing and migration. The load on the host is observed, and then threshold values



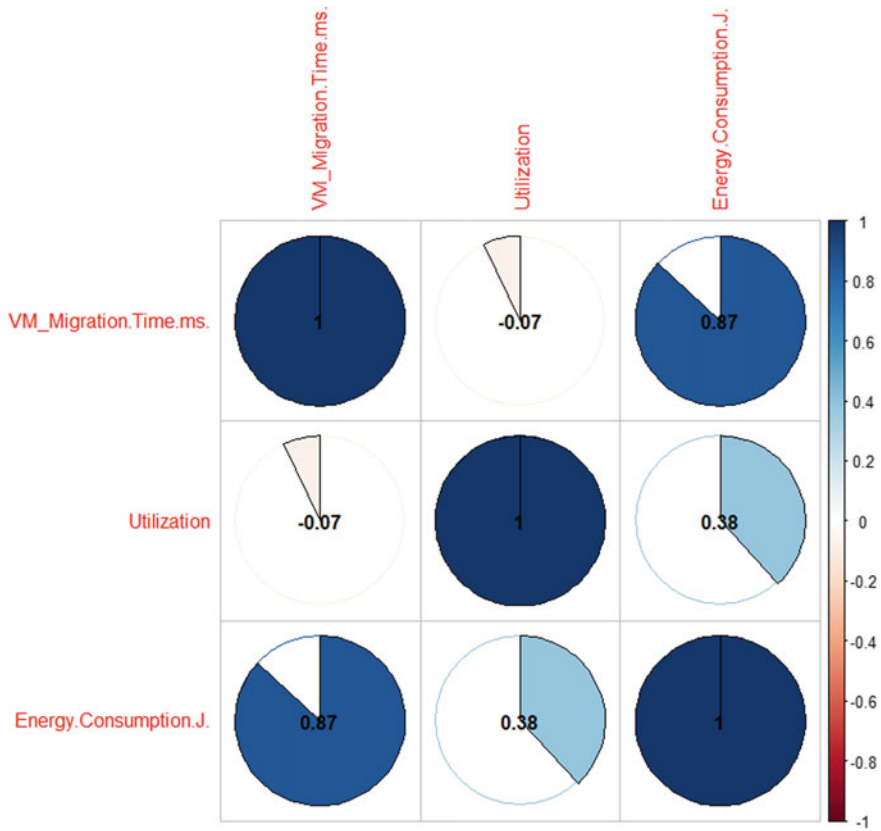


Fig. 5 Relationship plot

are maintained to check overloaded virtual machines. In addition, overloaded virtual machine host deterioration is checked. In case deterioration is detected, hosts which are not overloaded are considered and randomly the VMs are selected from those hosts. The randomly selected VMs migration time and the utilization are checked. The machine with the minimum migration time and the maximum utilization will be migrated to the destination host. The result obtained shows improvement at greater load and hence proves worth of study.

## References

1. M. Kaur, S. Sharma, R. Kaur, Optimization of job scheduling in cloud computing environment. *Int. J. Adv. Res. Comput. Sci. Softw. Eng.* **4**(7) (2014)
2. B. Jennings, R. Stadler, Resource management in clouds: survey and research challenges. *J. Netw. Syst. Manage.* **23**(3), 567–619 (2015)

3. A. Zhou, S. Wang, Z. Zheng, C.H. Hsu, M.R. Lyu, F. Yang, On cloud service reliability enhancement with optimal resource usage. *IEEE Trans. Cloud Comput.* **4**(4), 452–466 (2014)
4. J. Li, S. Su, X. Cheng, M. Song, L. Ma, J. Wang, Cost-efficient coordinated scheduling for leasing cloud resources on hybrid workloads. *Parallel Comput.* **44**, 1–17 (2015)
5. R. Nathuji, K. Schwan, Virtualpower: coordinated power management in virtualized enterprise systems. *ACM SIGOPS Oper. Syst. Rev.* **41**(6), 265–278 (2007). ACM
6. E. Pinheiro, R. Bianchini, E.V. Carrera, T. Heath, Load balancing and unbalancing for power and performance in cluster-based systems (2001)
7. J.S. Chase, D.C. Anderson, P.N. Thakar, A.M. Vahdat, R.P. Doyle, Managing energy and server resources in hosting centers. *ACM SIGOPS Oper. Syst. Rev.* **35**(5), 103–116 (2001)
8. D. Kusic, J.O. Kephart, J.E. Hanson, N. Kandasamy, G. Jiang, Power and performance management of virtualized computing environments via lookahead control. *Clust. Comput.* **12**(1), 1–15 (2009)
9. X. Zhu, D. Young, B.J. Watson, Z. Wang, J. Rolia, S. Singhal, B. McKee, C. Hyser, D. Gmach, R. Gardner, T. Christian, Integrated capacity and workload management for the next generation data center, in *ICAC08: Proceedings of the 5th International Conference on Autonomic Computing* (2008)
10. J.L. Berral, Í. Goiri, R. Nou, F. Julià, J. Guitart, R. Gavalda, J. Torres, Towards energy-aware scheduling in data centers using machine learning, in *Proceedings of the 1st International Conference on energy-Efficient Computing and Networking*, Apr 2010 (ACM, 2010), pp. 215–224
11. U. Deshpande, K. Keahey, Traffic-sensitive live migration of virtual machines. *Futur. Gener. Comput. Syst.* **72**, 118–128 (2017)
12. A. Beloglazov, R. Buyya, Optimal online deterministic algorithms and adaptive heuristics for energy and performance efficient dynamic consolidation of virtual machines in cloud data centers. *Concurr. Comput.: Pract. Exp.* **24**(13), 1397–1420 (2012)
13. B. Meroufel, G. Belalem, Adaptive time-based coordinated checkpointing for cloud computing workflows. *Scalable Comput.: Pract. Exp.* **15**(2), 153–168 (2014)
14. K. Li, Scheduling parallel tasks with energy and time constraints on multiple manycore processors in a cloud computing environment. *Futur. Gener. Comput. Syst.* **82**, 591–605 (2018)
15. A.V. Dastjerdi, R. Buyya, An autonomous time-dependent SLA negotiation strategy for cloud computing. *Comput. J.* **58**(11), 3202–3216 (2015)
16. J. Xu, S. Pears, A dynamic shadow approach to fault-tolerant mobile agents in an autonomic environment. *R.-Time Syst.* **32**(3), 235–252 (2006)
17. P.D. Patel, M. Karamta, M.D. Bhavsar, M.B. Potdar, Live virtual machine migration techniques in cloud computing: a survey. *Int. J. Comput. Appl.* **86**(16) (2014)
18. D. Duolikun, S. Nakamura, R. Watanabe, T. Enokido, M. Takizawa, Energy-aware migration of virtual machines in a cluster, in *International Conference on Broadband and Wireless Computing, Communication and Applications*, Nov 2016 (Springer, Cham, 2016), pp. 21–32
19. B. Zhao, X. Chen, J. Zhu, Z. Zhu, Survivable control plane establishment with live control service backup and migration in SD-EONs. *J. Opt. Commun. Netw.* **8**(6), 371–381 (2016)
20. D. Duolikun, S. Nakamura, T. Enokido, M. Takizawa, Energy-efficient replication and migration of processes in a cluster, in *2015 Ninth International Conference on Complex, Intelligent, and Software Intensive Systems*, July 2015 (IEEE, 2015), pp. 118–125
21. J. Sekhar, G. Jeba, Energy efficient VM live migration in cloud data centers 1 (2013)
22. F. Curzi, M. Ryan, U.S. Patent No. 9,459,856. U.S. Patent and Trademark Office, Washington, DC (2016)
23. N.R. Katsipoulakis, K. Tsakalozos, A. Delis, Adaptive live VM migration in share-nothing IaaS-clouds with LiveFS, in *2013 IEEE 5th International Conference on Cloud Computing Technology and Science*, vol. 2, Dec 2013 (IEEE, 2013), pp. 293–298
24. Q. Wu, F. Ishikawa, Q. Zhu, Y. Xia Energy and migration cost-aware dynamic virtual machine consolidation in heterogeneous cloud datacenters. *IEEE Trans. Serv. Comput.* (2016)