The Challenges in Virtual Machine Live Migration and Resource Management

Priyanka H
CSE R&D Centre,
Dr. Ambedkar Institute of Technology, Bengaluru, India

Mary Cherian
Dept. of CSE
Dr. Ambedkar Institute of Technology, Bengaluru, India

Abstract— **Cloud computing is becoming increasingly popular and is hosted in huge data centres. Such huge data centres have a huge number of physical servers that are, in effect, virtualized into a large number of Virtual Machines (VMs) managed by cloud infrastructure such as open stack, cloud stack, etc. Virtual Machine migration has become an important criterion for saving resources, growing resource utilization, and maintaining the quality of service (QoS) in cloud data centres with the increasing growth of cloud environments. Different techniques for migration of VMs are established for the efficient use of resources. This paper addresses the live migration of the virtual machine, with its components, drawbacks and resource management.**

Keywords **- Virtual machine migration, Cloud computing, Virtualization, Live migration, Resource Management.**

I. INTRODUCTION

Cloud computing is a methodology for conveying information technology-friendly software innovation; resources are retrieved from the internet by digital devices and applications, rather than storing in a database. In comparison to keeping records on an exclusive local hard drive or community space, cloud-based ability makes it possible to store records in a remote server. It is called cloud computing because the data being accessed can be found in "the cloud" and does not allow a customer to be in a specific location to access this. This form of the system allows members to operate remotely.Cloud Services have been used by many business and organizations. The latest on-demand accessibility of cloud-based computing services has contributed to a significant technical and financial revolution within the IT industry. Conservation of energy is a global problem for the present world. There is a demand for alternative energy resources from various IT industries; it is increasing because the traditional energy resources are utilized exponentially. Many IT innovations are therefore concentrating on designing energy-efficient frameworks, and processes at cloud data centres, although cloud storage is referred to as an inherently energy-efficient platform due to its resource scalability and multi-tenant ability. Highly populated data centers are typically over-supplied and generate vast enormous amounts of energy in a competitive atmosphere between the service providers to ensure 24/7 customer support accessibility.

Virtualization is a basic cloud computing element that divides a physical server into multiple single execution environments by installing a layer called a hypervisor or Virtual Machine Manager (VMM). Figure 1 shows the Virtualization, each execution situation, i.e. Virtual Machine (VM) which runs on any OS and applicationswithout any mutual interference. It makes it possible for users to use multiple host setups for managing multiple virtual machines. Virtual machines offer mobility and reliability in exchange for essential computing resources (memory, bandwidth, CPU). Virtualization makes it possible for the virtual machine to switch from one physical host to another. If due to shortage of resources, the server fails to provide Service Level Agreement (SLA) assurances, the hosting VM will move to another server with sufficient resources. Section 2 discusses the related work. Section 3 is about VM migration. Section 4 describes the resource management. Section 5 gives a comparison of different VM migration techniques. Section 6 discusses the research gaps. Section 7 describes the further work of the study and Section 8 concludes the entire paper.

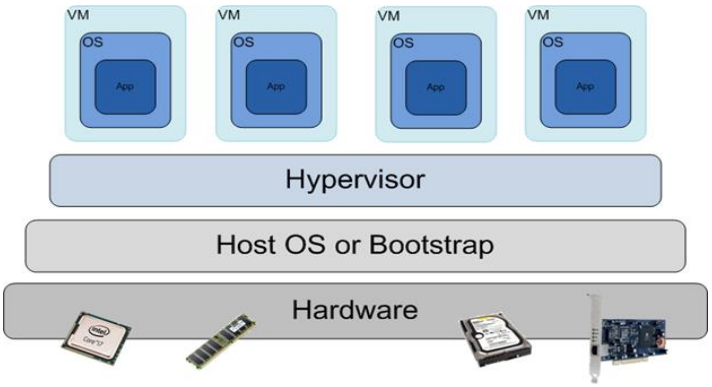

Figure 1. Virtualization

II. RELATED WORK

Zhen Xiao, and Qi Chen et al [1], the estimation of resource allocation status is updated based on anticipated future demands from VMs. It does not offer any demand or request forefficient prediction.
Shuang Chen, Yanzhi Wang, MassoudPedram et al [2], the method described in this paper, is required to address the analysis of power hierarchy in data center and also complex power models.
Huiwang, and HuagloryTianfield et al [3] VM positioning and HS policy for migration, called Space Aware Best Fit Decreasing (SABFD) is proposed. This carries out the migration process many times.
In [4], the authors proposed approximation unit embedded to forecast the system's potential loads, and then two schedulers were called to schedule, respectively, the anticipated and unpredicted loads. The Scheduler uses the technique of column generation to handle the problem of linear/quadratic integer optimization.
In [2], the Ant Colony Optimization (ACO) algorithm is proposed. The algorithm is represented through constant monitoring of resource usage by regional migration agents and then initiating virtual machine migration.
In [6], the authors suggested a task-based scheduling algorithm with techniques for "particle swarm optimization". Rather than transferring the entire virtual machine, the algorithm moved extra tasks from an overloaded VM
In [7], the authors suggested the principle of the queuing and allocation of bandwidth algorithms. The theory of the queuing modeled the relation between VMs. The algorithm enhanced the performance of live virtual machine.
In [8], the authors suggested algorithm is based on task allocation with workload transformation in VMs. The workload resizes the relocation of the virtual machine.
In [9], bee colony algorithm is proposed, which takes the job precedence of VMs standing in queues into consideration. The product reliability (QOS) is optimized by the proposed algorithm.
In [10], the authors introduced a meta-heuristic composite algorithm integration of ant colony with optimization(ACO) of particle swarm. Tominimize the calculation time and boost virtual machine migration time, the algorithm is attached to a PSO operator of ACO algorithm.
In [11], The proposed Branch-and-Bound based algorithm reduces the data center energy. Due to high computation time, it is not suitable for a large cloud data center.

III VM MIGRATION

A. *What does VM Migration mean?*

Migration of virtual machines seeks to shift a virtual machine from one physical device to another. It is part of operating applications for virtualization of hardware

1) *Types of VM migration*
   - *a) Cold migration* – Stop VM on the source machine, and again start on destination
   - *b) Warm migration* – Suspend VM on machine 1, copy across RAM and CPU registers, restart on machine 2 (some seconds later)
   - *c) Live migration* –It is of two types:
     - i) *Post Copy*: In this method [20], VM stops execution on origin machine and its execution environment (registry, memory pages required to start VM on the target machine, CPU,) is transferred to target machine and VM starts its execution, without the whole replacement of memory pages.
     - ii) *Pre-copy:* Pre-copy is used by hypervisors like Xen, KVM and VMware. It is one of the live migration techniques. There are two phases in this method.
       - *I) Warm-up phase***:** The memory pages are transferred from origin machine to target machine. The VM at the source is not stopped until the required number of memory pages is duplicated to the target machine. If the pages are changed, then it should be updated at the target machine. The dirty page condition will occur if the rate of copying page to the target host is less than the rate of sending a modified page.
       - *II) Phase of stop and copy***:** After the completion of the warm-up phase, the VM is stationary at the source machine and it starts executing at the destination machine. The disadvantage of this approach is that modified pages have to resend iteratively leading to inefficiency as it consumes more network bandwidth.

B. *Live virtual machine migration components:*

On VM migration [21][22], recognizing which content needs to be transferred or what needs to be flocked is highly critical. In the migration process, it is important to research the lower limit of operator interruption by checking, how migration manages CPU state, memory information, and space content.

1) Memory content. Data storage is a greater amount of information that includes data and guest OS RAM on the VM running processes. The VM is designed with a large amount of memory, but VM may not use it to its fullest, so there is no need to migrate unused memory. The following is the memory module to be transferred during the migration process:

1. Memory built by VM: The amount of actual physical memory provided by the hypervisor to the guest VM. This memory was used by the VM as their physical memory.
2. Memory distributed by the hypervisor: It is part of the memory configured by VM and is regularly used by VM, but its volume is smaller than the memory configured by VM.
3. Memory used by VM: This type of memory is used frequently through the Virtual machine operating system. The guest VM keeps track of memory pages.
4. Application Memory: This memory is used to run the applications within the VM. It is not compulsory that all allotted memory must be in the physical memory, but it is permitted to be in disk storage as well.

2) Storage content: For ensuring the quality of data, a huge amount of information needs to be transmitted. Also, huge work is required within the network for the hard disk document change. To minimize the waiting time, hypervisor must differentiate the unwanted storage occupants and idle space to prevent migration. This enables the specific type of processing information content to be transferred to lower the migration time as follows:

1. ***Virtual Disk Size:*** The VM is assigned a disk size of its usage, which is known as virtual disk size. The VM size is described upon the VM creation. The hypervisor allows to select available disk space upon VM creation or permits VM to expand dynamically.
2. ***Blocks used by VM*****:** Classifies the blocks which are used by the VM and lists the free storage area.
3. ***Blocks distributed by hypervisor*****:** When pre-allocation comes to an end, a VM is assigned for data storage via the hypervisor. If in casethe VM releases some blocks, the hypervisor will not reduce the assigned block sizebecause the VM stage is invisible to be checked. So avoidingthe idle space and garbage pool will significantly reduce the migration time.

IV RESOURCE MANAGEMENT

The quality of VM migration will be influenced by a range of factors. This section focuses on investigating the relationships between the efficiency of migration and the relevant factors. These are summarized as follows:

- *Bandwidth replication:* This is the bandwidth between the origin of network and the destination server. Migration in LAN system has a significantly higher bandwidth than WAN.
- *Characteristic of workload:* Moving applications in the migrated VM will affect the performance of migration. The systemwill become complex due to increased network size and CPU usage. The suitable migration strategy should be applied to the appropriate resources.
- *Effective host efficiency:* The available distribution resources and ideal servers are also critical for migration success since migration strategies depend on available resources.
- *VM parameters:* It implies fixed VM parameters consisting of memory size, utilized memory area, virtual disk width, virtual disk format, network buffer size, etc.
- *Virtual Machine Monitor (VMM):* Migration approach is widely utilized in VMM. The implementation is different for each VMM. This is accompanied by one of the migration technique.

A. Resource Allocation in Cloud Computing Systems forEnergy Cost Minimization

The energy cost minimization is further classified into two groups: 1) Binpacking-based static resource allocation and 2) Threshold-based dynamic resource consolidation.

1) Bin Packing-Based Static Resource Allocation[10]: Latest studies have shown that in cloud computing, physical machines (PMs) use a large amount of energy. Intuitively, the problem of static energy-conscious resource allocation can be represented as the problem of bin-packing, where VMs and PMs are the objects and bins in the problem of bin-packing. To reduce theusage costs the bin packing problem is divided into 2 steps:

i)First lower bound on the optimal solution (LB-OPT) is calculated

ii) Design of a polynomial resource allocation algorithm. By considering the multiple types of resources of PMs, an efficient VM allocation approach must be applied to improve the imbalanced utilization of the multiple resources.

i) *Threshold-Based Dynamic Resource Consolidation:* Live migration software has proven effective in tackling complex resource consolidation, enabling a VM to be transferred from one Physical Machine (PM) to another PM. During VM migration, the system should move VMs on the low resource utilized source PM to another target PM; this allows the source PM to switch off without consuming any power. On the other hand, the system should also avoid the target PM over-utilization. The threshold-based approach towards resource reduction has been studied to achieve this goal. This approach works by predetermining the first two thresholds, the higher the threshold $t_h$and the lower thresholdt_l. When the resource utilization of a PM p_i, exceeds t_h, the system will transfer some VMs on p_i to another PM for hotspot avoidance. When the resource utilization of p_i falls below t_l, the system will migrate all

of the VMs on p_i to another target PM for energy saving. In dynamic cloud systems, to predict the two thresholds t_h and t_l precisely, there exists a central manager that monitors and maintains information about all PMs and VMs. As an alternate approach to the central manager is, allowing the PMs to manage resources in a distributed manner, thereby improving system robustness. Recently developed probability-based distributed VM consolidation approaches only focused on reducing energy cost, ignoring its negative effect on violating QoS.

V. COMPARISON OF VM MIGRATION TECHNIQUES

We address the comparison of different techniques used for live migration of VMs with their basic techniques and their advantages and limitations.

TABLE I VIRTUAL MACHINES MIGRATION TECHNIQUES WITH THEIR PROS AND CONS.

Algorithms	Technique	Description	Results	Drawbacks
"Pre-copy"[2]	Technique of Pre-copy	The memory is first transferred and then the execution is transferred.	Down time < 1 sec.	More number of duplicate page is transmitted.
"Post-copy"[17]	Technique of Post-copy	The execution of the VM is transferred first and later memory is transferred.	Memory transferred in a single pass. Less network overhead	Down time is more.
Particle Swarm Optimization(PSO)[5]	Particle Swarm Optimization Technique	It does not move entirely VM, but only the extra tasks running on overloaded VM is moved.	Minimizes execution time of tasks. Time consumption and the energy consumption is reduced.	Partial optimization is obtained.
Bee colony algorithm [9]	load balancing using the bee colony algorithm	Dependent on the prioritization of tasks. The lowest priority task is selected for removal.	Minimum waiting time, Reduces response time. Maximizes throughput.	Energy consumption, VM Mips rate, and VM cost not considered.
Genetic (GA)[15]	load balancing of virtual machine resources using GA	It maintains the load on the physical machine. So it should not be overloaded	It provides the best load balance. The problem of load mismatch is solved.	Time-Consuming because it has no bounds.
non-live migration [20]	non-live migration	The entire VM stops its execution on the origin machine and it is transferred to the target machine.	Reduces the number of VM migrations.	Waiting time is more.
Modified Best Fit Decreasing(MBFD)[18]	Modified best fit decreasing algorithm	Virtual machine dynamic relocation helps in significant energy savings.	Minimizes Energy Consumption. Provides reliable QOS.	Network interface and disk storage not considered

VI RESEARCH GAPS

Limitations of virtual machine live migration

1. High WAN Bandwidth: Due to the huge image size, live virtual machine cannot migrate over low bandwidth WAN and generates high latency.
2. The fault of the network: In the post-copy, if the VM tries to retrieve pages which is unavailable, VM redirects to the source host and thus causes network error when VM stops.
3. Network memory status: Live transition monitors CPU action and data transfer among group of machines operating as VDSM (Virtual Distributed Shared memory) system.
4. Overlapping of post-copy migration technique :The overlapping of post-copy migration technique which occurs as a result of page errors. This is created for varying time intervals during VM performance on the target host. VM stops at the target host every time and it continues to wait for the retransmission from the source host for the necessary memory pages.
5. Reduce the overall migration cost: The data transfer need to be reduced (CPU state and memory pages) between host groups over the network.
6. Page fault detection replication: It is a major disadvantage of the post-copy strategy and it must be configured to decrease the quantity of data retransmission cycles from the source host. This would improve the quality of the technique for post-copying migration.
7. Workload of various types: If the VM has various types of workload, it consumes a huge quantity of memory.
8. Availability of resources: It helps to take an appropriate decision on how to allocate the resources and when to migrate the VMs.
9. Enclosing of network address: One of the issues with WAN migration, is that, the address of the VM is enclosed from the source server to the destination server complicating the WAN status and the WAN connected LANs. Other problems include networking faults[23], overloaded VMs, memory and data intensive applications[24, 25], bandwidth usage and cloud resources.

VII DIRECTIONS FOR THE FUTURE WORK

In future research, based on the limitations that are identified in the research gaps, powerful migration model need to be developed.

- For addressing heavy workloads, filtering and reducing technique can be used to allocate the data evenly among all the places so that workload can be shared equally. It uses less amount of memory.
- Another problem in live migration is low bandwidth; the network bandwidth can be well-utilized by allocating it dynamically.
- In order to reduce the complexity of the post-copy technique, memory and CPU transfer needs to be further optimized as a single VDSM system.
- Page failure detection is done by hiding actual page and creating virtual page, which can be implemented in the cloud environment to manage target node failure in the post-copy migration approach.
- In cloud environment, there are few network-conscious migration techniques available for VM migration. Pre-copying and post-copy migration approaches can be used as a hybrid technique. During the VM migration,this will yield best results if the memory reuse concept is used during hybridization.

VIII CONCLUSIONS

From this study, it can be decided that VM live migration is a significant part of the research that can increase the overall performance of data centers in the cloud environment. Optimal load balancing is a key feature of the data center. Cloud computing technologies including virtualization environment, live virtual machine migration modules, and resource management are explained in this paper. Several problems with VM migration strategies are addressed.Comparison of different migration techniques with their drawbacks is mentioned. The major issues in the live migration process are illustrated, indicating that an ideal model of VM migration has to be delivered.The research gaps in the VM migration and also future research directions are highlighted.

REFERENCES

[1] Zhen Xiao, Senior Member, IEEE, Weijia Song, and Qi Chen, "Dynamic Resource Allocation Using VirtualMachines for Cloud Computing Environment," IEEE transactions on parallel and distributed systems, vol. 24, no. 6, pp 1107-1117 IEEE, JUNE 2013
[2] L. Weining and F. Tao, "Live migration of virtual machine based onrecovering system and CPU scheduling," in 6th IEEE joint InternationalInformation Technology and Artificial Intelligence Conference,Piscataway, NJ, USA, May 2009, pp. 303-7.
[3] Huiwang andhuaglorytianfield, "Energy-Aware Dynamic Virtual Machine Consolidation for Cloud Datacenters", IEEE transactions, volume 6, 2018.
[4] Shahinvakilinia, behdadheidarpouandmohamedcheriet"Energy-Efficient Resource Allocation in Cloud Computing Environments"IEEE Access, special section on future networks: architectures, protocols, and applications,Vol 4, 2016.
[5] Ramezani, Fahimeh, Jie Lu, and FarookhKhadeerHussain, "Task-based system load balancing in cloud computing using particle swarm optimization," International Journal of Parallel Programming 42, no. 5 pp. 739-754,2014.
[6] Gutierrez-Garcia, J. Octavio, and Adrian Ramirez-Nafarrate, "Agent-based load balancing in Cloud data centers," Cluster Computing 18, no. 3 pp. 1041-1062, 2015.
[7] Zhu, Changpeng, Bo Han, Yinliang Zhao, and Bin Liu, "A Queueing-Theory-Based Bandwidth Allocation Algorithm for Live Virtual Machine Migration," In2015 IEEE International Conference on Smart City/SocialCom/SustainCom (SmartCity), pp. 1065-1072. IEEE, 2015.
[8] Bo Li, Jianxin Li, JinpengHuai, TianyuWo, Qin Li, LiangZhong (2009), "EnaCloud: AnEnegy-saving Application Live Placement Approach for Cloud Computing Environments", In IEEE International Conference on cloud Computing 2009, pp. 17-24.
[9] Babu, KR Remesh, Amaya Anna Joy, and Philip Samuel, "Load Balancing of Tasks in Cloud Computing Environment Based on Bee Colony Algorithm." In2015 Fifth International Conference on Advances in Computing and Communications (ICACC), pp. 89-93. IEEE, 2015.
[10] Komarasamy, and VijayalakshmiMuthuswamy, "A Novel Approach for Dynamic Load Balancing with Effective Bin Packing and VM Reconfiguration in Cloud," Indian Journal of Science and Technology 9, no. 11, 2016.
[11] Sarker, TusherKumer, and Maolin Tang, "Performance-driven live migration of multiple virtual machines in datacenters," In Granular Computing (GrC), 2013 IEEE International Conference on, pp. 253-258. IEEE, 2013.
[12] Hu, Jinhua, JianhuaGu, Guofei Sun, and Tianhai Zhao. "A scheduling strategy on load balancing of virtual machine resources in cloud computing environment." In Parallel Architectures, Algorithms and Programming (PAAP), 2010 Third International Symposium on, pp. 89-96. IEEE, 2010.
[13] M. Duggan, J. Duggan, E. Howley, and E. Barrett, "An autonomous network aware vm migration strategy in cloud data centres," in Proceedingof the International Conference on Cloud and Autonomic Computing(ICCAC-2016), Sept 2016, pp. 24–32.
[14] N. K. Sharma and G. R. M. Reddy, "A novel approach for multidimensional variable sized virtual machine allocation and migration at cloud data center," in 2017 9th International Conference on CommunicationSystems and Networks (COMSNETS), Jan 2017, pp. 383–384.
[15] Safwat A. Hamad and Fatma A. Omara,"Genetic-Based Task Scheduling Algorithm in Cloud Computing Environment",International Journal of Advanced Computer Science and Applications, Vol. 7, No. 4, pp 550- 556, 2016
[16] B.Hu, Z. Lei, Y. Lei, D. Xu and J. Li 2011. A TimeSeries Based Precopy Approach for Live Migration ofVirtual Machine. IEEE 17th International Conference on Parallel and Distributed Systems, pp 947-952.
[17] R. H. Michael, D. Umesh, and G. Kartik 2009. Post-copy live migration of virtual machines. ACM SpecialInterest Group on Operating Systems, pp14-26.
[18] Magalhães, Deborah, Rodrigo N. Calheiros, RajkumarBuyya, and Danielo G. Gomes. "Workload modeling for resource usage analysis and simulation in cloud computing." Computers & Electrical Engineering,Elsevier, pp 69-81, 2015.
[19] L. Weining and F. Tao 2009. Live migration of virtual machine based on recovering system and CPUscheduling. 6th IEEE joint International Information Technology and Artificial IntelligenceConference,Piscataway, NJ, USA, 303
[20] FahimehFarahnakian, Adnan Ashraf, TapioPahikkala, PasiLiljeberg, JuhaPlosila,IvanPorres, and HannuTenhunen, "Using Ant Colony System to Consolidate VMs for Green Cloud Computing", IEEE Transactions on Services Computing, Volume: 8 , Issue: 2 , March-April 1 2015 ,pp 187-198
[21] L. Haikun, J. Hai, L. Xiaofei, H. Liting, and Y. Chen, 2009. Live migration of virtual machine based onfull system trace and replay. 18th ACM International Symposium on High performance distributed computing, pp 101-110.
[22] Soramichi Akiyama, Takahiro Hirofuchi, Ryousei Takano and Shinichi Honiden, "MiyakoDori: A MemoryReusing Mechanism for Dynamic VM Consolidation", 2012 IEEE Fifth International Conference on CloudComputing
[23] R[42] K. Sato, H. Sato, and S. Matsuoka, "A model-based algorithm foroptimizing I/O intensive applications in clouds using vm-based migration," in 2009 9th IEEE/ACM International Symposium on Cluster Computing and the Grid, CCGRID 2009, pp. 466-471.

[24] J. T. Piao and J. Yan, "A network-aware virtual machine placement andmigration approach in cloud computing," 9th International Conferenceon Grid and Cloud Computing, GCC 2010, pp. 87-92. H. Michael, D. Umesh, and G. Kartik, "Post-copy live migration of virtual machines," SIGOPS Oper. Syst. Rev., vol. 43, pp. 14-26, 2009.
[25] H. Stuart, Beno, and H. t, "Improving the live migration process of largeEnterprise applications," 3rd International Workshop on Virtualization Technologies in Distributed Computing, Barcelona, Spain: ACM, 2009.