Virtual Machine Migration Techniques in Cloud Environment: A Survey

Rinal M. Chawda[1] Ompriya Kale[2]
[1]M. E.(C.E.) [2]Assistant Professor
[1, 2]L.J.I.E.T. , Ahmedabad

Abstract— Cloud is an emerging technology in the world of information technology and is built on the key concept of virtualization. Virtualization separates hardware from software and has benefits of server consolidation and live migration. Live migration is a useful tool for migrating OS instances across distant physical of data centers and clusters. It facilitates load balancing, fault management, low-level system maintenance and reduction in energy consumption. In this paper, we survey the major issues of virtual machine live migration. There are various techniques available for live migration and different parameters are considered for migration.

Key words: Cloud computing; Virtual machine, Live migration; Pre-copy; Post-copy; Network aware Migration

I. INTRODUCTION

With the rapid development of internet technology and network technology, more and more people use the internet to obtain information, shopping and entertainment. The growth of the amount of data and the number of user requests will be explosive, which require the computing and processing of the servers higher. The servers need response to client requests within the shortest possible time in order to improve the user experience.

A new computing model cloud computing generates in this case. Cloud computing is an emerging computing model based on the development of distributed computing, parallel processing and grid computing [1]. It distributes the computing tasks to the resource pool made from a large number of computers, and it makes the various application systems use the computing power, storage space and a variety of software and services as they need.

Cloud computing delivers services like software or applications (SaaS – Software as a Service), infrastructure (IaaS - Infrastructure as a service), and platform (PaaS - Platform as a service). Computing is made available in a Pay-As-You-Use manner to users. Some common examples are Google's App Engine [1], Amazon's EC2 [2], Microsoft Azure [3], IBM SmartCloud [4].

Virtualization technology is key concept of cloud computing. Virtualization is currently becoming increasingly popular and valuable in cloud computing environments due to the benefits of server consolidation, live migration, and resource isolation.

In this paper we survey on the performance technologies of the VM live migration. We discuss live migration techniques was applied to Cloud computing. This paper is organized as follows. Section 2 gives a brief introduction of Virtual Machine Migration (VMM). Section 3 describes some Live VMM Techniques. We conclude our work in section 4 with future directions.

II. BACKGROUND

A. Virtual Machine Migration

Virtualization is a technology for cloud computing environments. It allows partitioning one physical machine into multiple virtual machines that runs concurrently and also shares the same physical resources. Virtual Machines (VM) can be provisioned on-demand, replicated and migrated. It has the ability to migrating virtual machine from one physical machine to another machine for the purpose of load balancing and also to make the physical machine fault tolerant. Virtual Machine consolidations can be used to reduce power consumption of cloud data centers. For instance, through server consolidation, multiple (virtual) servers can be allowed to run simultaneously on a single physical server [6].

B. Live Virtual Migration

Live migration [8] is emerging as an impressive technology to efficiently balance load and optimize VM deployment across physical nodes in a data center. With the help of live migration,

VMs on a physical node can be transferred to another without shutting down the system.

This is useful in various scenarios.

- VMs on a failing physical node can be migrated to a healthy one.
- Idle VMs on a node can be migrated to others for optimizing resource utilization.
- VMs on a stressed physical node can be migrated across various nodes for load balancing.

On a broad level, live migration process can be classified into two steps - (i) switching the control to the destination (ii) transferring the data (memory/disk) to the destination. The two most common methods of live migration can be easily differentiated based on different iterations of the steps mentioned above.

- Pre-copy - Memory is transferred first and then the execution is transferred .
- Post-copy - Execution is transferred first and then the memory .

1) Pre-copy

The approach behind pre-copy is to transfer the memory to the destination over a series of iterations. The newly written pages are transferred in each iteration and this process is repeated until either the limit on iteration reaches or the final data is too small for causing any network transfer overhead.

1) Memory and VCPUs are reserved on the destination host.
2) When the migration is issued, a check on page writes is initiated and all the RAM contents are transferred to the destination. This is the first iteration.

3) In the subsequent steps, only the pages that have been dirtied since the last iteration are transferred until the iteration limit is reached or the memory of dirty pages in an iteration is low enough.
4) The execution of VM on source is stopped and CPU state, registers, Virtual devices state and last memory pages are transferred to the destination.
5) VM is resumed at the destination. Pre-copy is well proven for read intensive workloads but in case of write intensive VMs, large set of dirty pages result in performance degradation. In worst cases, pre-copy won't even work. This is possible when the data write rate is more than the network bandwidth.

2) Post-copy

In contrary to pre-copy, post copy transfers the VCPU and devices state on the destination in the first step and starts the execution on the destination in the second.

1) Stop the VM at the source.
2) VCPU registers and states of devices are copied to the destination VM.
3) Resume the execution at the destination.
4) If the VM tries to access a not yet fetched page, then a network page fault occurs and the page is transferred to the destination (on-demand paging). This is the most trivial form of post-copy.

Network page faults indeed bring down the performance and increases total VM migration time.

The following metrics are usually used to measure the performance of live migration.

– *Preparation:* In this phase, appropriate resources are reserved on the destination VM and various operations are performed on the source which varies with different live migration approaches.
– *Downtime:* This is time during which the VM on the source host is suspended.
– *Resume:* This step involves the instantiation of VM on the destination host with the same state as suspended source VM. The time for the completion of this phase varies with different approaches.
– *Total time:* The total time taken in completion of all these phases is called Total Migration time and the time taken in second phase is Downtime.[6] Live Virtual Machine Migration Technique

As described above, main two traditional techniques (post-copy & pre-copy)for migration has there advantage and disadvantage. Various variations have been applied to post-copy/pre-copy to improve performance and reduce time. In addition to the on-demand paging, a bulk copy runs in the background which tries to transfer the memory as soon as possible. Adaptive pre paging is another technique where we intelligently identify the pages which may be accessed by the VM in near future and transfer them beforehand. This helps in quick transfer of the VMs working set.

C.Sagana and all [6] present technique to improve the performance of live migration. The key issue in live migration is the total migration time and the downtime. They perform an optimized iterative pre-copy algorithm is used to reduce the dirty rate of VM with the help of Xen. WSClock page replacement algorithm has been proposed along with traditional pre-copy approach that includes preprocessing phase for reducing the amount of transferred memory page and total migration time.

Tin-Yu Wu and all [7] implemented a technique that decreases downtime and total migration time for live migration. They proposed scheme sets the threshold values based on memory modification prediction technique to determine whether the pages must be transferred in the stop-and-copy phase. By integrating the predicted probability of memory modification with our proposed related dirty rate, we can predict the probability of memory modification, find out the pages with the highest number of modifications, and determine the pages suitable to be transferred in the iterative pre-copy phase.

Shashank Sahni & Vasudeva Varma [8] present, hybrid approach that takes the best of both the traditional methods of live migration - pre and post-copy. In this approach, in addition to processor state, they bundle a lot of useful state information. This includes devices and frequently accessed pages of the VM, aka the working set. This drastically reduces the number of page faults over network while they actively transfer memory. Additionally, on every page fault over the network they transfer a set of pages in its locality in addition to the page itself. They propose a prototype design on VM/Qemu and present a comparative analysis of pre-copy, post-copy and our hybrid approach. And ensure good performance in which pre/post copy migration technique fails.

– *Network aware VM Migration strategy*

In Network aware VM Migration strategy, migration of VM performed based on Network traffic and Network latency. And also various strategies apply for migration.

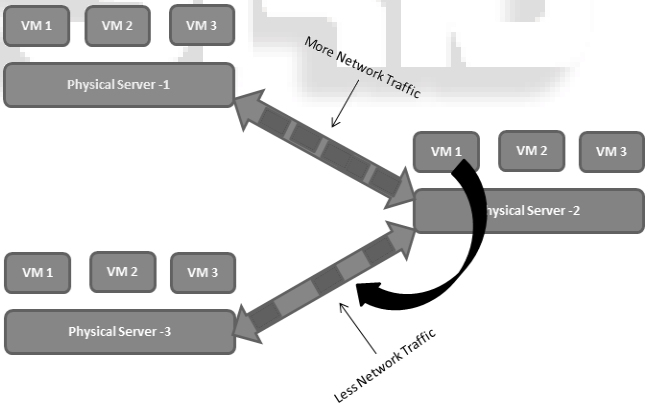

Fig. 1: Network aware Migration [Before Migration] [9]

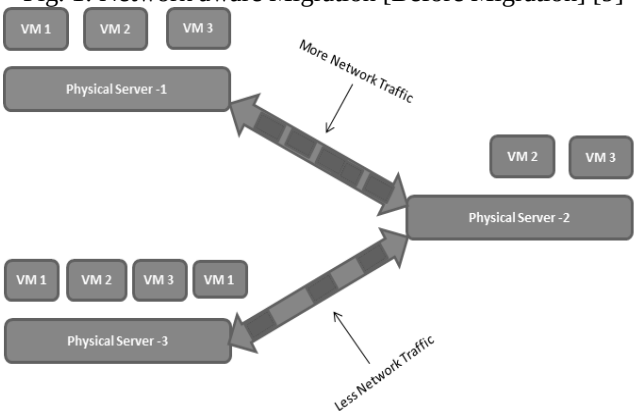

Fig. 2: Network aware Migration [After Migration] [9]

Jing Tai Piao and all [10] present an approach, VM placement and migration for data intensive application, to minimize data transfer time consumption. The proposed approach places the VMs on target physical machines with consideration of the network conditions between the physical machines and the data storage. Thus, this technique helps in optimize the overall application performance.

Xiaoqiao Meng and all [11] presents an approach of manipulating VM placement to address the scalability concern in modern data center networks. This paper proposes using traffic-aware virtual machine placement problem (TVMPP) to improve the network scalability. And this technique achieves significant performance improvement under different traffic pattern and network topology.

Yin Li and all [12] develop a data center deployment scheme for an enterprise to meet its client requirements under uncertain network status. They designed a two-time-scale solution for these problem, long time scale, and short time scale. The data center resizing problem has been considered when the enterprise demands increase such that the current resource configuration cannot satisfy the requirements of its client.

Kento Sato and all [13] present an algorithm for data intensive application.

To find minimum total file access time DAG shortest path search problem is used and also Network throughput, file size and location. Using markov model decision has been taken to determine optimal location of VM to access file.

Alexander Stage and all [14] present network topology aware scheduling models for VM live migrations and classify VM workloads. For each class of VM workloads migration and resource scheduling policy are proposed and network topology & bandwidth are taken into consider.

Method	Reason of VM Migration	Selection of overloaded Physical Machine	Target Physical Machine	Performance metrics use for evaluation
Network aware VM Placement & Migration approach [10]	The data transfer time crosses a certain threshold (T_{SLA}) due to unstable n/w.	When Unexpected network latency or congestion occurs between Host of VM and Data center.	Target physical machine is chosen according to the current network conditions.	The threshold can be determined by a time-related Service Level Agreement (SLA) between the cloud facility provider and the cloud user.
Traffic-aware VM Placement Problem (TVMPP) [11]	VMs on host > the decided the number of VMs .	Request for No. of VM increase then predefine VM.	Randomly select a core switch as an intermediate destination.	Analysis on traffic patterns and the network topology in data centers affect the potential network scalability benefit by optimally solving the TVMPP.
Optimization of data center deployment & VM assignment [12]	When network latency between clients and data centers changes drastically VM migration needed.	If a certain proportion of links change their latency significantly, the assignment optimization and VM migration need to be redone.	Target data center based on down time matrix and total migration cost.	Based on time (sort/long) scale solution are provided.
A Model-Based Algorithm for Optimizing I/O Intensive Applications in Clouds using VM-Based Migration [13]	For data intensive application to minimizing the expected value of file access time VM migration required.	Using markov model, selection of VM which to be migrated is performed. [4(3.2)]	Algorithm is use to find optimal location of VM.	DAG shortest path search problems, on the assumption that the network throughput between sites and the size and locations of target files.
Network-aware migration control and scheduling of differentiated virtual machine workloads [14]	VM workload classifier is use to predict VM migration.	Based on the classified workloads (VMs), the allocation planner predicts over or under load situations that require migrations of VMs and proposes VM re-allocations to other servers.	Allocation planner is used find target host.	
Network-Aware Coordination of Virtual Machine Migrations in Enterprise Data Centers and Clouds [15]	To achieve load balancing and fault tolerance, the workload rise in an application may require Multiple VM migration.	Randomly select host machine	A bin packing algorithm to organize bandwidth resources from all the network links, and allocate them to different migration tasks.	Link sharing is performed and software simulator is use to show that total time for migration task can be minimized.

Table. 1: Comparison Of Network Aware Vm Migration Techniques

Haifeng Chen and all [15] gives technique to perform multiple migration which is helpful for fully utilize available resources in the whole network to achieve the fast completion. For that network link sharing has been

addressed by software simulation, we have proposed a bin packing algorithm to deal with the global resource assignment.

III. RESEARCH CHALLENGES

A. VM Migration SLA based

There is a limitation of migration technology, when it is used for SLA because the optimized data access might still go over the time requirement in the SLA [10].

B. High dirty rate memory page

The problems like the pre-copy (more page fault and higher total migration time) occur when high dirty rate of memory pages although memory threshold technique applied [7].

IV. CONCLUSION

This paper perform a survey of live migration of virtual machine techniques. Live migration involves transferring a running virtual machine across distinct physical hosts. There are many techniques which attempt to minimize the down & total migration time and to provide better performance in low bandwidth environment. There is less number of network aware migration techniques available. In future, work can be done on it.

REFERENCES

[1] Google, "Google App Engine", (2012), [online]. Available: cloud.google.com

[2] Amazon, "Amazon Elastic Compute Cloud (Amazon EC2)", (2012), [online]. Available: aws.amazon.com/ec2/

[3] Microsoft, "Windows Azure.", (2012), [online]. Available: windowsazure.com

[4] IBM, "SmartCloud." (2012), [online]. Available: http://www.ibm.com/cloud-computing/social/us/en/

[5] Ye, Kejiang, et al. "Live migration of multiple virtual machines with resource reservation in cloud computing environments." *Cloud Computing (CLOUD), 2011 IEEE International Conference on*. IEEE, 2011.

[6] Sagana, C., M. Geetha, and R. C. Suganthe. "Performance enhancement in live migration for cloud computing environments." *Information Communication and Embedded Systems (ICICES), 2013 International Conference on*. IEEE, 2013.

[7] Wu, Tin-Yu, et al. "An enhanced approach for estimating writable working set by pre-copy and prediction." *Information Networking (ICOIN), 2013 International Conference on*. IEEE, 2013.

[8] Sahni, S., & Varma, V. (2012, October). A Hybrid Approach To Live Migration Of Virtual Machines. In *Cloud Computing in Emerging Markets (CCEM), 2012 IEEE International Conference on* (pp. 1-5). IEEE.

[9] Leelipushpam, P. Getzi Jeba, and J. Sharmila. "Live VM migration techniques in cloud environment—A survey." *Information & Communication Technologies (ICT), 2013 IEEE Conference on*. IEEE, 2013.

[10] Piao, Jing Tai, and Jun Yan. "A network-aware virtual machine placement and migration approach in cloud computing." *Grid and Cooperative Computing (GCC), 2010 9th International Conference on*. IEEE, 2010.

[11] Meng, Xiaoqiao, Vasileios Pappas, and Li Zhang. "Improving the scalability of data center networks with traffic-aware virtual machine placement." *INFOCOM, 2010 Proceedings IEEE*. IEEE, 2010.

[12] Li, Yin, Min Yao, and Chuang Lin. "Joint study on optimizations of data center deployment, VM assignment and migration." *Quality of Service (IWQoS), 2013 IEEE/ACM 21st International Symposium on*. IEEE, 2013.

[13] Sato, Kento, Hitoshi Sato, and Satoshi Matsuoka. "A model-based algorithm for optimizing i/o intensive applications in clouds using vm-based migration."*Cluster Computing and the Grid, 2009. CCGRID'09. 9th IEEE/ACM International Symposium on*. IEEE, 2009.

[14] Stage, Alexander, and Thomas Setzer. "Network-aware migration control and scheduling of differentiated virtual machine workloads." *Proceedings of the 2009 ICSE Workshop on Software Engineering Challenges of Cloud Computing*. IEEE Computer Society, 2009.

[15] Haifeng Chen, Hui Kang, Guofei Jiang, Yueping Zhang Stage, Alexander, and Thomas Setzer. "Network-aware migration control and scheduling of differentiated virtual machine workloads." *Proceedings of the 2009 ICSE Workshop on Software Engineering Challenges of Cloud Computing*. IEEE Computer Society, 2009.

